



**FACULTAD DE INGENIERÍA, ARQUITECTURA Y
URBANISMO.**

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

TESIS

**ANÁLISIS COMPARATIVO DE ADAPTIVE
BOOSTING Y DECISION TREE PARA PREDECIR
LA MOROSIDAD DE CUOTAS SOCIALES DEL
COLEGIO DE INGENIEROS DEL PERU CONSEJO
DEPARTAMENTAL LAMBAYEQUE**

**PARA OPTAR EL TÍTULO PROFESIONAL DE INGENIERO
DE SISTEMAS**

Autor:

**Bach. Silva Parraguez Máximo Gabriel
ORCID: <https://orcid.org/0000-0001-6496-7710>**

Asesor:

**Mg. Mejia Cabrera Heber Ivan
ORCID: <https://orcid.org/0000-0002-0007-0928>**

**Línea de Investigación:
Infraestructura, Tecnología y Medio Ambiente**

Pimentel – Perú 2022

APROBACIÓN DEL JURADO

TÍTULO DE TESIS:

**ANÁLISIS COMPARATIVO DE ADAPTIVE BOOSTING Y DECISION TREE
PARA PREDECIR LA MOROSIDAD DE CUOTAS SOCIALES DEL COLEGIO
DE INGENIEROS DEL PERÚ CONSEJO DEPARTAMENTAL LAMBAYEQUE.**

Bach. Silva Parraguez Máximo Gabriel

Autor

Ing. Mejía Cabrera Heber Iván.

Asesor

Ing. Mejía Cabrera Heber Iván.

Presidente de Jurado

Ing. Ramos Moscol Mario Fernando.

Secretario de Jurado

Ing. Tuesta Monteza Victor Alexci.

Vocal de Jurado

DEDICATORIA:

A:

Dios, por darme la dicha y oportunidad de llegar a finalizar la carrera, y por estar presente siempre conmigo en cada paso que doy, por darme fuerzas en mi corazón e iluminar mi mente, y también por haber puesto en mi camino a algunas personas que han sido mi mayor motivación, soporte y compañía durante todo este periodo de estudio.

Mis padres, por haber estado siempre conmigo en todo momento, sobretodo por el apoyo económico y moral que me brindaron en todo momento, por sus consejos, sus valores, por la motivación constante que me ha permitido ser una persona de bien, saber mis prioridades, enseñanzas que me servirán para toda mi futura vida profesional.

Mis maestros, por que a lo largo de la vida academica universitaria brindaron sus amplios conocimientos y cultivaron su ética profesional.

AGRADECIMIENTOS:

A mis padres, por su motivación, apoyo y comprensión.

A mi asesor, el Ing. Heber Iván Mejía Cabrera por su orientación y colaboración en el desarrollo de este trabajo de investigación.

RESUMEN:

Esta investigación lleva como título “ANÁLISIS COMPARATIVO DE ADAPTIVE BOOSTING Y DECISION TREE PARA PREDECIR LA MOROSIDAD DE CUOTAS SOCIALES DEL COLEGIO DE INGENIEROS DEL PERÚ CONSEJO DEPARTAMENTAL LAMBAYEQUE” tuvo como finalidad emplear y comparar dos algoritmos de predicción, las cuales fueron Adaptive Boosting (AdaBoost) y Árbol de Decisiones (Decision Tree), para llegar a la conclusión de conocer que algoritmo es más eficiente para predecir o pronosticar la morosidad.

El interés por la presente investigación surge porque hasta el momento, si bien hay estudios parecidos al presente trabajo, los resultados finales no son iguales para todos, es decir, en cada uno de los estudios, la técnica que mejor se comporta siempre depende de los datos y por ende al tipo de estudio que se está realizando, entonces surge así el interés de saber cual de ambas técnicas se comportará más eficientemente en el pronóstico de morosidad en el Colegio de Ingenieros CD Lambayeque.

Entonces, para poder determinar que técnica es la mejor, se evaluó la exactitud, sensibilidad, especificidad y la curva ROC, y se empleó el desarrollo de ETL para la extracción de los datos de los colegiados y sus pagos respectivos por cada mes, y se tomaron los datos más representativos después de varios análisis, todo esto para que sean correctamente procesados por las técnicas, y así obtener resultados apropiados, para que la toma de decisiones sea lo más certera posible.

Al realizar todo el proceso del método propuesto se obtuvieron como resultados que ambas técnicas aplicadas al conjunto de datos asignados son muy parecidos, pero con una pequeña mejora en la técnica de Decision Tree contra Adaboost, con 86.73% contra 86.71%, además cabe destacar que Decision Tree tardó en promedio 0.70 segundos, contra 2.49 segundos de Adaboost.

Por lo tanto, se llegó a la conclusión de que haciendo uso de la herramienta R donde fueron implementadas ambas técnicas, se demostró de que Decisión Tree ganó por poco a Adaboost en Accuracy.

Palabras Clave: Incumplimiento de Pago, Morosidad, Adaptive Boosting, Adaboost, Arbol de Decisión, C5.0, Aprendizaje Automático.

ABSTRACT:

This research is entitled "COMPARATIVE ANALYSIS OF ADAPTIVE BOOSTING AND DECISION TREE TO PREDICT DELAY IN SOCIAL FEES OF THE COLEGIO DE ENGENIEROS DEL PERÚ LAMBAYEQUE DEPARTMENTAL COUNCIL" was intended to use and compare two prediction algorithms (Adaptive Boosting) and Decision Tree, to reach the conclusion of knowing which algorithm is more efficient to predict or forecast delinquency.

The interest in this research arises because so far, although there are studies similar to the present work, the final results are not the same for all, that is, in each of the studies, the technique that behaves best always depends on the data and therefore to the type of study that is being carried out, then the interest arises to know which of both techniques will behave more efficiently in the prognosis of delinquency in the CD Lambayeque College of Engineers.

Then, in order to determine which technique is the best, the accuracy, sensitivity, specificity and the ROC curve were evaluated, and the ETL development was used to extract the data from the collegiate and their respective payments for each month, and They took the most representative data after several analyzes, all this so that they are correctly processed by the techniques, and thus obtain appropriate results, so that decision-making is as accurate as possible.

When carrying out the entire process of the proposed method, the results were obtained that both techniques applied to the assigned data set are very similar, but with a small improvement in the Decision Tree technique against Adaboost, with 86.73% against 86.71%, it should also be noted that Decision Tree took an average of 0.70 seconds, against Adaboost's 2.49 seconds.

Therefore, it was concluded that by using the R tool where both techniques were implemented, it was shown that Decision Tree narrowly beat Adaboost in Accuracy.

Keywords: Default, Late Payment, Adaptive Boosting, Adaboost, Decision Tree, C5.0, Machine Learning.

ÍNDICE

I. INTRODUCCIÓN	14
1.1. Realidad Problemática	14
1.2. Trabajos Previos.	18
1.3. Teorías relacionadas al tema.	26
1.3.1. Inteligencia de negocio.....	26
1.3.2. Modelos de predicción.....	27
1.3.3. Toma de decisiones.	27
1.3.4. Morosidad.	28
1.3.5. Técnicas de predicción.....	28
1.3.6. AdaBoost:	29
1.3.7. Redes Neuronales:	33
1.3.8. Random Forest:	36
1.3.9. Support Vector Machines:	39
1.3.10. Decisión Tree:	40
1.4. Formulación del Problema.	41
1.5. Justificación e Importancia del estudio.	42
1.6. Hipótesis.	42
1.7. Objetivo.	42
1.7.1. Objetivo General.....	42
1.7.2. Objetivos Específicos.	43
II. MATERIAL Y MÉTODO	43
2.1. Tipo y Diseño de la Investigación.	43
2.1.1. Tipo de Investigación.....	43
2.1.2. Diseño de la Investigación.....	43
2.2. Población y Muestra.	44
2.3. Variables, Operacionalización.	44
2.3.1. Variable Independiente.....	44
2.3.2. Variable Dependiente	44
2.3.3. Operacionalización de Variables.....	45
2.4. Técnicas e Instrumentos de recolección de datos, validez y confiabilidad.	46
2.4.1. Técnicas de recolección de datos.....	46

2.4.2. Instrumentos de recolección de Datos.	46
2.5. Procedimiento de Análisis de Datos.	47
2.5.1. Procedimientos para la recolección de datos.	47
2.5.2. Análisis estadísticos e interpretación de los datos	47
2.6. Criterios éticos.	47
2.7. Criterios de rigor científico.	48
2.7.1. Confiabilidad:	48
2.7.2. Validación:.....	48
III. RESULTADOS.	49
3.1. Resultados en Tablas y Figuras.	49
3.2. Discusión de Resultados.....	61
3.3. Aporte Práctico.....	62
3.3.1. Seleccionar las técnicas de predicción a evaluar.	63
3.3.2. Procesar información histórica de base de datos.....	66
3.3.3. Realizar el Proceso ETL para la Extracción de datos relevantes..	79
3.3.4. Aplicar las Técnicas Adaptive Boosting y Decision Tree.	89
3.3.5. Evaluar los resultados Obtenidos.....	137
IV. CONCLUSIONES Y RECOMENDACIONES.....	141
4.1. Conclusiones.....	141
4.2. Recomendaciones:	143
V. REFERENCIAS	144
VI. ANEXOS.....	153

ÍNDICE DE TABLAS

Tabla 1. Rendimiento diagnóstico de la evaluación del riesgo de osteoporosis en múltiples aprendizajes de máquinas y el método clínico OST.....	21
Tabla 2 Operacionalización de variables.	45
Tabla 3 Resultados Finales de Ambas Técnicas	50
Tabla 4 Precisión de ambas técnicas.	51
Tabla 5 Exactitud de Ambas Técnicas.....	52
Tabla 6 Sensibilidad de Ambas Técnicas.	53
Tabla 7 Especificidad de Ambas Técnicas.....	54
Tabla 8 AUC de ambas Técnicas.	55
Tabla 9 Tiempo de Estimación de Morosidad de Ambas Técnicas por cada mes.	58
Tabla 10 Características de las principales técnicas de árboles de decisión.....	65
Tabla 11 Descripción de tablas del modelo Entidad – Relación.	77
Tabla 12 Leyenda de Evaluación de Likert	80
Tabla 13 Puntuación para elegir el modelo dimensional más adecuado al caso de estudio.....	80
Tabla 14 Requisito N° 1	81
Tabla 15 Requisito N° 2.	82
Tabla 16 Requisito N°3.	82
Tabla 17 Requisito N°4.	82
Tabla 18 Requisito N°5.	82
Tabla 19 Requisito N°6.	83
Tabla 20 Requisito N°7.	83
Tabla 21 Requisito N°8.	83
Tabla 22 Requisito N° 9.	84
Tabla 23 Requisito N° 10.	84
Tabla 24 Requisito N°11.	84
Tabla 25 Requisito N° 12.	84
Tabla 26 Requisito N° 13.	85
Tabla 27 Requisito N° 14.	85
Tabla 28 Requisito N° 15.	85
Tabla 29 Campos usados para elaborar el modelo dimensional.	86

Tabla 30 Elección de dimensiones	86
Tabla 31 Archivos .csv para los algoritmos.....	91
Tabla 32 Dividiendo instancias en Grupo para Modelo 1.....	96
Tabla 33 Datos Real y Dato Arrojado por la division de Grupo (RIGHT y LEFT) y ERROR	97
Tabla 34 Hallando WeightedError para Modelo 1	98
Tabla 35 Pesos Actualizados para Modelo 1.....	100
Tabla 36 Dividiendo instancias en Grupos para Modelo 2.....	101
Tabla 37 Datos Real y Dato Arrojado por la división de Grupo (RIGHT y LEFT) y ERROR para Modelo 2	102
Tabla 38 Hallando WeightedError para Modelo 2.....	103
Tabla 39 Pesos Actualizados para Modelo 2.....	105
Tabla 40 Datos Real y Dato Arrojado por la división de Grupo (RIGHT y LEFT) y ERROR para Modelo 2.	106
Tabla 41 Hallando WeightedError para Modelo 3.....	107
Tabla 42 Instancia con los resultados predicción de los Modelos creados en cada Iteración.....	108
Tabla 43 Sumatoria de predicciones de Modelos Creados por cada Instancia..	110
Tabla 44 Tabla Final de Error comparando la salida Original, con la salida Predicha por los modelos.....	111
Tabla 45 Matriz de Confusión Final.	112
Tabla 46 Error de prueba en porcentaje.	114
Tabla 47 Identificando el mejor ajuste del modelo de Adaboost al caso de estudio (evaluando accuracy).....	115
Tabla 48 Datos empleados para ejemplo de C5.0.....	121
Tabla 49. Cantidad de datos de la clase SI y NO para ejemplo C50	122
Tabla 50. Cantidad de datos de la clase SI y NO para ejemplo C50 de la variable Genero.	123
Tabla 51. Datos de la variable Edad.....	124
Tabla 52. Variables con sus ganancias respectivas.	126
Tabla 53. Nuevos datos para calcular la entropía y la ganancia.....	127
Tabla 54. Nueva información de variables y sus ganancias.	128
Tabla 55. Datos para test de predicción del árbol C5.0 y sus resultados (ESTADO	

DE PREDICCIÓN).....	130
Tabla 56. Matriz de confusión de resultados de Predicción de ejemplo de C5.0	133
Tabla 57 Comparando C5.0 sin modificaciones, C5.0 con boosting (10 y 100 iteraciones) y C5.0 con Poda.	136
Tabla 58 Representación de una matriz de confusión según Kohavi & Provost.	138
Tabla 59 Interpretación de los valores generado por la matriz de confusión.	138

ÍNDICE DE FIGURAS

Figura 1. Precisión y Probabilidad de detección (PD) para proyectos tomados. Fuente: Kaur, A. Kaur, I. (2016) “An empirical evaluation of classification algorithms for fault prediction in open source projects”	20
Figura 2. Diagrama de Flujo de AdaBoost.....	33
Figura 3. Diagrama de una Neurona.	34
Figura 4. Diagrama de flujo para entrenar y reconocimiento de la red neuronal.	35
Figura 5. Diagrama de flujo detallado de la red neuronal artificial para el entrenamiento.	35
Figura 6. Diagrama de flujo detallado para el reconocimiento de la red neuronal artificial.	36
Figura 7. Esquema del algoritmo Random Forest.Fuente: (Borisov et al. 2005) .	39
Figura 8. El diagrama muestra el proceso para la implementación de la herramienta de la máquina de vectores de soporte en un Asesor Experto. Recuperado de: https://www.mql5.com/es/articles/584	40
Figura 9. Análisis Comparativo de indicador Precisión (Accuracy) entre resultados promediados de todos los meses usando Adaboost y Decision Tree.	51
Figura 10. Análisis Comparativo de indicador Exactitud (Precision) entre resultados promediados de todos los meses usando Adaboost y Decision Tree.	53
Figura 11. Análisis Comparativo de indicador Sensitivity (Sensibilidad) entre resultados promediados de todos los meses usando Adaboost y Decision Tree.	54
Figura 12. Análisis Comparativo de indicador Especificity (Especificidad) entre resultados promediados de todos los meses usando Adaboost y Decision Tree.	55
Figura 13. Análisis Comparativo de indicador AUC entre resultados promediados de todos los meses usando Adaboost y Decision Tree. Fuente: Elaboración propia.	56
Figura 14. Curva Roc: Respecto a Adaboost. Fuente: Elaboración Propia.	57
Figura 15. Curva Roc: Respecto a Decision Tree. Fuente: Elaboración Propia. ...	57
Figura 16. Análisis Comparativo del tiempo de estimación de predicción de Morosidad (segundos) entre resultados promediados de todos los meses usando Adaboost y Decision Tree. Fuente: Elaboración Propia.	59
Figura 17. Árbol de decision generado en el mes de septiembre por C5.0(Decision Tree).....	60

Figura 18. Diagrama con Nombre: CIP. Elaborado por el Grupo de Investigación	71
Figura 19. Diagrama con Nombre: DIAGRAM_AWB. Elaborado por el Grupo de Investigación.	72
Figura 20. Diagrama con Nombre: DIAGRAM_aCTIVOSfIJOS. Elaborado por el Grupo de Investigación.	72
Figura 21. Diagrama con Nombre: DIAGRAM_UBIGEO. Elaborado por el Grupo de Investigación.	73
Figura 22. Nombres de algunas tablas con la que contaba la base de datos original. Elaborado por el grupo de Investigación.	73
Figura 23. Columnas con la que cuenta la hoja de cálculo mencionado. Elaborado por el Grupo de investigación.....	74
Figura 24. Diagrama Entidad – Relación (Con Datos Ordenados) con nombre: CIP.	77
Figura 25. Pasando datos de fuentes de origen a fuente de destino.....	78
Figura 26. Modelo Físico de Modelo Dimensional CIP_DM.	88
Figura 27. Comparación con Resultados arrojados en R bajo los mismos datos y argumentos	113
Figura 28. Cargando Datos y separando en Entrenamiento (70%) y Test (30%).	116
Figura 29. Creando el Modelo de Adaboost, Evaluandolo con el Test y Generando Matriz de Confusion.	116
Figura 30. Aplicando AUC - Curva Roc	116
Figura 31. Resultado del árbol de decisión generado por R.....	133
Figura 32. Resultado de matriz de confusión y el Accuracy generado por R. ...	133
Figura 33. <i>Cargando Datos y separando en Entrenamiento (70%) y Test (30%).</i>	136
Figura 34. Creando el Modelo de C5.0, Evaluandolo con el Test y Generando Matriz de Confusion	137
Figura 35. Aplicando AUC - Curva Roc	137
Figura 36. Ficha de observación del rendimiento del algoritmo de Adaboost del mes de enero. Fuente: Elaboración propia	140

I. INTRODUCCIÓN

1.1. Realidad Problemática.

Hoy en día, las carencias de las empresas están definidas por el entorno global en el que se ubican, el cual constituye un entorno muy competitivo que requiere la toma constante de decisiones importantes, y con una posibilidad muy baja de errores que involucren atentar contra la viabilidad de una empresa u organización. (Villanueva 2015).

Para ello, la organización debe contar con información relacionada con el medio que lo rodea, pues esta información, diseñada para descubrir nuevas oportunidades de negocio y aclarar el sistema de respuesta a las amenazas, es fundamental para la toma de decisiones de la empresa.(Garzón 2015).

Cuando un gerente enfrenta decisiones en una organización, encima de entender la situación en la que se enfrenta, también debe poseer la facultad de estimar, examinar, recolectar alternativas, tener en cuenta variables, etc. Para encontrar una solución razonable, es decir, trate de tomar decisiones fundamentadas en la razón. (Villanueva 2015).

En un entorno donde la gente piensa que la base de datos es la principal herramienta para proporcionar información sobre cualquier departamento o área, nació la inteligencia de negocios, que es un grupo de herramientas que hace que toda la información almacenada sea significativa. Lo realmente valioso es que ahora se puede hacer cualquier tipo de análisis que consideremos más adecuado y nos ayude a enfocarnos en las diferentes áreas de análisis de la organización para cumplir con sus objetivos. Sin embargo, a pesar de utilizar las mismas herramientas, cada empresa aún puede ajustar sus informes y análisis a sus propias necesidades, creando así una planificación especial para abordar ciertas áreas de oportunidad.(Castro 2015).

En un ambiente empresarial muy variable y altamente competitivo, aquellas empresas que pueden sobrevivir son aquellas que adoptan una postura activa,

es decir, ajustan sus estrategias para adaptarse a los cambios y anticipar nuevas tendencias. Por tanto, el valor de las empresas actuales radica en su posibilidad para generar y gestionar conocimiento de manera renovadora. En este aspecto, la información se está transformando en uno de los bienes más apreciados de la organización. (DIR&GE 2017).

Pero peor que no contar con información a la mano es tener demasiada información y no saber qué poder realizar con ella. La inteligencia de negocios (BI: Business Intelligence) es una forma de solucionar este problema, pues a través de esta información se pueden generar programas de apoyo efectivo para la toma de decisiones, pronósticos, pronósticos e informes, que se convierten sin lugar a duda, en una gran ventaja. La competencia es feroz, todo lo cual sin duda significa un paso en el negocio. Lo fundamental de la inteligencia de negocios es la información, y como uno de sus mayores beneficios el que se puede utilizar en el proceso de toma de decisiones. Actualmente, existe diversos software de inteligencia de negocios con aplicaciones parecidas que se pueden utilizar en distintas áreas funcionales de las empresas, tales como: compras, ventas, almacén, producción, marketing, finanzas, recursos humanos, contabilidad, etc., muchas empresas han aprendido que los sistemas de inteligencia de negocios (BI) se ha beneficiado de la implementación y se prevé que, con el tiempo, se convertirá en algo fundamental para todas las empresas. (Sánchez 2003).

Se prevé que con el pasar del tiempo se transformará en una necesidad indispensable para toda organización, como lo evidencia la encuesta realizada por CIO de Garther, que para el 2018 Business intelligence and analytics se convirtió en el área tecnológica más importante ayudando a las empresas a diferenciarse de sus competidores.(Gartner 2018)

Las empresas tienen cada vez más datos. Contradictoriamente, esto no asegura que puedan poseer información más útil referente a su rubro de negocio. Entonces surge la pregunta de cómo hacer de todos estos datos, un medio para tomar las decisiones correctas para el futuro de la empresa. Sin embargo, es inútil si carece de un sistema de gestión potente que permita ordenarlos y situarlos en contexto. Una de las principales funciones del modelo de datos es

brindar un conocimiento más profundo, por ejemplo, puede mostrarse los patrones de comportamiento del mercado, clientes o competidores. El próximo escalon de funcionalidad del sistema será la detección de amenazas y oportunidades a manera temprana. (DIR&GE 2017).

El sistema de Business Intelligence actual es el resultado de la transformación de las herramientas de generación de informes, que luego incluyen análisis, monitoreo y hoy están integradas con la capa de predicción. Este último supone un verdadero salto cualitativo. (DIR&GE 2017).

Ahora bien, como se mencionó anteriormente, la Inteligencia de Negocios ayuda en muchos aspectos empresariales.

No se debe olvidar que uno de los factores principales para que una empresa logre un buen desempeño es su capacidad de generar efectivo, y esta capacidad es directamente proporcional a la efectividad de la gestión de la empresa para lograr sus objetivos. No hay duda de que los préstamos morosos se ven fuertemente afectados por el ciclo económico, ya que los préstamos en mora han disminuido significativamente durante el auge. (Brachfield 2013).

Por lo expuesto en lo anterior, se da a entender claramente que la morosidad en una organización es causa fundamental para el fracaso seguro de ésta. Ahora bien, en el Consejo Departamental de Lambayeque del Colegio Ingenieros del Perú (CIP CD-Lambayeque), se realizan diferentes actividades relacionadas con los asociados de la orden profesional y que en algunas oportunidades generan una deficiente toma de decisiones y por ende el descontento de los miembros colegiados.

Según un reporte de la situación financiera hasta fines de diciembre del año 2017, uno de los grandes inconvenientes que manifiesta esta institución son las cuentas por cobrar al personal, directores y Gerentes, sumando un total de S/. 6,324,575.41 (en la nota 03). Suma muy perjudicial para el colegio de Ingenieros.

Ahora bien, se puede cometer errores en la toma de decisiones

organizacionales, sin embargo (Delen, Moscato, and Toma 2018) afirma que se puede solucionar en tiempo real incorporando advertencias predictivas de Business intelligence and analytics ayudando a los especialistas en la toma de decisiones.

Por ello es que se requiere implantar una solución donde los encargados de tomar decisiones respecto a la morosidad, las hagan certeramente, por lo que es importante tomar la data histórica que poseen, procesarla y analizarla para poder brindar información a priori sobre la morosidad en esta institución.

Es por ésto que surge la necesidad de analizar aquellos datos llamados históricos con la finalidad de descubrir información valiosa que permita al negocio estar siempre adelante de la competencia y poder tomar decisiones concretas, acertadas y de forma inmediata, es por ello la utilización de la “La inteligencia de negocios hace referencia al desarrollo de transformar datos en conocimiento y éstos en operaciones para establecer una ventaja en los negocios muy competitiva”. (The Data Warehouse Institute 2016).

Las empresas de los principales países ya están implementando inteligencia de negocios para aumentar los ingresos, reducir los costos y adaptar los productos a los clientes. Tanto es así, una encuesta realizada por Accenture Consulting en 2017 mostró que el 69% de las empresas ya están utilizando modelos predictivos y utilizando diversas técnicas para predecir las actitudes, hábitos y preferencias de sus clientes. Inclusive, la revolución del Big Data está animando a las empresas a utilizar estos datos para aplicarlos a problemas actuales, como la prevención de accidentes de tráfico, la delincuencia o los hábitos saludables. (Quantic 2014).

Pero para aplicar estas técnicas junto a Inteligencia de Negocios, se debe tener en cuenta cuáles son las mejores, y aquí está el problema, el hecho que la mejor técnica de predicción para un conjunto de datos en particular, no compromete la precisión y exactitud para otro conjunto de datos cuyos atributos son lógicamente diferente de la otra. (Rathor 2017).

Existen diferentes técnicas de predicción, por ejemplo: Redes Neuronales – Perceptrón, Random Forest, Boosting – AdaBoost, Support Vector Machines

(SVM), Algoritmos Genéticos, Entre Otros decritos en Los antecedentes de estudios.

1.2. Trabajos Previos.

Díaz (2016) en su Tesis:

“TÉCNICAS DE MINERÍA DE DATOS PARA PREDICCIÓN DEL DIAGNÓSTICO DE HIPERTENSIÓN ARTERIAL”. El propósito de su investigación consistió en hallar relaciones y patrones en los datos, lo que permite crear y utilizar modelos para representar el conocimiento, basados en reglas de agrupamiento y árboles de decisión. Los resultados muestran que el método de reglas de asociación es la más precisa para el pre-diagnóstico de hipertensión, con un nivel de confianza del 98,6%.

Montalvo (2016) en su Tesis:

“ANÁLISIS COMPARATIVO DE TÉCNICAS DE MINERÍA DE DATOS PARA LA PREDICCIÓN DE VENTAS” Se propone una comparación analítica entre diversas técnicas utilizadas en el diseño de modelos de predicción de series de tiempo en minería de datos. La creación de un modelo de minería de datos no es un problema, pero puede evaluar qué tecnología es eficaz o tiene un mejor rendimiento en determinados problemas. Las técnicas que se compararon de minería de datos son: regresión, agrupamiento, red neuronal, series de tiempo; utilizado para el área de ventas. La conclusión es que, en respuesta a esta realidad, las técnicas de series temporales se han ajustado a las investigaciones realizadas.

Vicentin (2016) en su Tesis

“PREDICCIÓN DE PRECIOS MEDIANTE MODELIZACIÓN MULTIVARIADA DE SERIES DE TIEMPO. UNA APLICACIÓN AL SECTOR LÁCTEO ARGENTINO” propone una comparación de metodologías VECM y ANN para una predicción más confiable de los precios al productor primario lácteo, primero la metodología VECM es el manejo de la metodología VAR y se utiliza el VECM para agrupar dichas relaciones y poder predecir los valores de las variables involucradas, por lo tanto, para poder predecir desde de un VECM en primer lugar es verificar el orden de incorporación de las series de tiempo

implicadas. La segunda metodología ANN no es necesario tener conocimiento previo sobre las relaciones entre las variables de entradas y también de las salidas, por lo general la metodología ANN son parcialmente flexible y permite realizar generalizaciones e inferir fuera de la muestra. Las ANN pueden ser de distintas formas y son aplicables a diferentes tipos de datos, especialmente a las series de tiempo. Concluyendo que la metodología ANN es más flexible en los supuestos de partida y aplicables también a otros tipos de datos y además es el más apropiado para predecir dicho conjunto de datos.

Jiménez (2007) en su Tesis:

“MODELOS AVANZADOS PARA LA PREDICCIÓN A CORTO PLAZO DE LA PRODUCCIÓN ELÉCTRICA EN PARQUES EÓLICOS”. El objetivo principal es elaborar un nuevo grupo de modelos avanzados para la proyección a corto plazo de la generación de energía de parques eólicos, de modo que los datos disponibles se puedan utilizar para mejorar las posibles previsiones. Los modelos utilizados por el autor para predecir la generación de energía de los parques eólicos son: modelo Arima, sistema de inferencia difusa y red neuronal artificial; se concluyó que los modelos más sensibles a los valores atípicos son el modelo Arima y los modelos fundamentados en redes neuronales. Por otro lado, los modelos fundados en sistemas de inferencia difusos con técnicas de agrupación sustractiva parecen verse menos afectados por valores atípicos.

Kaur, I. & Kaur, A., (2016) en su investigación:

“AN EMPIRICAL EVALUATION OF CLASSIFICATION ALGORITHMS FOR FAULT PREDICTION IN OPEN SOURCE PROJECTS”, en este documento, se han utilizado seis modelos de machine learning para pronosticar la calidad del software en cinco softwares de código libre. Los proyectos de código libre o abiertos son diferentes de los proyectos industriales la cual son fundamental para la investigación ya que los resultados de estos se pueden comparar y se puede repetir la validación. Los proyectos de fuente abierta tienden hacer más creatividad y llegan a tener menos defectos a medida que los defectos se van encuentran y se fijan rápidamente. En este trabajo los conjuntos de datos utilizados son proyectos Java de código abierto como: PMD, EMMA, Find Bugs, Trove y Dr Java. Las seis técnicas de clasificación más conocidas han

sido escogidas para este estudio actual y son: Random Forest, Naive Bayes, Bagging, J48, Regresión logística e IB1. Los resultados muestran que Random Forest and Bagging produce buenos resultados, mientras que Naive Bayes es menos preferible para la predicción.

Classifier/project	PMD		Find Bugs		EMMA		TROVE		Dr Java	
	Acc.	PD	Acc.	PD	Acc.	PD	Acc.	PD	Acc.	PD
Logistic regression	0.77	0.83	0.64	0.48	0.74	0.63	0.99	0.90	0.80	0.16
Bagging	0.68	0.75	0.85	0.80	0.74	0.58	0.97	0.60	0.81	0.26
Naïve Bayes	0.64	0.88	0.62	0.24	0.63	0.16	0.70	0.90	0.46	0.95
IB1	0.78	0.87	0.74	0.64	0.72	0.63	0.98	0.60	0.78	0.47
J48	0.72	0.72	0.85	0.76	0.63	0.44	0.98	0.63	0.83	0.44
Random Forest	0.73	0.73	0.88	0.80	0.77	0.67	1.00	0.90	0.82	0.41

Figura 1. Precisión y Probabilidad de detección (PD) para proyectos tomados.
Fuente: Kaur, A. Kaur, I. (2016) "An empirical evaluation of classification algorithms for fault prediction in open-source projects".

Kim, S. K., Yoo, T. K., Oh, E., & Kim, D. W. (2013) en su investigación:

"OSTEOPOROSIS RISK PREDICTION USING MACHINE LEARNING AND CONVENTIONAL METHODS", en este documento, fueron los primeros en realizar comparaciones del rendimiento de la predicción de osteoporosis entre el aprendizaje automático y la herramienta de decisión clínica (OST), desarrollando y validando modelos de aprendizaje automático para identificar con mayor precisión el peligro de tener osteoporosis en mujeres posmenopáusicas, modelos basados en técnicas populares tales como: máquinas de vectores de soporte (SVM), bosques aleatorios (RF), redes neuronales artificiales (ANN) y regresión logística (LR) basadas en varios predictores asociados con baja densidad ósea, y también lo compara con la capacidad de una herramienta de decisión clínica convencional, la herramienta de autoevaluación de la osteoporosis (OST). Los resultados muestran que SVM tiene un área significativamente mejor de AUC y la propiedad del destinatario ROC que ANN, LR y OST. La validación en el conjunto de prueba mostró que SVM predijo riesgo de osteoporosis con un AUC de 0.827, precisión del 76.7%, sensibilidad del 77.8% y especificidad del 76.0%.

Tabla 1.

Rendimiento diagnóstico de la evaluación del riesgo de osteoporosis en múltiples aprendizajes de máquinas y el método clínico OST.

	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
SVM	0.827	76.7	77.8	76.0
RF	0.824	76.5	76.6	76.5
ANN	0.807	75.2	76.6	74.4
LR	0.809	74.5	77.8	72.7
OST	0.806	74.0	75.4	73.2

Fuente: Kean, Keun y otros. (2013) “Osteoporosis Risk Prediction Using Machine Learning and Conventional Methods”.

Rathor, A. (2017) en su investigación:

“A REVIEW AT MACHINE LEARNING ALGORITHMS TARGETING BIG DATA CHALLENGES”, se centró en la clasificación de los técnicas ML y determinar el técnica más eficiente con la mayor precisión y exactitud. Además de establecer el rendimiento de diferentes técnicas en conjuntos de datos grandes y pequeños. Usaron las siguientes técnicas Neural Networks, Jrip, J48, k nearest neighbor, Random Forest, support vector machine (SVM). Según los resultados se obtuvo que: 1. Con conjuntos de datos más altos, el análisis predictivo es más confiable. 2. Con conjuntos de datos más pequeños, Neural Networks, JRip y J48 muestran una reducción drástica en el porcentaje de errores. 3. Con un conjunto de datos más pequeño SVM y RF muestra alta exactitud y precisión. 4. Se encontró que considerando la precisión para un conjunto de datos más grande, SVM refleja la predicción más alta. 5. Se muestra que SVM es el técnica con la más alta precisión. 6. SVM tiene el mayor nivel de precisión con un tiempo de 0.04 segundos. 7. La red neuronal funciona bien con el gran grupo de datos a comparación del grupo con datos pequeños. 8. el técnica SVM muestra la clasificación más alta y el cuanto mayor es el conjunto de datos, mayor es la precisión. 9. Los técnicas de aprendizaje automático SVM, NB y RF pueden ofrecer alta precisión independientemente del número de atributos e instancias de datos.

Val,N., (2016) en su investigación:

“CLASIFICACIÓN DEL NIVEL DE POBREZA EN LA ENCUESTA ENAHO APLICANDO REGRESIÓN LOGÍSTICA Y ÁRBOLES DE CLASIFICACIÓN CON EL ALGORITMO C5.0” empleó técnicas de minería de datos de aprendizaje supervisado los cuales fueron: regresión logística y árbol de decisión con el algoritmo C5.0, con la finalidad de predecir el nivel de pobreza de los hogares del Perú empleando los datos de la ENAHO 2015. Obtuvo como resultados que ambos modelos predictivos muestran una alta y similar precisión, obteniendo porcentajes de 94.1% para C5.0 y 93,8% para regresión logística.

Trujillo,D., (2017) en su investigación:

“APLICACIÓN DE METODOLOGÍAS MACHINE LEARNING EN LA GESTIÓN DE RIESGO DE CRÉDITO” Tuvo como finalidad evaluar la probabilidad de emplear estos métodos en la gestión del riesgo de crédito. Así mismo, se estimaron y compararon varios modelos de calificación crediticia mediante algoritmos de aprendizaje automático. Por otro lado, se comparan los resultados de estos modelos con la tecnología de modelado de puntuación crediticia tradicional de regresión logística para analizar las ventajas y el progreso del uso de modelos de aprendizaje automático en comparación con la tecnología de modelado tradicional.

Mathew, V., Toby, T., Singh, V., Rao, B. M., & Kumar, M. G. (2017) en su investigación:

“PREDICTION OF REMAINING USEFUL LIFETIME (RUL) OF TURBOFAN ENGINE USING MACHINE LEARNING” Se seleccionó diez técnicas de aprendizaje automático para comparar que tan precisa es la predicción. Los diferentes técnicas se compararon para obtener el modelo de predicción que tiene la predicción más cercano de restante del ciclo de vida útil en términos de número de ciclos de vida. También compara el rendimiento de las técnicas de aprendizaje automático para la predicción. Se usaron los siguientes técnicas Regresión lineal [Linear Regression], máquina de vectores de soporte (SVM) [Support Vector Machine], Árbol de decisión [Decision Tree], K-vecinos más cercanos [K-Nearest Neighbors], Bosque aleatorio [Random

Forest], el algoritmo K Means [The K Means algorithm], Método de aumento de gradiente (GBM) [Gradient Boosting Method], AdaBoost, aprendizaje profundo [Deep Learning] y Anova. Según los resultados se observó que el rendimiento de los diez técnicas fueron consistentes en los cuatro conjuntos de datos diferentes, generando exactitud proporcional para los diferentes técnicas probadas. Se trazaron los cuadrados medios de raíz de los errores y se obtuvo que la técnica de bosque aleatorio presenta los mejores resultados.

Lee, Y., & Min, O. (2017) en su investigación:

“COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS TO URBAN TRAFFIC PREDICTION” se investiga las demandas de tráfico en Seúl, se generaron los datos sintéticos de un simulador microscópico SUMO bien hecho, y finalmente se aplica varios tipos de algoritmos de aprendizaje automático para descubrir el pronóstico de la transmisión de gases y la proyección del tiempo de viaje. Las técnicas usados son Random Forest, Gradient Boosting, K-nearest Neighbor, Multi-Layer Perceptron. En cuanto a los resultados se obtuvo que la técnica de Gradient boosting es superior a otras tres técnicas. También se obtuvo que la técnica K-nearest neighbor es menos adecuado que otros técnicas.

Portugal, I., Alencar, P. & Cowan, D. (2017) en su investigación:

“THE USE OF MACHINE LEARNING ALGORITHMS IN RECOMMENDER SYSTEMS: A SYSTEMATIC REVIEW” se realizó una revisión sistemática de varias técnicas de machine learning en sistemas de recomendación, uno de los objetivos principales fue el de identificar tendencias en el uso o investigación de técnicas de aprendizaje automático en sistemas de recomendación; se describieron un total de 46 técnicas de machine learning.

Brownlee, J. (2013) en su investigación:

“A TOUR OF MACHINE LEARNING ALGORITHMS”, realiza una descripción de varios algoritmos de aprendizaje automático, los divide en 11 grandes grupos, e incluye en cada grupo, varios algoritmos, lo cual ayuda mucho a ser entendidos.

Alfaro, E., & Elizondo, D. (2008) en su investigación:

“BANKRUPTCY FORECASTING: AN EMPIRICAL COMPARISON OF ADABOOST AND NEURAL NETWORKS”, Se realizó una comparativa referente a la exactitud de pronóstico de dos tecnologías (AdaBoost y Red Neuronal) en un grupo de organizaciones de Europa, teniendo en cuenta variables predictivas habituales como ratios financieros, también variables cualitativas como la dimensión de la organización, estructura legal y actividades. Los resultados señalaron que AdaBoost mejora la precisión de la implementación de la red neuronal.

Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., & Han, M. (2018) en su investigación:

“A REVIEW OF DATA-DRIVEN APPROACHES FOR PREDICTION AND CLASSIFICATION OF BUILDING ENERGY CONSUMPTION”, Este artículo revisa los enfoques dominantes basados en datos utilizados en el análisis de energía de edificios bajo diferentes arquetipos y granularidades, incluidos las técnicas de predicción (Red Neuronal Artificial, Máquina de Soporte Vectorial, regresión estadística, árbol de decisión y algoritmo genético) y los métodos de clasificación (K -mean clustering, autoorganización de mapas y agrupación de jerarquías). En los resultados se obtuvo que en la categoría de enfoques basados en datos para la predicción, las RNA obtienen una posición primaria en un gran número de aplicaciones que van desde la previsión de carga, la estimación de potencial de retroadaptación. En cuanto a SVM, el método tiene un simple proceso de capacitación que requiere algunas entradas, mientras que la baja eficiencia de cálculo se ha convertido en un obstáculo en su aplicación, especialmente para el análisis de energía de construcción a gran escala y alta granularidad. La regresión estadística hace una acción muy importante referente a la evaluación de parámetros significativos, el mapeo energético y la evaluación comparativa a escala urbana. Sin embargo, la inexactitud en la predicción a corto plazo y las posibles correlaciones imprevistas entre los predictores seleccionados socavan en gran medida la efectividad de los modelos de regresión en aplicaciones prácticas. En cuanto a DT, se considera una técnica factible para la predicción del consumo de

energía y la construcción de benchmarking de energía siempre que estos casos no involucren datos secuenciales y no lineales. La principal preocupación es que los resultados de DT se basan principalmente en las expectativas y, por lo tanto, la validez de muchos modelos de DT que se cuestionan. GA sobresale al hacer frente a los datos sofisticados y puede proporcionar soluciones óptimas y multiobjetivos, pero GA sufre de baja eficiencia computacional y entregas no únicas. En cuanto a los enfoques basados en datos para la clasificación, la agrupación de K-means, el mapa autoorganizado y la agrupación de jerarquías fueron ejemplificados con enfoques en el reconocimiento de patrones de curvas de carga, mapeo de energía, evaluación comparativa y retroadaptación a macroescala.

Yeh, I., & Lien, C., (2009) en su investigación:

“THE COMPARISONS OF DATA MINING TECHNIQUES FOR THE PREDICTIVE ACCURACY OF PROBABILITY OF DEFAULT OF CREDIT CARD CLIENTS” tuvo como objetivo el caso de los pagos predeterminados de los clientes en Taiwán y compara la precisión predictiva de la probabilidad de incumplimiento entre los seis métodos de extracción de datos. En esa investigación utilizaron las técnicas de: análisis discriminante, regresión logística, clasificador de Bayes, vecino más cercano, redes neuronales artificiales y árbol de clasificación. Se empleó una variable binaria, pago predeterminado (Sí = 1, No = 0), como la variable de respuesta.

Nagalla, R., Pothuganti, P. & Pawar, D. (2017) en su investigación:

“ANALYZING GAP ACCEPTANCE BEHAVIOR AT UNSIGNALIZED INTERSECTIONS USING SUPPORT VECTOR MACHINES, DECISION TREE AND RANDOM FORESTS”, Este artículo trata de la aplicación de tres técnicas de data mining no paramétricas ampliamente utilizadas, y son, Árboles de decisión (DT), Máquinas de vectores de soporte (SVM) y Bosques aleatorios (RF) para predecir el comportamiento de aceptación de brecha del conductor. Este estudio evalúa y compara el rendimiento de SVM, los bosques aleatorios y el técnica del árbol de decisiones para clasificar y predecir la decisión de aceptación / rechazo de la brecha del conductor en las intersecciones no controladas. Los datos en tres intersecciones de cuatro

patas se recolectaron y se usaron para desarrollar los modelos. Dado que las tres son técnicas no paramétricas, no es necesario suponer en la distribución de datos subyacente. En los resultados se obtuvo que si bien las SVM son insensibles al desequilibrio de clase, el árbol de decisiones generado por el técnica CART proporciona información crítica sobre el desarrollo realizado para tomar decisiones empleado por el conductor. Los bosques aleatorios y el árbol de decisión establecen implícitamente la importancia concerniente a los distintos factores que influye la decisión del conductor. Además, los puntajes de habilidad utilizados para validar los modelos revelaron que los modelos SVM y DT se desempeñaron casi de manera similar, mientras que el modelo RF superó a SVM y DT. Si bien las SVM parecen funcionar mejor que los Árboles de decisión, la diferencia no es significativa. Como las SVM son insensibles al desequilibrio de clase en los datos de entrenamiento, se espera que tengan un mejor rendimiento que otros modelos. Sin embargo, las SVM también tienen algunas desventajas sobre los árboles de Decisión. En primer lugar, los modelos Árbol de decisión y Bosque aleatorio brindan información sobre la importancia relativa de las variables, SVM no puede brindar esta información.

1.3. Teorías relacionadas al tema.

1.3.1. Inteligencia de negocio.

Stackowiak, Rayman, Greenwald (2007) En su libro “Soluciones de inteligencia empresarial y almacenamiento de datos de Oracle”, afirma que, la inteligencia empresarial puede definirse como el correcto acceso a los datos correctos imprescindibles para tomar la decisión comercial correcta en el momento adecuado. Los datos pueden ser crudos o podrían haber sido analizados de alguna manera. Tener acceso a dicha información permite la gestión del negocio por hechos en lugar de confiar principalmente en la intuición.

Según Data Warehouse Institute conceptualiza a la inteligencia de negocio de la siguiente manera:

Son los procesos, técnicas y herramientas que convierten los datos en información, transforman la información en conocimiento y ésta a su vez

transforman el conocimiento en objetivos o decisiones para guiar a la organización.

En el documento escrito por (Yglesias, 2008) sobre la inteligencia de negocio operacional, define que:

“La inteligencia de negocios operacional es una perspectiva distinta a una nueva descendencia de sistemas de BI. En este sistema, la toma de decisiones no es solo para la gestión, sino también para la mejora de las operaciones comerciales, para que la organización se permita influir de manera más eficiente y así obtener mejor provecho.”

1.3.2. Modelos de predicción.

En el documento escrito por (Mende, Koschke, Peleska, 2011) sobre “la utilidad de un modelo de predicción de defectos durante las pruebas de integración de HW / SW: un estudio de caso retrospectivo”, define que:

“Los modelos de predicción por defecto intentan identificar los componentes propensos a errores, de modo que estos pueden probarse antes o más en profundidad, y así mejorar el costo-efectividad durante la prueba. Dichos modelos se han investigado extensamente, pero como y cuando son aplicables en la práctica todavía se debate. La aplicabilidad depende de muchos factores, y sostenemos que no se puede analizar sin un escenario específico en mente”.

1.3.3. Toma de decisiones.

En el documento escrito por (Solano, 2003) sobre la toma de decisiones gerenciales, define que:

“La importancia de la toma de decisiones está estrechamente ligada con la postura de quien toma las decisiones en la empresa; así por ejemplo, decidir el mejor momento para efectuar depósitos bancarios de manera regular puede ser sustancial para un mensajero de la empresa, pero no para un gerente”.

En el documento escrito por (Gutiérrez, 2004) sobre la toma de decisiones, define que:

“El proceso de toma de decisiones consiste en la elección entre opciones

o métodos para dar solución a distintas situaciones de la vida. Se puede plantear en diferentes entornos: empresarial, familiar, emocional, laboral, etc., o sea, la decisión siempre está en curso, y la diferencia entre cada decisión radica en el proceso o método de decisión. La toma de decisiones radica fundamentalmente en escoger una de las opciones disponibles para solucionar problemas actuales o potenciales (aunque no exista evidencia de conflictos potenciales).”

1.3.4. Morosidad.

En el documento escrito por (Gonzales y Gómez, 2014) sobre la morosidad, un acuciante problema financiero de nuestros días, define que:

“Desde el punto de vista empresarial, la morosidad son atrasos en el cumplimiento de las obligaciones contractuales, por lo que, consiste en una falta de cumplimiento del acuerdo de pago en la fecha programada. El daño que el acto ilícito puede causar a los afiliados es suficiente para acreditar la aprobación de una Directiva europea, en donde el objetivo primordial es poder disminuirla.”

En el documento escrito por (Vallcorba y Delgado, 2007) sobre Determinantes de la morosidad bancaria en una economía dolarizada, el caso de Uruguay, define que:

“En este marco, el cambio de la falta de compromiso de pago bancario también muestra una conducta cíclica, acompañado del período económico global. En la estado alto de la fase, los hogares y las organizaciones no están sujetos a restricciones crediticias, los ingresos y las ventas continúan creciendo y tienen menos dificultades para cumplir con sus deberes de pago financieros”.

1.3.5. Técnicas de predicción.

En el documento escrito por (Gonzales, 2009) sobre Técnicas de predicción económica, define que:

“Intentan predecir eventos que aún no han ocurrido con la mayor precisión posible. Para hacer estos pronosticos, la adivina emplea una bola de cristal como herramienta, aunque teniendo en cuenta la historia pasada. Por otro lado, los pronósticos realizados por econométristas o estadísticos

se basan en un estudio claro de la información mostrada por eventos que han ocurrido más o menos en un pasado no tan lejano”.

1.3.5.1. Técnicas.

En el documento escrito por (More, 2013) sobre Métodos y técnicas de la investigación científica, define que:

“Es definido como el procedimiento y herramienta que son utilizados para, investigar, observar, entrevistar, adquirir conocimientos y todo lo que obtenemos de ella.”

En el documento escrito por (Latorre y Seco del pozo, 2013) sobre Estrategias y técnicas metodológicas, define que:

“Es conceptualizado como un proceso algorítmico y un conjunto limitado de pasos ordenados y fijados. La secuencia está predeterminada y ordenada, y su correct desarrollo puede resolver de manera segura la tarea o el problema planteado; por ejemplo, sacar la raíz cúbica, realizar sumas, coser botones, integrar, multiplicar, atar cordones de zapatos, realizar cirugías, confeccionar trajes, reparar o reemplazar llantas de automóviles, clases derivadas, hacer cerámica, etc”.

1.3.5.2. Predicción.

En el documento escrito por (Martínez, 2016) sobre Predicción científica y prescripción en mejora genética vegetal en cuanto ciencia aplicada de diseño: El caso de la mejora de frutales del género Prunus, define que:

“La predicción científica puede ser comprendido como la principal finalidad de la ciencia tradicional del mejoramiento genético como ciencia del diseño: es crucial comprender el futuro posible para así realizar la creación de diseños (en este caso más de 12 años para completar) en un árbol de fruta de hueso”.

1.3.6. AdaBoost:

AdaBoost es un algoritmo de machine learning creado por Schapire Robert y Yoav Freund. En sí "AdaBoost" se refiere al impulso adaptativo o estimulación adaptativa. El algoritmo se adapta a los siguientes aspectos: la clasificación posterior echa se adapta para favorecer la clasificación mal realizada de la clasificación anterior. AdaBoost cuenta con mucha

sensibilidad a casos aislados y ruido en los datos. Aunque, para ciertas circunstancias, es menos adecuado para el sobreajuste (sobreentrenamiento) que la mayoría de los algoritmos de machine learning. (Calvo, 2016).

1.3.6.1. Notación Matemática:

AdaBoost se utiliza mejor para aumentar la productividad de los árboles de decisión en problemas de clasificación binaria. AdaBoost fue originalmente llamado AdaBoost.M1 por los desarrolladores de la técnica. Más recientemente, puede denominarse AdaBoost discreto porque se usa para la clasificación en lugar de la regresión. (...). El algoritmo más adecuado y, por lo tanto, el más común utilizado con AdaBoost son los árboles de decisión con un nivel. Debido a que estos árboles son tan cortos y solo contienen una decisión para la clasificación, a menudo se les llama muñones de decisión. (Brownlee 2016)

Cada instancia en el grupo de datos que se entrenará es ponderada. El peso inicial se establece en:

$$weight(x_i) = \frac{1}{n}$$

Donde x_i es la i^{ma} instancia de entrenamiento y n es el número de instancias de entrenamiento.

Se utiliza un clasificador débil (tocón de decisión) sobre los datos de entrenamiento utilizando las muestras ponderadas. Solo se admiten problemas de clasificación binarios (de dos clases), por lo que cada grupo de decisión toma una decisión sobre una variable de entrada y genera un valor de +1.0 o -1.0 para el valor de la primera o la segunda clase. La tasa de clasificación errónea se calcula para el modelo entrenado. Tradicionalmente, esto se calcula como:

$$error = \frac{correct - N}{N}$$

Donde *error* es la tasa de clasificación errónea, *correct* es el conteo de instancias de entrenamiento pronosticadas correctamente por el modelo y *N* es el número total de instancias de entrenamiento.

Esto se modifica para utilizar la ponderación de las instancias de entrenamiento:

$$error = \frac{\sum_{i=1}^n (w_i * p_{error_i})}{\sum_{i=1}^n w}$$

Cuál es la suma ponderada de la tasa de errores de clasificación, donde w es el peso para la instancia de entrenamiento i y p_{error} es el error de predicción para la instancia de entrenamiento i que es 1 si se clasifica erróneamente y 0 si se clasifica correctamente.

Se calcula un valor de etapa (stage) para el modelo entrenado que proporciona una ponderación para cualquier predicción que haga el modelo. El valor de etapa para un modelo entrenado se calcula de la siguiente manera:

$$\begin{aligned} \text{i)} \quad & stage = \ln\left(\frac{1-error}{error}\right) \\ \text{ii)} \quad & stage = \frac{1}{2} \ln\left(\frac{1-error}{error}\right) \end{aligned}$$

Se puede utilizar cualquiera de los dos fórmulas, ya que lo que variaría es el coeficiente de actualización de peso ($stage$). (Alfaro, s.f.).

Donde $stage$ es el valor de etapa utilizado para ponderar las predicciones del modelo, $\ln()$ es el logaritmo natural y $error$ es el error de clasificación errónea del modelo. El efecto del peso de la $stage$ es que los modelos más precisos tienen más peso o contribución a la predicción final. Los pesos de entrenamiento se actualizan dando más peso a las instancias predichas incorrectamente, y menos peso a las instancias predichas correctamente. Por ejemplo, el peso de una instancia de entrenamiento (w) se actualiza usando:

$$w = w * e^{stage * p_{error}}$$

Donde w es el peso para una instancia de entrenamiento específica, e es la constante numérica del número de Euler elevado a una potencia, $stage$ es la tasa de error de clasificación para el clasificador débil y p_{error} es el error que hizo el clasificador débil para predecir la variable de salida para la instancia de entrenamiento. evaluado como:

$$\begin{aligned} p_{error} &= 0 \text{ IF } y == p \\ p_{error} &= 1 \text{ IF } y \neq p \end{aligned}$$

Donde y es la variable de salida para la instancia de entrenamiento y p es la predicción del aprendiz débil. Esto tiene el efecto de no cambiar el peso si la

instancia de entrenamiento se clasificó correctamente y hacer el peso un poco más grande si el alumno débil clasificó incorrectamente la instancia.

Los modelos débiles se agregan secuencialmente, entrenados usando los datos de entrenamiento ponderados. El proceso continúa hasta que se haya creado un número preestablecido de alumnos débiles (un parámetro de usuario) o no se puedan realizar mejoras adicionales en el conjunto de datos de capacitación. Una vez completado, te quedas con un grupo de alumnos débiles, cada uno con un valor de escenario.

Las predicciones se realizan calculando el promedio ponderado de los clasificadores débiles. Para una nueva instancia de entrada, cada aprendiz débil calcula un valor predicho como +1.0 o -1.0. Los valores predichos son ponderados por cada valor de etapa de los estudiantes débiles. La predicción para el modelo de conjunto se toma como la suma de las predicciones ponderadas. Si la suma es positiva, entonces se predice la primera clase, y si es negativa se predice la segunda clase.

1.3.6.2. Pseudocódigo:

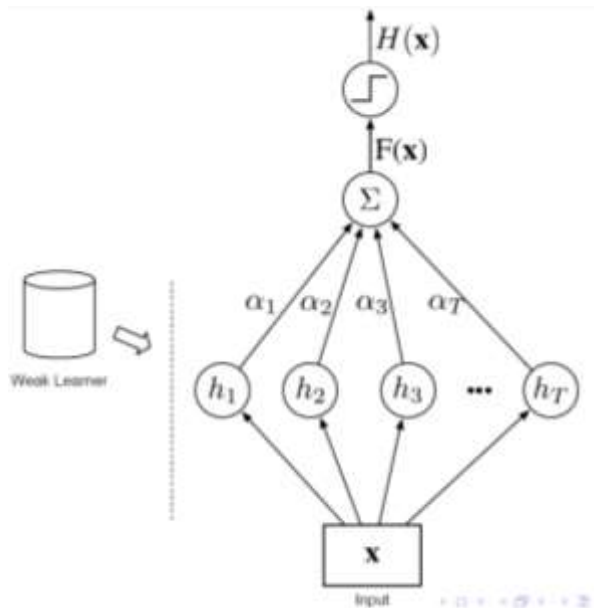


Figura 2. Diagrama de Flujo de AdaBoost.

1.3.7. Redes Neuronales:

Son un método de la Inteligencia Artificial que está usualmente definida como una red compuesta de un gran número de procesadores simples (neuronas) que están masivamente interconectadas, que operan en paralelo y aprenden de la experiencia (ejemplos), estas son las características primarias conocidas de un sistema neuronal biológico que son fácilmente aprovechables en una red neuronal artificial. (Reyes 2017).

1.3.7.1. Notación Matemática:

El modelo más simple de una ANN es una red de una neurona, por este motivo se representa en la Figura 3 sus componentes principales:

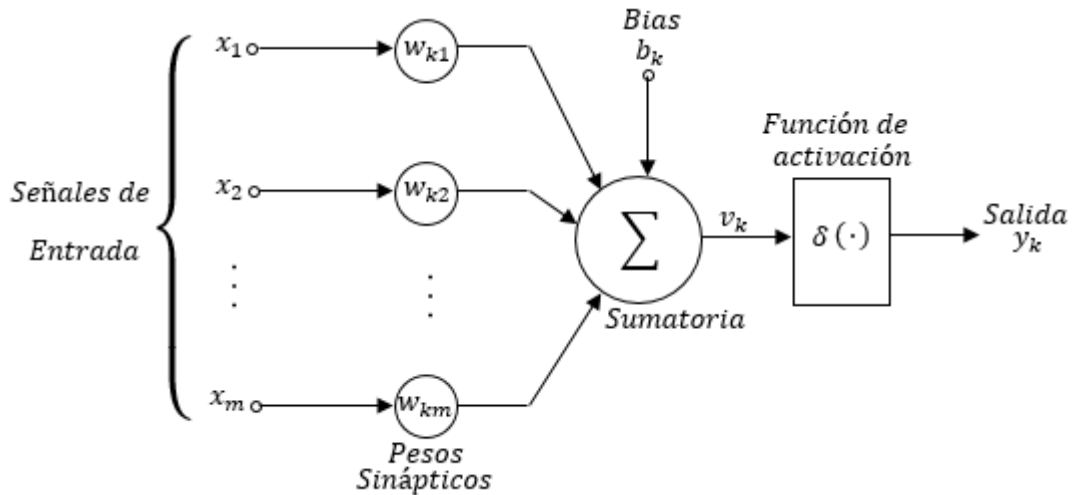


Figura 3. Diagrama de una Neurona.

Donde m es la cantidad de muestras, k el indicador de la neurona, x_1, x_2, \dots, x_m las señales entradas de la neurona, $w_{k1}, w_{k2}, \dots, w_{km}$ el peso sináptico de cada entrada de la neurona k que puede ser una entrada del sistema o estar interconectada a otra neurona, b_k es un bias aplicado externamente cuya función es incrementar o disminuir el ingreso a la red de la función de conexión, una sumatoria encargada de sumar las entradas de la neurona ponderadas por su respectivo peso, v_k denominado como el estado interno de la neurona que es el resultado de la sumatoria, y_k que es la salida de la neurona, descrita por (1) y $\delta(\cdot)$ es la función de activación que se encarga de limitar la amplitud del rango de salida a un valor finito, normalmente normalizado en el rango de $[0, 1]$ o $[-1, 1]$ [50]. Existen diversos tipos de funciones de activación, todas ellas no lineales, de las cuales la más usada para modelos de una capa es el limitador duro y para modelos multi-capas se suelen usar funciones continuas como Gaussiana o Sigmoidea. En términos matemáticos se puede describir la neurona k por medio de :

$$y_k = \delta \left(\sum_{i=1}^m (x_i * w_i) + b_k \right)$$

1.3.7.2. Pseudocódigo:

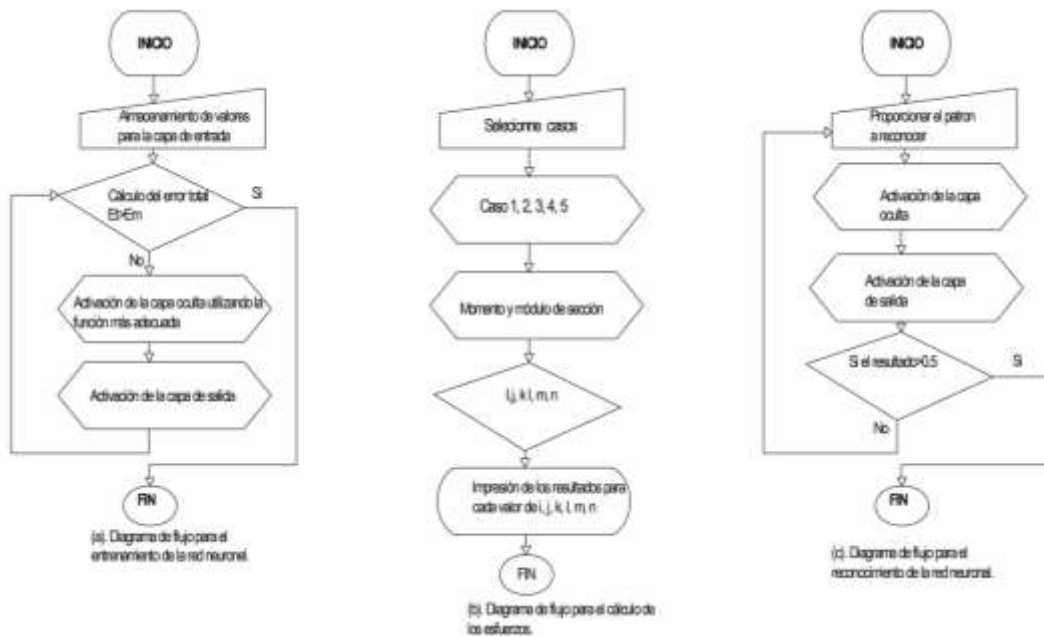


Figura 4. Diagrama de flujo para entrenar y reconocimiento de la red neuronal.

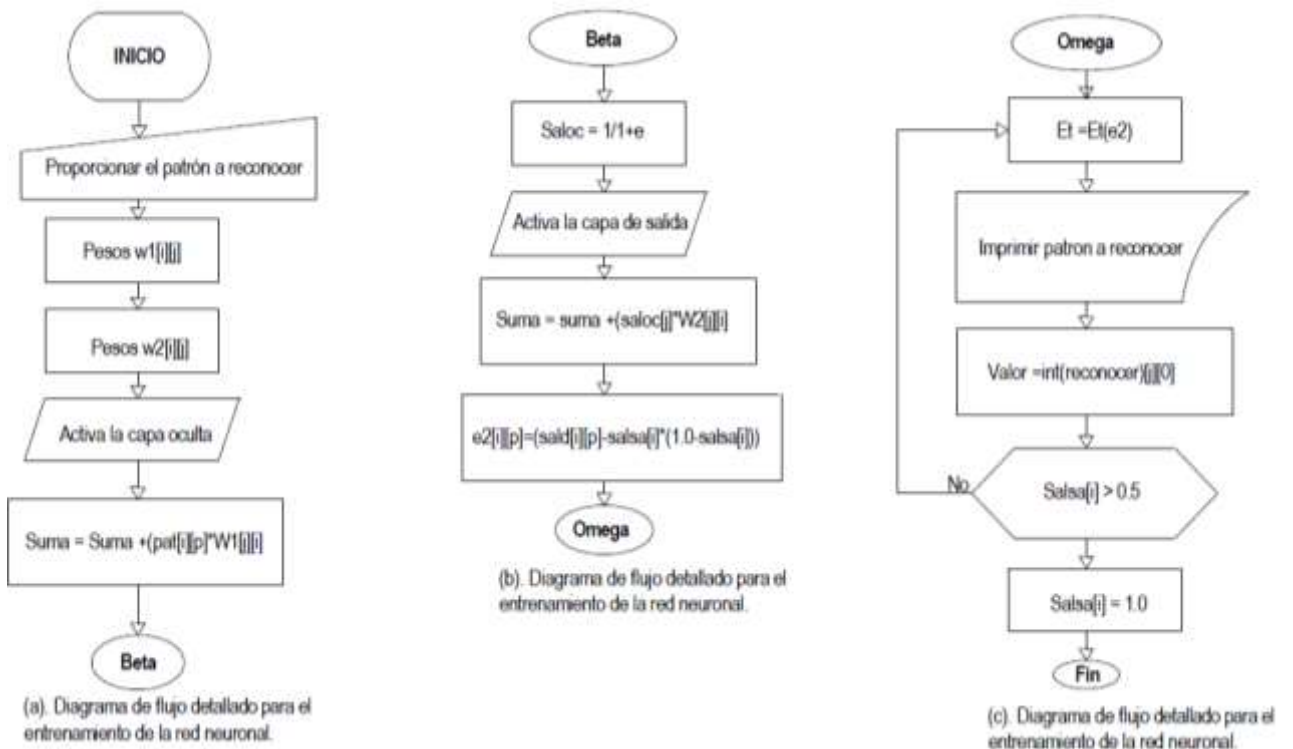


Figura 5. Diagrama de flujo detallado de la red neuronal artificial para el entrenamiento.

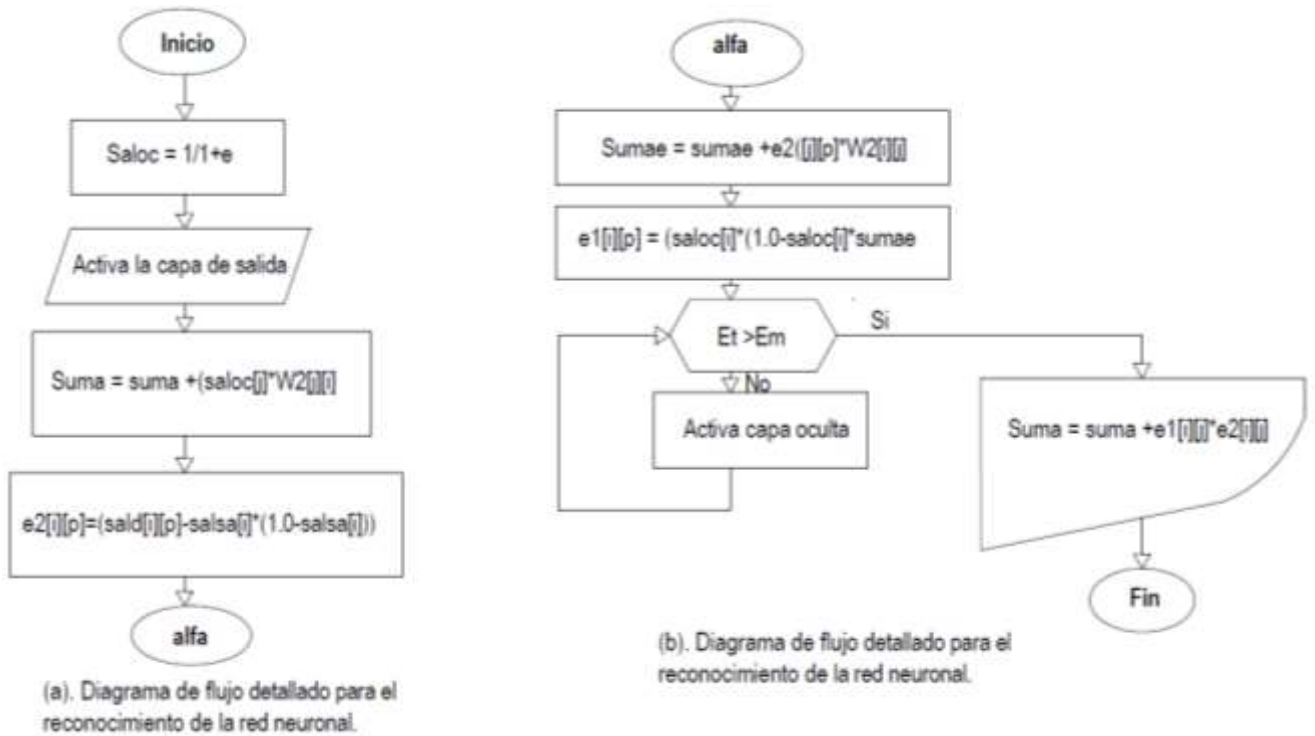


Figura 6. Diagrama de flujo detallado para el reconocimiento de la red neuronal artificial.

1.3.8. Random Forest:

Según (Soltero and Bodas 2005) en su artículo señala:

“Está basado en el desarrollo de varios árboles de clasificación. Así, para catalogar los objetos del vector de entrada, el vector se coloca debajo de cada árbol del bosque. Cada uno de éstos árboles generará un ordenamiento, y el bosque elegirá este agrupamiento, manteniendo en cuenta el árbol con más votos entre todos los árboles del bosque.

El desarrollo de cada árbol es el siguiente:

- Si el total de casos en el grupo de entreno es N, N casos se prueban de forma aleatoria a partir de los datos originales, pero es necesario reemplazarlos. Será éste un grupo de entreno para la elaboración de árboles.
- Si el número de variables de entrada es M, se especifica el número $m \ll M$ para cada nodo, seleccione de forma aleatoria m variables del grupo M y utilice la mejor contribución de m para hacer la división del nodo. Durante el crecimiento del bosque, el valor de m permanece constante.
- Cada árbol crece lo más ampliamente posible sin ninguna poda.”

1.3.8.1. Notación Matemática:

En la etapa donde se entrena, el algoritmo trata de mejorar los indicadores de la función de división de los ejemplares de entrenamiento:

$$\theta_j^* = \operatorname{argmax}_{\theta_j \in \tau_j} I_j$$

Para eso se emplea ésta fórmula de ganancia de información:

$$I_j = H(S_j) - \sum_{i \in \{1,2\}} \frac{|S_j^i|}{|S_j|} H(S_j^i)$$

Entre ellos, representa el grupo de muestras en el nodo que se va a dividir y son los dos grupos de muestras creados a partir de la división. Esta función evalúa la entropía del grupo y es dependiente del tipo de problema que queramos resolver. En cuanto a la regresión, usamos la función de distribución de probabilidad continua para obtener la siguiente expresión:

$$I_j = \sum_{v \in S_j} \log(|\Lambda_y(v)|) - \sum_{i \in \{1,2\}} \sum_{v \in S_j^i} \log(|\Lambda_y(v)|)$$

Dónde está la matriz de co-varianza condicionada.

Una característica que se puede emplear para encontrar la mejor división de nodos es un clasificador simple binario lineal. Por cada uno de los nodo j :

$$h(v, \theta_j) \in j = 1, 2, \dots$$

Entre ellos, el vector indica la muestra que se ingresa y es el parámetro a mejorar en el nodo. Específicamente, compararemos si un determinado valor es menor o mayor que un umbral. Luego los valores deben mejorarse en cada nodo mediante combinaciones de búsquedas arduas. Aún así, es demasiado importante hallar un acuerdo entre esta mejora y el mantenimiento de la forma aleatoria de ciertos indicadores del algoritmo. Luego, en primer lugar, puede deberse a que no encontramos suficientes combinaciones y el split fue demasiado malo. En segundo lugar, se debe conservar un cierto grado de indicadores aleatorios, para así no conseguir un árbol de correlación y desaprovechar el resultado de generalización.

Finalmente, el resultado final se obtiene teniendo como base los resultados que nos brindan los árboles. El desarrollo consiste en que cada hoja de cada uno de los árboles sobre la que cae una muestra invisible puede

considerarse un pronóstico (reparto de probabilidad). Solamente tome la media aritmética de cada distribución para conseguir el resultado del bosque completo.

1.3.8.2. Pseudocódigo:

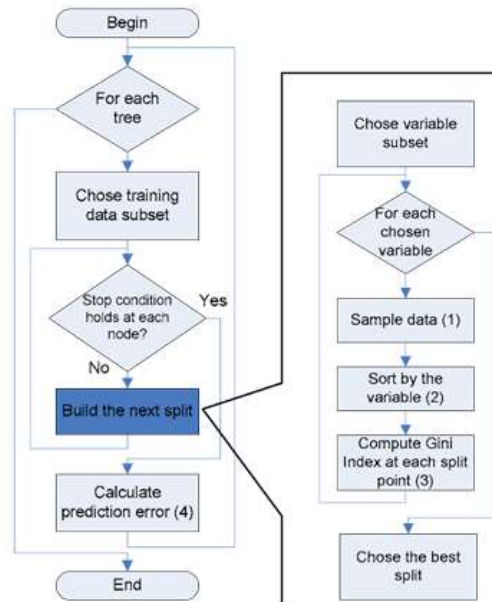


Figura 7. Esquema del algoritmo Random Forest. Fuente: (Borisov et al. 2005)

1.3.9. Support Vector Machines:

El SVM (máquina de soporte vectorial) aprende la cabida de decisión de dos tipos diferentes de puntos de entrada. Como clasificador de clase única, la especificación proporcionada por los datos vectoriales de soporte puede crear un límite de decisión en torno al rango de los datos de aprendizaje con poco conocimiento de los datos fuera del límite. Estos datos se asignan a un espacio de características de mayor dimensión a través de núcleos gaussianos u otros tipos de núcleos, y en este espacio se busca el intervalo máximo entre clases. Al regresar al espacio de entrada, esta función de límite puede dividir los datos en todas las clases diferentes, y cada clase constituye un grupo. (Betancour 2005)

1.3.9.1. Pseudocódigo:

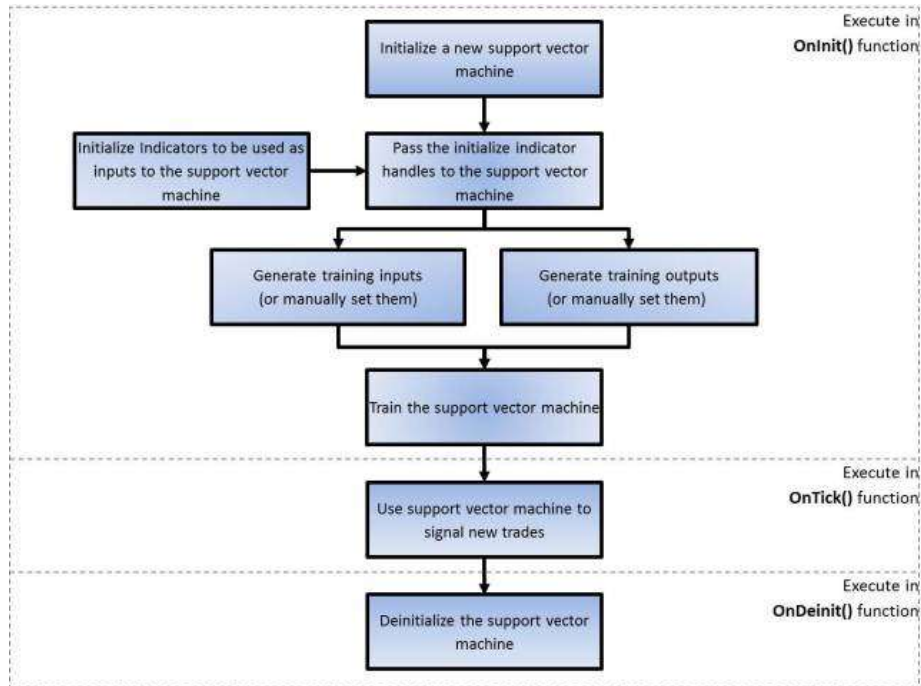


Figura 8. Proceso para la implementación de la herramienta de la máquina de vectores de soporte en un Asesor Experto. Recuperado de: <https://www.mql5.com/es/articles/584>

1.3.10. Decisión Tree:

Son un modelo predictivo basado en una colección de datos que se pueden clasificar y representan una serie de condiciones continuas para resolver un problema específico. Son excelentes herramientas que pueden ayudarlo a tomar la decisión correcta entre muchas posibilidades. Su conformación admite elegir una y otra vez diversas preferencias para sacar provecho a distintas opciones de decisión posibles. Los árboles de decisión suelen ser binarios, o sea, tienen dos opciones, aunque esto no quiere decir que no pueda haber un árbol con tres o inclusive más opciones. Como su nombre indica, se representa en forma de árbol, seguido de varios nodos exactos que indican puntos de decisión, y de ellos surgen ramas que representan distintas probabilidades u opciones. (Jara, Giral, and Martínez 2016).

El árbol de clasificación tiene como finalidad elaborar un modelo que pronostica el valor de la variable objetivo teniendo como base varias

variables de entrada y es una de las técnicas más poderosas dentro de la clasificación supervisada.(Calvo 2016).

1.3.10.1. Pseudocódigo:

Según (Luis 2019) el pseudocódigo para la creación de árboles de decisión es:

1. Corroborar los casos iniciales.
2. A cada atributo
 1. Hallar normalizada la ganancia de información de la partición de a
3. Permitir que a_mejor sea la característica con la ganancia de información normalizada superior.
4. Construir un nodo de decisión que particione a_mejor.
5. Iterar en las sub-listas conseguidas por partición de a_mejor, y adicionar aquellos nodos como los hijos del nodo.

1.3.10.2. Entropía:

La entropía es laa medición empleada en la teoría de la información. El propósito es calcular la cantidad promedio de información comprendida en el flujo de datos, mejor dicho, lo que proporciona cierta información sobre un hecho o datos específicos. (Nestor 2016).

1.3.10.3. Ganancia:

Entendemos por ganancia de información como la medida de relevancia que tiene un atributo dentro de un grupo de datos. Un atributo con una ganancia alta será muy relevante en el grupo de datos. (Calvo 2016).

1.3.10.4. Poda de un árbol de clasificación:

Podar los nodos del árbol de clasificación incluye borrar el sub-árbol anidado en el nodo, convertirlo en hojas y luego asignarle la tipología más general del ejemplo de entrenamiento que se considera en el nodo.(Nestor 2016).

1.4. Formulación del Problema.

¿Qué técnica de predicción será la más eficiente en la predicción de morosidad?

1.5. Justificación e Importancia del estudio.

El presente trabajo de investigación pertenece a la línea de investigación de Infraestructura, Tecnología y Medio Ambiente de la Escuela Profesional de Ingeniería de Sistemas, y esta línea tiene fuerza en la aplicación de la base matemática, principios de técnicas y la teoría de la Ciencia de la Computación en el modelamiento y diseño de sistemas computacionales para alcanzar las necesidades deseadas.

Éste trabajo es también importante porque presenta pertinencia, ya que los resultados de la investigación aportarán un nuevo conocimiento sobre el comportamiento de estas técnicas en el contexto de la morosidad, además los resultados de esta investigación aportarán significativamente al trabajo del doctorado en el que se pretende ayudar, así mismo los resultados del presente trabajo brindará nuevas ideas o inquietudes para realizar nuevos trabajos de investigación posteriores y así ir desarrollando esta línea de investigación.

Tecnológicamente el trabajo se justifica a la viabilidad debido a que los algoritmos que se usaron en este trabajo de investigación son algoritmos que han sido presentados al mundo científico y han sido probados e inclusive se mostraron resultados en trabajos de investigación en contextos parecidos a esta investigación, y además estas técnicas están disponibles para poder trabajarlas. También es viable porque se contó con la base de datos y la autorización del colegio de ingenieros para realizar la investigación. Es viable también porque los recursos económicos que se utilizaran no son muy excesivos y están al alcance del investigador.

1.6. Hipótesis.

La técnica de Adaboost presenta mejores resultados de precisión en la predicción de morosidad.

1.7. Objetivo.

1.7.1. Objetivo General.

Comparar las Técnicas de Predicción de Adaptive Boosting y Decision Tree en la predicción de morosidad.

1.7.2. Objetivos Específicos.

- a) Seleccionar las técnicas de predicción a evaluar.
- b) Procesar la información histórica de base de datos.
- c) Realizar el proceso de ETL para la extracción de datos relevantes.
- d) Aplicar las técnicas Adaptive Boosting y Decision Tree.
- e) Evaluar los resultados obtenidos.

II. MATERIAL Y MÉTODO

2.1. Tipo y Diseño de la Investigación.

2.1.1. Tipo de Investigación.

La presente investigación es Tecnológica Aplicada pertenece a una investigación de tipo Cuantitativa, debido a que se trabajará con valores cuantificables como, niveles de precisión, exactitud. Incluso, permite efectuar preguntas específicas como ¿Qué tan precisas?, ¿Cuál es el nivel de exactitud?, etc.

El tipo de investigación Cuantitativa, según Sampieri, Fernández & Baptista (2006), “Los métodos cuantitativos utilizan la recopilación de datos para tratar suposiciones basadas en mediciones numéricas y análisis estadístico para implantar patrones de comportamientos y experimentar teorías.” En este trabajo, según datos que serán recolectados, veremos si la hipótesis será verdadera o falsa, dependiendo de los resultados que también se obtendrán.

2.1.2. Diseño de la Investigación.

En el presente trabajo de investigación se evaluará la precisión de dos técnicas seleccionadas para la comparación, mediante la implementación en una solución de inteligencia de negocios; por lo que es preciso implantar un tipo de diseño de investigación cuasi experimental.

El Diseño de la Investigación es Cuasi-Experimental, según Sampieri, Fernández & Baptista (1997) menciona que “El diseño cuasi-experimental también dirigen deliberadamente mínimo una variable independiente para conocer su rol y conexión con una o muchas variables dependientes, pero pueden ser diferentes del experimento "real" en términos de confiabilidad o seguridad, y puede tener un impacto en la equivalencia del principio del grupo.

En un diseño cuasiexperimental, los sujetos no se asignan aleatoriamente a grupos ni se emparejan, por el contrario, estos grupos se forman antes del realizar el experimento, y son grupos completos (el motivo de su generación y la forma en que se forman es independiente del experimento o no tiene nada que ver con éste)".

2.2. Población y Muestra.

En el estudio realizado por varios paper se llegó a identificar 99 técnicas de predicción (**Ver Anexo 03**).

La muestra para este trabajo de investigación son dos técnicas: Adaptive Boosting(AdaBoost) y Árbol de Decisión (Decision Tree), y fue determinada por el investigador principal.

2.3. Variables, Operacionalización.

2.3.1. Variable Independiente

Técnicas de predicción.

2.3.2. Variable Dependiente

Predicción de la Morosidad de cuotas sociales Consejo Departamental Lambayeque.

2.3.3. Operacionalización de Variables.

Tabla 2

Operacionalización de variables.

Variables	Dimensiones	Indicadores	Unidad de Medida	Formula	Descripcion	Metodo de Recolección de Datos
Independientes: Técnicas de Predicción	Percentage Split (Porcentaje Dividido)	Accuracy (Precisión)	%	$(VP+VN) / Total$	VP= es la cantidad de positivos que fueron clasificados correctamente como positivos por el modelo. VN= es la cantidad de negativos que fueron clasificados correctamente como negativos por el modelo. FN= es la cantidad de positivos que fueron clasificados incorrectamente como negativos. FP= es la cantidad de negativos que fueron clasificados incorrectamente como positivos.	- Fichas de Observación -Software Automatizados
		Precision (Exactitud)	%	$VP/Total$ clasificados positivos		
		Sensitivity (Sensibilidad)	%	$VP/Total$ Positivos		
		Especificity (Especificidad) AUC (Área Bajo la Curva)	% decimal	$VN/Total$ Negativos $(Sensibilidad+Especificidad) / 2$		
Dependiente: Predicción de Morosidad de cuotas sociales del Colegio de Ingenieros del Perú consejo departamental de Lambayeque.	Tiempo	Tiempo para generar estimación (morosidad)	segundos	TE = TS	TE = Tiempo para generar estimación (morosidad) TS = Tiempo en segundos.	- Fichas de Observación -Software Automatizados

Fuente: Elaboracion Propia

2.4. Técnicas e Instrumentos de recolección de datos, validez y confiabilidad.

2.4.1. Técnicas de recolección de datos.

a) Observación.

Según Campos y Cobarrubias & Lule Martinez, (2012) menciona:

“Observar es la manera más lógica y sistémica de información conocida visualmente y verificable; o sea, captura todo lo que sucede en el mundo real de la forma más clara para analizarlo, describirlo o explicarlo desde un punto de vista científico”. (pag,5).

Mediante esta técnica se busca observar el comportamiento y procesos involucrados para identificar y predecir la morosidad.

b) Análisis Documental.

Según Aenor (1994), la norma UNE 50 -113, menciona que:

“Operación que incluye la comprobación del documento para encontrar sus elementos básicos y las relaciones entre ellos”.

Se utilizó esta técnica al analizar muchos artículos científicos y tesis.

c) Experimento.

Según Bueno (2003) menciona que:

“Los experimentos admiten al investigador intervenir en el proceso para revelar aquellos atributos y relaciones relevantes para su investigación. A lo largo del experimento, las condiciones se pueden alterar para identificar el comportamiento observado por el sujeto bajo este cambio”. (pag. 41-42).

Se realizó la experimentación, al momento de ejecutar los algoritmos con los diferentes datos e indicadores, y se fue modificando algunas condiciones para su ejecución y mejores resultados (iteraciones, cantidad de variables, etc.).

2.4.2. Instrumentos de recolección de Datos.

El instrumento que utilizará esta investigación es una ficha de observación, en donde podemos especificar la relación que se origina entre las variables frente a los eventos.

2.5. Procedimiento de Análisis de Datos.

2.5.1. Procedimientos para la recolección de datos.

- a) Acceder a la base de datos.
- b) Definir Indicadores de Morosidad.
- c) Realizar el proceso de Limpieza de Datos.
- d) Realizar Procedimientos de Extracción, Transformación y Carga de datos a utilizar.
- e) Los datos serán puestos en un almacén de Datos, específicamente en un DataMart.
 - e.1) Aplicar la Técnicas de Predicción AdaBoost.
 - e.2) Aplicar la Técnica de Predicción Decision Tree.
- f) Pasar los Datos arrojados a un Excel para su análisis.
- g) Evaluar los resultados para discernir que técnica obtuvo mejores resultados para este caso de estudio.

2.5.2. Análisis estadísticos e interpretación de los datos

El Análisis estadísticos de los datos se basa principalmente en:

- a) Emplear tablas, para analizar resultados de arrojan dichas técnicas de minería de datos.
- b) Emplear gráficos de estadística, para corroborar los diversos resultados mostrados por las técnicas de minería de datos.

2.6. Criterios éticos.

La ética utilizada en esta investigación está relacionada con algún evento que pueda estar involucrado en este trabajo. Por tanto, se adoptan los siguientes estándares éticos:

a. Criterio de Confiabilidad

Los datos y la información personal requeridos para este trabajo deben obtenerse de una forma legal y profesional para no ocasionar daños a las personas implicadas; como se indica en la ley N ° 29733: "Ley de Protección de Datos Personales", en su sección IV: "Obligaciones del Titular y Administrador de Bases de Datos Personales", en su artículo 28, establece: La recolección de información personal de manera ilegal, desleales o fraudulentas será sancionada por la República Peruana.

b. Confidencialidad:

Asegurar la identidad de la agencia de protección y la identidad del informante involucrado en la investigación.

c. Objetividad:

El análisis de la situación encontrada se basa en criterios técnicos e imparciales.

d. Originalidad:

Se citan las fuentes bibliográficas de la información.

e. Veracidad:

La información mostrada es verdadera, cuidando la confidencialidad.

2.7. Criterios de rigor científico.

2.7.1. Confiabilidad:

Se realizó cálculos estadísticos para encontrar el nivel de consistencia dentro de la herramienta de recopilación de datos.

2.7.2. Validación:

Se retifican los instrumentos usados para la recolección de datos y la validez de resultados empleando Juicio de Expertos.

III. RESULTADOS.

3.1. Resultados en Tablas y Figuras.

En primer lugar se realizó una prueba con todos los meses y se obtuvieron los resultados de la *Tabla 3*.

En donde se puede demostrar que según la Evaluación de Porcentajes (30% para entrenamiento y 70% para test), en general los resultados muestran que, al realizar un promedio, Decision Tree se comportó con mejores resultados que el algoritmo de AdaBoost para este caso de estudio.

Desde ahí se ve que es un algoritmo que ofrece muy buenos resultados, solo ganándole por poco de diferencia al algoritmo de Adaboost.

Tabla 3
Resultados Finales de Ambas Técnicas

	POR PORCENTAJE SPLIT (PORCENTAJE DIVIDIDO)													
	TIEMPO DE CONSTRUCCION DEL MODELO (seg.)		TIEMPO DE PREDICCION (seg.)		Accuracy (Precisión) (%)		Precision (Exactitud) (%)		Sensibilidad (%)		Especificidad (%)		AUC-Curva ROC (decimal)	
	AdaBoos t	Decision Tree	AdaBoos t	Decision Tree	AdaBoos t	Decision Tree	AdaBoos t	Decision Tree	AdaBoos t	Decision Tree	AdaBoos t	Decision Tree	AdaBoos t	Decision Tree
Enero	43.68	2.64	2.48	0.84	86.79	86.82	87.87	89.16	80.65	83.27	90.76	89.12	0.857	0.860
Febrero	49.60	5.94	2.46	0.68	86.30	86.45	87	89.01	79.25	83.31	90.92	88.52	0.851	0.85
Marzo	52.61	1.84	2.50	0.66	87.03	85.79	88.22	85.54	82.06	76.94	90.37	91.74	0.862	0.84
Abril	53.53	2.79	2.59	0.86	86.73	87.15	86.45	88.74	78.71	83.21	92.17	89.83	0.854	0.86
Mayo	71.01	2.82	2.54	0.62	86.94	86.79	86.19	88.97	77.52	83.37	93.19	89.06	0.854	0.86
Junio	92.36	3.05	2.47	0.94	86.30	86.15	85.78	87.78	77.26	81.55	92.39	89.25	0.848	0.85
Julio	65.57	2.31	3.14	0.66	86.67	86.55	85.06	87.55	75.49	80.98	94.21	90.3	0.848	0.86
Agosto	89.56	5.72	2.50	0.64	90.36	91.06	90.88	93.66	83.16	88.89	94.41	92.28	0.887	0.91
Septiembre	70.8	2.89	2.27	0.64	86.27	85.79	84.14	87.32	73.07	80.5	95.06	89.3	0.84	0.85
Octubre	62.17	2.23	2.31	0.64	86.21	86.85	84.42	85.04	74.19	75.32	94.32	94.62	0.842	0.85
Noviembre	57.73	2.62	2.19	0.61	86.06	86.36	83.91	84.33	73.57	74.39	94.63	94.58	0.841	0.84
Diciembre	63.39	3.14	2.4	0.56	84.88	84.97	82.27	82.06	71.76	71.18	94.32	94.89	0.83	0.83
PROMEDIO	64.33	3.17	2.49	0.70	86.71	86.73	86.02	87.43	77.22	80.24	93.06	91.12	0.85	0.86

Fuente: Elaboración propia

Ahora bien, evaluando detalladamente cada indicador de la Variable Independiente tenemos:

Accuracy (precisión):

La Accuracy (Precisión) correspondiente a ambos algoritmos todos los meses, se obtuvo lo siguiente:

Tabla 4
Precisión de ambas técnicas.

Accuracy (Precisión) (%)		
MES	ADABOOST	DECISION TREE
Enero	86.79	86.82
Febrero	86.30	86.45
Marzo	87.03	85.79
Abril	86.73	87.15
Mayo	86.94	86.79
Junio	86.30	86.15
Julio	86.67	86.55
Agosto	90.36	91.06
Septiembre	86.27	85.79
Octubre	86.21	86.85
Noviembre	86.06	86.36
Diciembre	84.88	84.97
PROMEDIO	86.71	86.73

Fuente: Elaboración propia

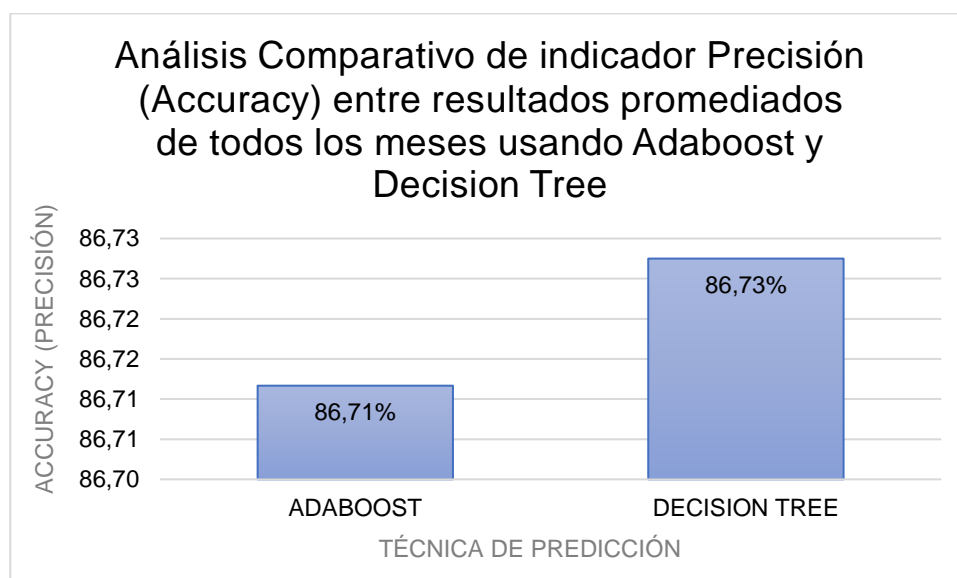


Figura 9. Análisis Comparativo de indicador Precisión (Accuracy) entre resultados promediados de todos los meses usando Adaboost y Decision Tree.

Se observa tanto en la *TABLA 4* como en la *Figura 9* que la técnica de Decision Tree con 86.73% gana por poco a la técnica de Adaboost con 86.71% en cuanto a Precisión (Accuracy), es decir, le gana poco en porcentaje de datos totales predichos correctamente.

Precision (Exactitud):

En cuanto a la Exactitud (Precision), se obtuvieron los resultados siguientes:

Tabla 5

Exactitud de Ambas Técnicas.

Precision (Exactitud) (%)		
MES	AdaBoost	Decision Tree
Enero	87.87	89.16
Febrero	87	89.01
Marzo	88.22	85.54
Abril	86.45	88.74
Mayo	86.19	88.97
Junio	85.78	87.78
Julio	85.06	87.55
Agosto	90.88	93.66
Septiembre	84.14	87.32
Octubre	84.42	85.04
Noviembre	83.91	84.33
Diciembre	82.27	82.06
PROMEDIO	86.02	87.43

Fuente: Elaboración propia

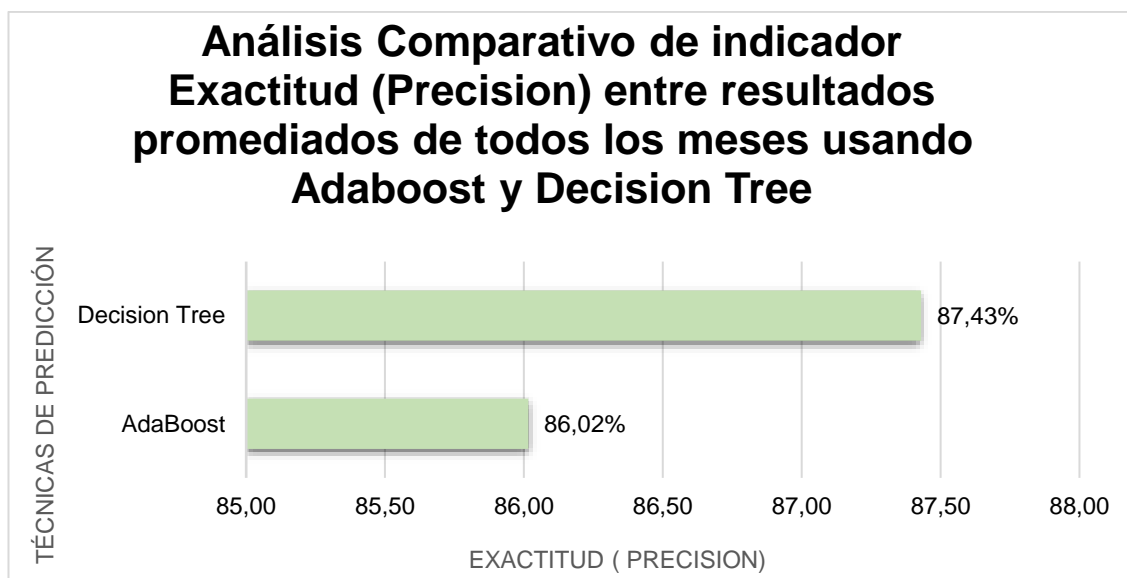


Figura 10. Análisis Comparativo de indicador Exactitud (Precision) entre resultados promediados de todos los meses usando Adaboost y Decision Tree.

Se observa tanto en la TABLA 5 como en la Figura 10 que la técnica de Decision Tree es la que tiene mas Exactitud que la técnica de Adaboost, es decir, le gana en porcentaje de datos positivos predichos correctamente con 87.43% contra 86.02%.

Sensitivity (Sensibilidad):

Al analizar la Sensibilidad, se obtuvieron los resultados siguientes:

Tabla 6

Sensibilidad de Ambas Técnicas.

MES	Sensibilidad (%)	
	AdaBoost	Decision Tree
Enero	80.65	83.27
Febrero	79.25	83.31
Marzo	82.06	76.94
Abril	78.71	83.21
Mayo	77.52	83.37
Junio	77.26	81.55
Julio	75.49	80.98
Agosto	83.16	88.89
Septiembre	73.07	80.5
Octubre	74.19	75.32
Noviembre	73.57	74.39
Diciembre	71.76	71.18
PROMEDIO	77.22	80.24

Fuente: Elaboración propia.

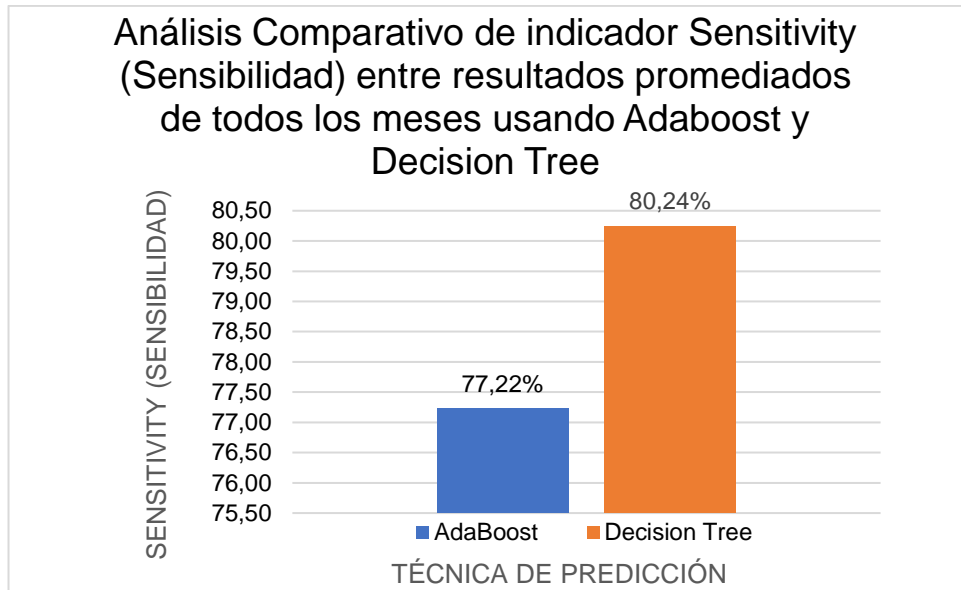


Figura 11. Análisis Comparativo de indicador Sensitivity (Sensibilidad) entre resultados promediados de todos los meses usando Adaboost y Decision Tree.

Se observa tanto en la TABLA 6 como en la Figura 11 que la técnica de Decision Tree es la que tiene mas Sensibilidad que la técnica de Adaboost, es decir, le gana en porcentaje de casos positivos que fueron correctamente identificadas por la técnica. Decision Tree gana con 80.24% contra 77.22 %.

Especificity (Especificidad):

Analizando la Especificidad, se obtuvieron los resultados siguientes:

Tabla 7
Especificidad de Ambas Técnicas.

MES	Especificidad (%)	
	AdaBoost	Decision Tree
Enero	90.76	89.12
Febrero	90.92	88.52
Marzo	90.37	91.74
Abril	92.17	89.83
Mayo	93.19	89.06
Junio	92.39	89.25
Julio	94.21	90.3
Agosto	94.41	92.28
Septiembre	95.06	89.3
Octubre	94.32	94.62
Noviembre	94.63	94.58
Diciembre	94.32	94.89
PROMEDIO	93.06	91.12

Fuente: Elaboración propia.

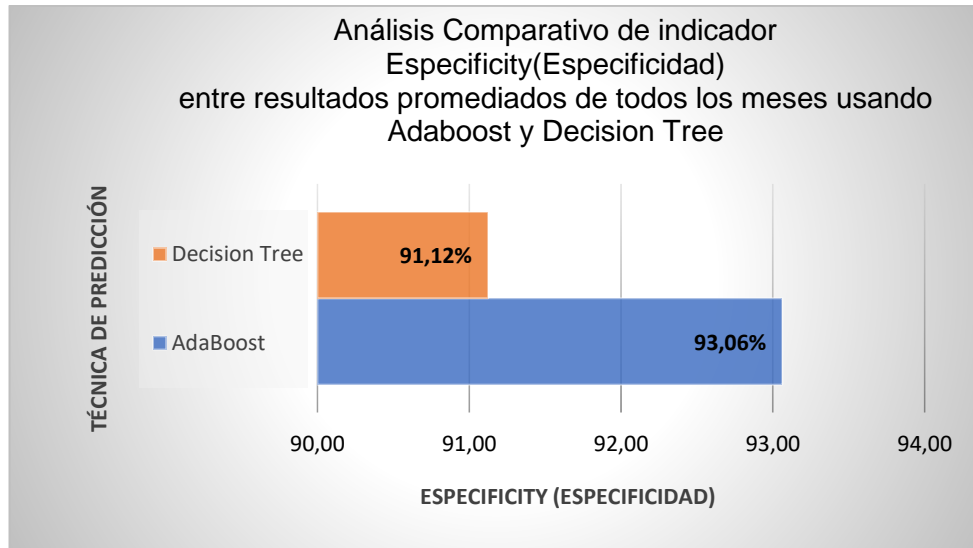


Figura 12. Análisis Comparativo de indicador Especificity (Especificidad) entre resultados promediados de todos los meses usando Adaboost y Decision Tree.
Fuente: Elaboración propia.

Se observa tanto en la *Tabla 7* como en la *Figura 12* que la técnica de Adaboost es la que tiene mas Especificidad que la técnica de Decision Tree, es decir, le gana en porcentaje de casos negativos que la técnica ha predicho correctamente. Adaboost tiene 93.06% contra 91.12 % de Decision Tree en Especificidad.

AUC:

Respecto al Área bajo la Curva (Curva Roc), la herramienta R arroja los resultados presentes en la *Tabla 8* y *Figura 13*:

Tabla 8

AUC de ambas Técnicas.

AUC-Curva ROC (decimal)		
MES	AdaBoost	Decision Tree
Enero	0.857	0.860
Febrero	0.851	0.85
Marzo	0.862	0.84
Abril	0.854	0.86
Mayo	0.854	0.86
Junio	0.848	0.85

Julio	0.848	0.86
Agosto	0.887	0.91
Septiembre	0.84	0.85
Octubre	0.842	0.85
Noviembre	0.841	0.84
Diciembre	0.83	0.83
PROMEDIO	0.85	0.86

Fuente: Elaboración Propia.

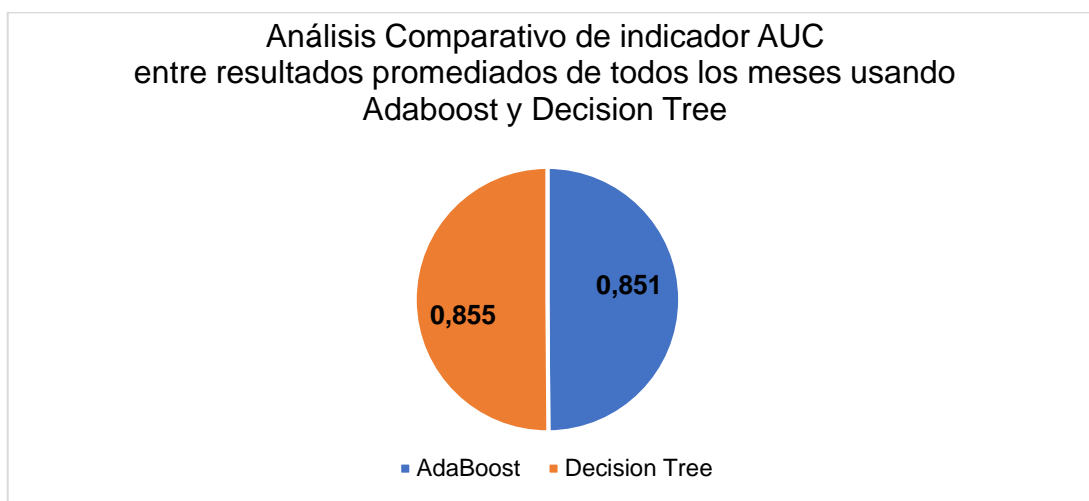


Figura 13. Análisis Comparativo de indicador AUC entre resultados promediados de todos los meses usando Adaboost y Decision Tree.

Fuente: Elaboración propia.

En el AUC, quien está mas cerca a 1 es Decisión Tree, por lo que se reafirma que le gano por realmente poco a Adaboost con 0.855 contra 0.851.

La Herrmienta R brinda diagramas de la Curva Roc - AUC, a continuación, en la *Figura 14* y *Figura 15* se observa como ejemplo las Curvas ROC – AUC generadas por la Herramienta del Mes de Enero, las demás figuras se encuentran en el Anexo 14. **(Ver Anexo 14).**

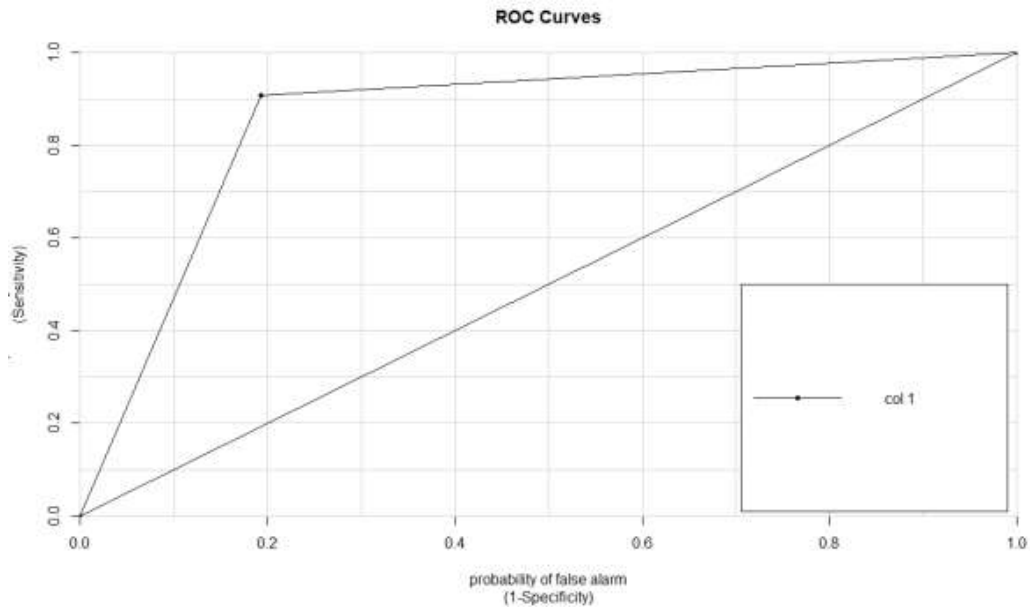


Figura 14. Curva Roc: Respecto a Adaboost. Fuente: Elaboración Propia.

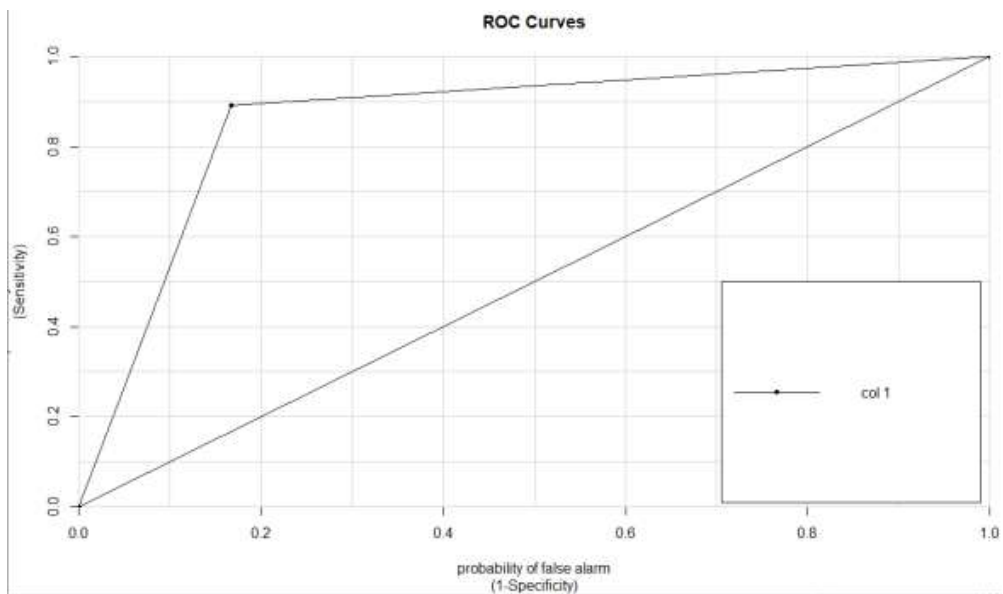


Figura 15. Curva Roc: Respecto a Decision Tree. Fuente: Elaboración Propia.

Ahora bien, evaluando detalladamente el indicador de la Variable Dependiente tenemos:

Tiempo de Estimación de Morosidad:

Para calcular el tiempo que tarde en predecir cada algoritmo se utilizó la función *proc.time()* de R, este tiempo será diferente en cada ordenador por lo que depende de las características que se tenga en la computadora, este

estudio se realizó con una laptop Lenovo Z40-70 con procesador Intel Core i5, con 12 gb de memoria Ram y con 500 gb de almacenamiento.

El Tiempo en el que tarde la herramienta para predecir morosidad (no para entrenar) en un conjunto de datos que se le otorge a la técnica para que prediga (testeo) brindaron los siguientes resultados:

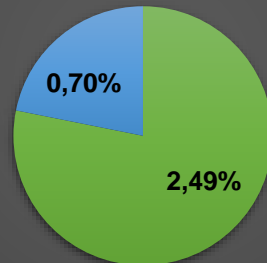
Tabla 9

Tiempo de Estimación de Morosidad de Ambas Técnicas por cada mes.

TIEMPO DE ESTIMACIÓN PREDICCIÓN DE MOROSIDAD		
(seg.)		
MES	AdaBoost	Decision Tree
Enero	2.48	0.84
Febrero	2.46	0.68
Marzo	2.50	0.66
Abril	2.59	0.86
Mayo	2.54	0.62
Junio	2.47	0.94
Julio	3.14	0.66
Agosto	2.50	0.64
Septiembre	2.27	0.64
Octubre	2.31	0.64
Noviembre	2.19	0.61
Diciembre	2.4	0.56
PROMEDIO	2.49	0.70

Fuente: Elaboración Propia.

Análisis Comparativo del tiempo de estimación de predicción de Morosidad (segundos) entre resultados promediados de todos los meses usando Adaboost y Decision Tree



■ Adaboost ■ Decision Tree

Figura 16. Análisis Comparativo del tiempo de estimación de predicción de Morosidad (segundos) entre resultados promediados de todos los meses usando Adaboost y Decision Tree. Fuente: Elaboración Propia.

En la *Tabla 9* y en la *Figura 16*, se observa la considerable diferencia en lo que tarda en dar la predicción Adaboost con 2.49 segundos a comparación de Decision Tree con 0.70 segundos.

Si bien los Resultados mostrados anteriormente son muy parecidos, el tiempo y recursos que consumen ambas técnicas es considerable. También en este sentido, Decision Tree es superior a Adaboost.

La Herramienta R grafica el árbol de Decision generado por cada mes usando la técnica de Decision Tree (C5.0), un ejemplo de ello se muestra en la *Figura 17* del mes de Septiembre:

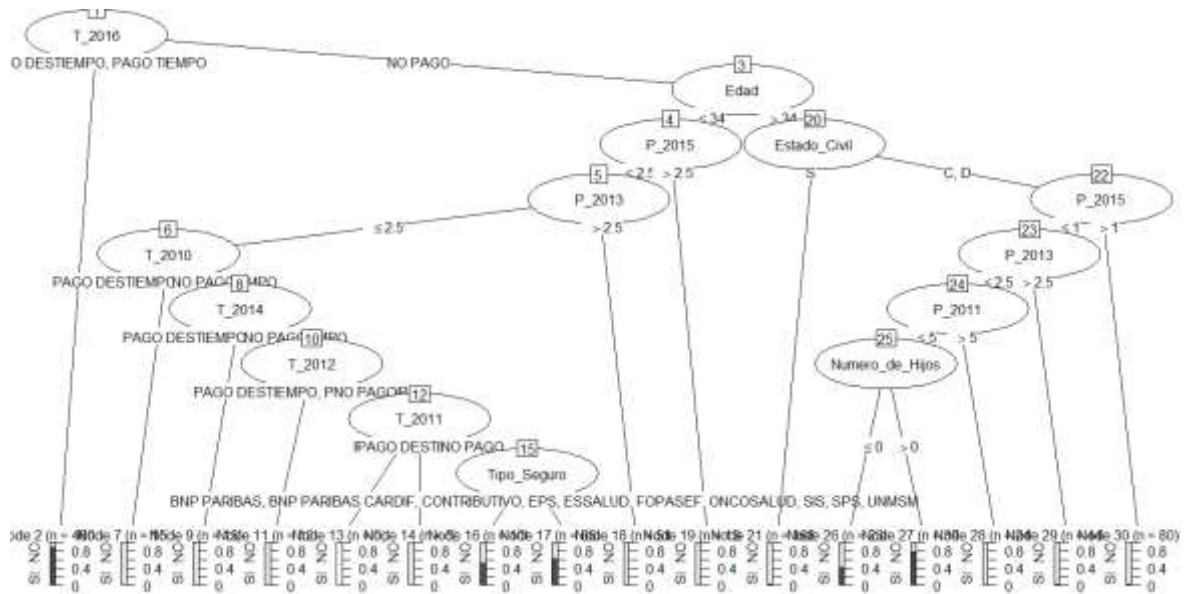


Figura 17. *Árbol de decisión generado en el mes de septiembre por C5.0(Decision Tree).*

Todos los Árboles Completos se pueden encontrar en el Anexo 13. **(Ver Anexo 13).**

3.2. Discusión de Resultados.

Como se demostró en los resultados, en casi la totalidad de evaluaciones, y más aún, promediando la precisión, Decisión Tree gana con 86.73% contra 86.71% de Adaboost. Decision tree también fue muy veloz al realizar la predicción, ya que, en promedio, lo realizó en tan solo 0.70 segundos, contra 2.49 segundos de Adaboost.

A diferencia de los Estudios, por ejemplo, de (Alaka et al. 2018), donde Decisión Tree esta entre los 3 mejores y en el estudio de (Jones, Johnstone, and Wilson 2015) Adaboost esta también entre los mejores.

Así mismo hay casos y/o estudios que no corroboran los resultados que se obtuvieron en este trabajo de investigación, por ejemplo (Heo and Yang 2014), concluye según sus pruebas realizadas, que en su conjunto de datos, la técnica que mejores resultados presenta es Adaboost, o el de (Alfaro and Elizondo 2008), que concluye lo mismo.

Por lo qué, entonces, se reafirma lo que indico (Rathor 2017), donde indica que ninguna técnica es mejor que los demás en un ámbito general, sino, que el mejor va a depender del caso en el que se esté estudiando y sus circunstancias por lo que se realizó un estudio para determinar los mejores técnicas predictivas.

Por último, la hipótesis que se mencionó al principio del presente trabajo de investigación es errónea, ya que, habiendo realizados las pruebas, Decisión Tree es la mejor técnica para realizar predicción en el presente caso de estudio.

3.3. Aporte Práctico.

a) Introducción.

Como se mencionó en el Capítulo I de este estudio, en la parte de la situación problemática, la toma de decisiones en las empresas es demasiado importante para no poner en riesgo la viabilidad de ésta. Para lograr ello, existe la Inteligencia de Negocios o BI (Business Intelligence), grupo de herramientas que le brindan una interpretación a toda la información reunida, de tal forma que es mucho más fácil de entender e interpretar esa información y así poder tomar decisiones.

También se mencionó que, si bien la Inteligencia de Negocios ayuda a tomar decisiones, ayudaría mucho más usar técnicas de predicción, ya que juntos brindan la opción de detectar amenazas y oportunidades para anticiparse a la toma de decisiones.

Este trabajo de Investigación es parte de un trabajo de investigación más grade, de un doctorado, cuyo título es “Sistema de Inteligencia de Negocios Sustentado en un Modelo Estratégico de Análisis Predictivo para mejorar la Toma de Decisiones en el Colegio de Ingenieros del Perú – CD Lambayeque” y está a cargo del Ing. Denny John Fuentes Adrianzén.

En este trabajo se menciona que según lo analizado por varios autores se evidencia que se siguen utilizando modelos descriptivos y de predicciones convencionales o tradicionales, y no existen suficientes referentes teóricos y prácticos relacionados con la Inteligencia de Negocios sustentados en un modelo estratégico de análisis predictivo que ayude en forma eficiente a la hora de tomar decisiones.

Por lo que se propone como objetivo el aplicar dos técnicas de predicción para poder predecir la morosidad, saber qué colegiados no pagaran en un determinado tiempo (ejemplo: colegiados que no pagarán en un determinado mes), e identificar cuál de estas técnicas es más eficiente y así ayude a la buena toma de decisiones del Colegio de Ingenieros del Perú – Consejo Departamental de Lambayeque.

b) Desarrollo.

El estudio de los diferentes artículos para poder armar el top fue realizado por un grupo de investigadores que usó la misma data para realizar la predicción de morosidad, pero con diferentes técnicas, el cual a partir de ahora en este trabajo se les llamará: El Grupo de Investigación. El cuál esta conformado por los siguientes investigadores: Alberto Guevara Barreto, Juan Carrasco Manay, Víctor Manuel Martínez Panta, Jefferson Huamán Bernilla, Flor Diaz Macalupu, Darwin Alain Vázquez Cercado, Elmer Anthony Monja Sandoval y Maximo Gabriel Silva Parraguez.

3.3.1. Seleccionar las técnicas de predicción a evaluar.

Como se mencionó, existe bastantes técnicas de predicción que juntamente con la Inteligencia de Negocios, ayuda de forma considerable para la toma de decisiones proyectándose a futuro. Para poder identificar la mayoría de estas técnicas de machine learning, se realizó una revisión de diversas fuentes (sean o no aplicados a un ámbito parecido a este caso de estudio, la idea es primero identificar la mayoría de las técnicas).

De todo lo revisado, las publicaciones mas completas y específicas para dar a conocer la mayoría de técnicas de machine learning para realizar predicción es la investigación de (Portugal et al. 2017) y la de (Brownlee 2013). En la primera investigación describe un total de 46 técnicas clasificándolas según la literatura que revisaron (**Ver Anexo 04**), en la segunda investigación, el autor realiza una descripción de varios algoritmos clasificándolos en 11 grandes grupos, dicha clasificación realizada es muy detallada (**Ver Anexo 05**).

Al finalizar el análisis, se obtuvo una lista de 99 técnicas como población total (**Ver Anexo 06**). Ahora bien, el problema es identificar cual de todas estas técnicas son las mejores para poder emplearlas en el ámbito de predicción de morosidad, en esta investigación se determinó si ciertos colegiados serán morosos o no; teniendo en cuenta esto, se eligió el algoritmo que más se adecue al ámbito de estudio.

Top de mejores técnicas de predicción:

Entonces, debido a la cantidad de técnicas que existen, se ha elaborado una revisión bibliográfica para elegir cuales son las más representativas y también las más apropiadas para el presente caso de estudio de predicción de morosidad.

El proceso que se siguió para poder elegir las técnicas que se usarán fue el de leer artículos científicos e ir armando un top conforme los resultados que se reflejaban en esos estudios realizados; cabe recalcar que estos estudios son parecidos al de la presente investigación. Los artículos mostrados en el Anexo 7 (**Ver Anexo 7**) se consideraron los más importantes para poder realizar el Top por la complejidad de su estudio realizados y por sus resultados.

Al observar los resultados, las técnicas más usadas y que mejores resultados han obtenido en estudios parecidos a la presente investigación, son los siguientes:

- 1) SUPPORT VECTOR MACHINES
- 2) ARTIFICIAL NEURAL NETWORK
- 3) DECISION TREE
- 4) RANDOM FOREST
- 5) ADABOOST
- 6) k-NEAREST NEIGHBOUR
- 7) GENETIC ALGORITHMS
- 8) ROUGH SETS
- 9) NAIVE BAYES
- 10) BAGGING

Para el caso del presente trabajo, el investigador principal asignó a conveniencia las técnicas de Adaptive Boosting (AdaBoost) y Árbol de Decisiones (Decisión Tree) para conocer su eficiencia dentro de ámbito de predicción (prediciendo morosidad).

Elección del tipo de técnica más adecuado de Árbol de Decisión y de AdaBoost.

Dentro de Árbol de Decisión y AdaBoost existen diversas variantes propias en cada una de estas técnicas, para saber cuáles se usarán en esta investigación, se procedió a realizar una revisión bibliográfica y se eligieron por sus características y por ser las más apropiadas al presente caso de estudio.

i. Árbol de Decisiones.

Según la literatura revisada, se identificaron varios tipos de árboles de decisión:

Tabla 10

Características de las principales técnicas de árboles de decisión

Algoritmo	Variables Predictoras	Tipo de división	Criterio de División	Casos <i>missing</i>	Método de Poda
CART (1984)	Continuas / Discretas	Binaria	Impureza (Gini index)	SI	Post-
ID3 (1979)	Discretas	n-aria	Ganancia de Información (Entropía)	NO	NO
C4.5 (1993)	Continuas / Discretas	Binaria / n-aria	Gain ratio (Entropía)	SI	Pre-/Post
J4.8 (Weka)	Continuas / Discretas	Binaria / n-aria	Gain ratio (Entropía)	SI	Pre-/Post
C5.0	Continuas / Discretas	Binaria / n-aria	Gain ratio (Entropía)	SI	Pre-/Post
CHAID (1975)	Discretas	n-aria	X ²	SI	Pre- (nivel de significancia)

Fuente: Pérez, J.M. (2006). Tesis Doctoral. Universidad del País Vasco

Se tomó como candidatos a algoritmos más recientes, estos son el C4.5 y el C5.0, y se procedió a su comparativa. Según (Rodrigo 2017) indica las mejoras que posee C5.0 respecto a su predecesor C4.5:

- La medida de pureza empleada para las divisiones del árbol es la entropía.
- Los árboles se pueden convertir en modelos basados en reglas.
- Emplea un algoritmo de boosting más próximo a AdaBoost que a Gradient Boosting.
- Por defecto, el algoritmo de boosting se detiene si la incorporación de nuevos modelos no aporta un mínimo de mejora.
- Incorpora una estrategia para la selección de predictores (Winnowing) previo ajuste del modelo.
- Permite asignar diferente peso a cada tipo de error.

Por lo expuesto en lo anterior, se utilizará el algoritmo C5.0 debido a las razones mencionadas y porque se adecua a la finalidad de esta investigación.

ii. AdaBoost.

Según la literatura disponible revisada, existen un sin número de algoritmos de Boosting, variantes de Adaboost, en esta investigación solo se revisará el algoritmo Adaptive Boosting (Adaboost) propiamente dicho, pero éste a su vez comprende diversidad de algoritmos (MATHWORKS 2018) los más representativos son:

- AdaBoost.M1 y AdaBoost.M2: técnicas/algoritmos originales para la clasificación de binarios y multiclases respectivamente.

Teniendo en cuenta que la finalidad del modelo de predicción del algoritmo es determinar si ciertos colegiados pagarán o no en determinado mes (binario), la técnica que más se adecua a esta investigación (por no ser multiclase) es el algoritmo de Adaboost.M1.

3.3.2. Procesar información histórica de base de datos.

Para poder llevar a cabo la investigación y aplicar las técnicas para la predicción, se usó la base de datos perteneciente al Colegio de Ingenieros CD Lambayeque.

La recopilación histórica de la base de datos fue realizada por el Grupo de Investigación.

El primer paso para conseguir la información que se utilizará para este estudio es definir variables de predicción. Estas variables deben contener información del caso que se está enfrentando, por lo que para poder hallar estas variables se realizó un análisis de estudios parecidos al presente trabajo, encontrándose las siguientes:

En la investigación de (Yeh and Lien 2009), revisaron otros estudios y al final se usó las siguientes 23 variables como variables para predicción:

X1.- Monto del crédito otorgado (dólar NT): incluye tanto el crédito individual al consumidor como su crédito familiar (suplementario).

X2.- Género (1 = para representar masculino y 2 = para representar femenino).

X3.- Educación (1 = para escuela de posgrado, 2 = para representar universidad, 3 = para escuela secundaria y 4 = otros).

X4.- Estado civil (1 = representa casado, 2 representa soltero y 3 = para otros).

X5.- Edad (indicada en años).

X6 al X11.- Historial de pagos anteriores. Rastreamos los registros de pagos mensuales anteriores (de abril a septiembre de 2005) de la siguiente manera: X6 = el pago realizado en septiembre de 2005; X7 = el pago realizado en agosto de 2005; . . .; X11 = el pago realizado en abril de 2005. La forma en que se mide el estado de reembolso es: -1 = pagar debidamente; 1 = incumplimiento de pago de un mes; 2 = incumplimiento de pago de dos meses; . . .; 9 = incumplimiento de pago de nueve meses o más.

X12 al X17.- Monto del estado de cuenta (dólar NT). X12 = total del estado de cuenta de septiembre de 2005; X13 = total del estado de cuenta de agosto de 2005; . . .; X17 = total del estado de cuenta de abril de 2005.

X18 al X23.- Importe del pago previo (dólar NT). Donde X18 = total pagado en el mes de septiembre de 2005; X19 = total pagado en el mes de agosto de 2005; . . .; X23 = total pagado en el mes de abril de 2005.

En el trabajo de Investigación de (Trujillo 2017) no menciona el total de variables de predicción a utilizar, pero si las más representativas:

Antigüedad_cliente: La mayor antigüedad como cliente de todos los intervinientes.

Antigüedad_empleo: La mayor de la antigüedad en el empleo de todos los intervinientes.

Edad_hijo_menor: La menor de las edades de los hijos de todos los titulares.

Edad: La mayor de las edades de todos los titulares.

Endeudamiento: Ratio de las deudas totales de todos los intervinientes entre el patrimonio de todos los intervinientes.

Estado_civil: Estado Civil del primer titular.

Fecha: Fecha de apertura del contrato.

Importe_inicial: Importe por el que se concede la operación.

Importe_patrimonio: Suma del importe total del patrimonio, agrupado para todos los titulares y todos los avalistas.

Ingresos: Ingresos anuales de todos los intervinientes de la operación, entendidos como la suma de los Ingresos del trabajo y de los Ingresos de otras rentas de la declaración de bienes.

LTV: Ratio del límite de la operación sobre la tasación de las garantías.

Numero_deudas: Número total de deudas informadas en la declaración de bienes del titular. **Numero_aval:** Número de avalistas.

Numero_tit: Número de titulares.

Numero_hijos: Número máximo de hijos de todos los titulares.

Deudas: Suma del importe total de las deudas declaradas en la declaración de bienes del titular.

Tasa_esfuerzo: Ratio de la cuota anual de la operación sobre los ingresos brutos anuales de todos los titulares.

Total_tas: Suma del valor de las últimas tasaciones de los bienes asociados al contrato.

IND_MORA: Indica si la operación ha registrado algún evento de default material a lo largo de la vida (variable que queremos predecir al final del ejercicio).

Identificador: Número que identifica al contrato.

Destino: Código de destino de la financiación del contrato categorizado. El código puede hacer alusión a: Nuevos inmuebles, Refinanciaciones u Otros.

Incidencias_en_sistema: Indica si han existido incidencias en el sistema en los últimos 12 meses anteriores a la formalización para alguno de los intervinientes.

Contratato_indefinido: Indica si alguno de los dos primeros titulares tiene contrato indefinido.

Incumplimiento: Indica si han existido incumplimientos medios superiores a 20 días en los 12 meses previos a la formalización para alguno de los titulares.

Nomina_domic: Indica si alguno de los intervinientes tiene un abono por concepto de nómina domiciliada en alguna cuenta a la vista.

Profesión: Profesión del primer titular categorizada.

Tomando en cuenta el análisis realizado de estos estudios, se determinó que en el caso de la presente investigación se tendrán en cuenta estas variables de predicción:

- a) **Género:** (Se definió por lo siguiente: M = masculino, F = Femenino).
- b) **Estado Civil.** (Se definió de la siguiente manera: S = Soltero, C =Casado, D =Otros/Se desconoce).
- c) **Edad:** La edad de cada colegiado (representada en Años).
- d) **Número de Hijos:** Cantidad de Hijos que posee el colegiado (representada en Números).
- e) **Historial de Pagos:** Es el historial de pagos pasados, se tomaron datos desde el año 2000 hasta el año 2017, ya que cuenta con mayor número de registro de pagos. Tomando como modelo la investigación de (Yeh and Lien 2009), la información será dividida en meses (En hojas de cálculo), siendo así: Hoja1(Enero): Años establecidos (2000-2017) que representaría el pago que realizó cada colegiado en Enero de esos años.....Hoja12(Diciembre): Años establecidos (2000-2017) que representaría el pago que realizó cada colegiado en Diciembre de esos años.

f) Historial de Puntualidad para cada pago: Representa la cantidad de días atrasados al realizar un pago determinado. Se obtendrá restando el día exacto en que se realizó el pago de la cuota mensual menos la fecha que vence el pago. Y se representa de la siguiente manera: Número negativo = Cantidad de días que pago antes de la fecha de vencimiento, Número Positivo = Cantidad de días que se demoró en realizar el pago, Número Cero (0) = Representa que no se cuenta registro de ningún pago.

g) Profesión/Especialidad: Representa la profesión exacta de cada colegiado.

Según (Ferrero, 2018) afirma que, si se tienen variables predictoras que es de interés, debemos asegurarnos de incluirlos en el modelo final.

Es así que estas dos variables a continuación, se tuvieron en cuenta debido a su importancia de cada una de ellas:

h) Categoría: Representa la categoría a la que pertenece cada colegiado. Se representó de la siguiente manera: (Vitalicio, Ordinario). Se considero muy importante porque, cada categoría de colegiado no paga un mismo monto de cuota mensual. Los “vitalicios” pagan, (hasta la fecha de esta investigación) s/. 2.50 y los “ordinarios” pagan s/. 20.00. Ésto puede influir en la morosidad habiendo dos perspectivas:

- Por ser vitalicio y pagar tan poco, puede que no sea moroso y este al día en sus pagos, ó,
- Por ser vitalicio y pagar tan poco, puede no pagar a tiempo por tan poca cantidad, y en determinado tiempo (mes de pago), realizar todos los pagos de su cuota mensual y ponerse al día, habiendo sido moroso todo ese tiempo.

i) Tipo de Seguro: Representa el tipo de seguro que posee el colegiado. Al no poder conseguir información de ingresos mensuales de cada colegiado, se tuvo en cuenta esta variable predictora, ya que existen seguros caros y otros no tanto, por lo que da una idea de la capacidad económica de cada uno de los colegiados, es por ello que se consideró importante.

Teniendo claro entonces las variables que se necesita usar, se procedió a analizar la data histórica de pagos de los colegiados con el fin de encontrar la información requerida.

Esta data histórica se encontraba en dos orígenes de datos, el primer origen de datos estaba en el motor de Sql Server y contaba con más de 12 gigabyte de información. Y el segundo origen de datos útil estaba en una hoja de cálculo (Excel) que se nos fue entregado al Grupo de Investigación.

En cuanto al primero origen de datos (base de datos transaccional original en SQL) se contaba con 4 diagramas ya generados, los cuales se muestran (con el nombre tal cual estaba guardado) en la *Figura 18*, *Figura 19*, *Figura 20*, *Figura 21*:

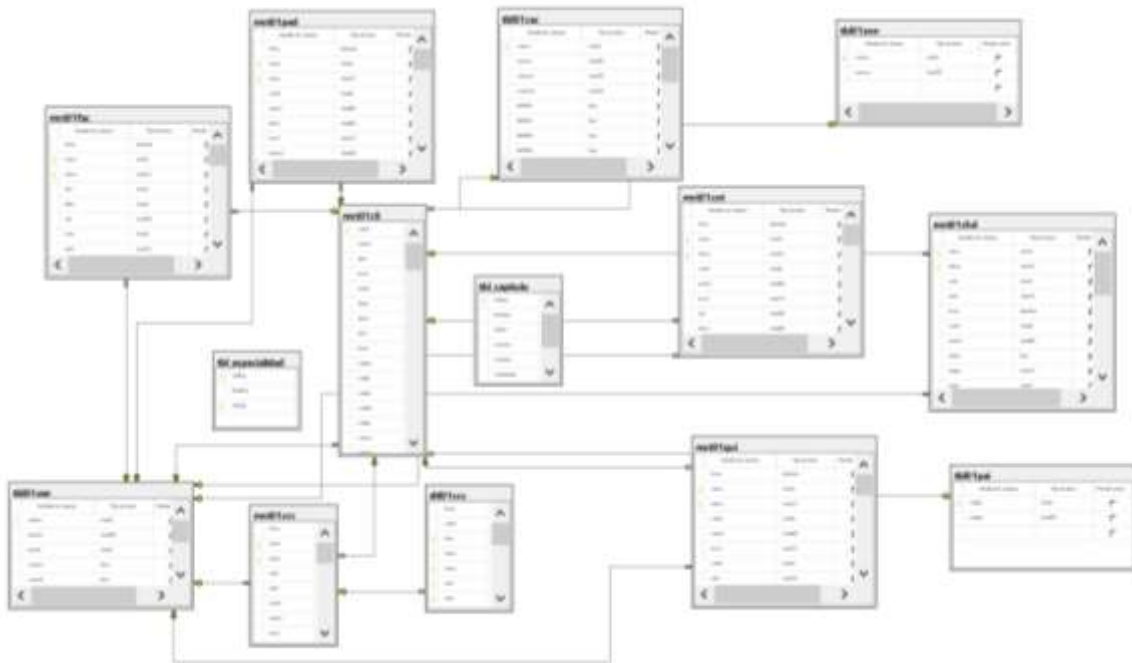


Figura 18. Diagrama con Nombre: CIP. Elaborado por el Grupo de Investigación

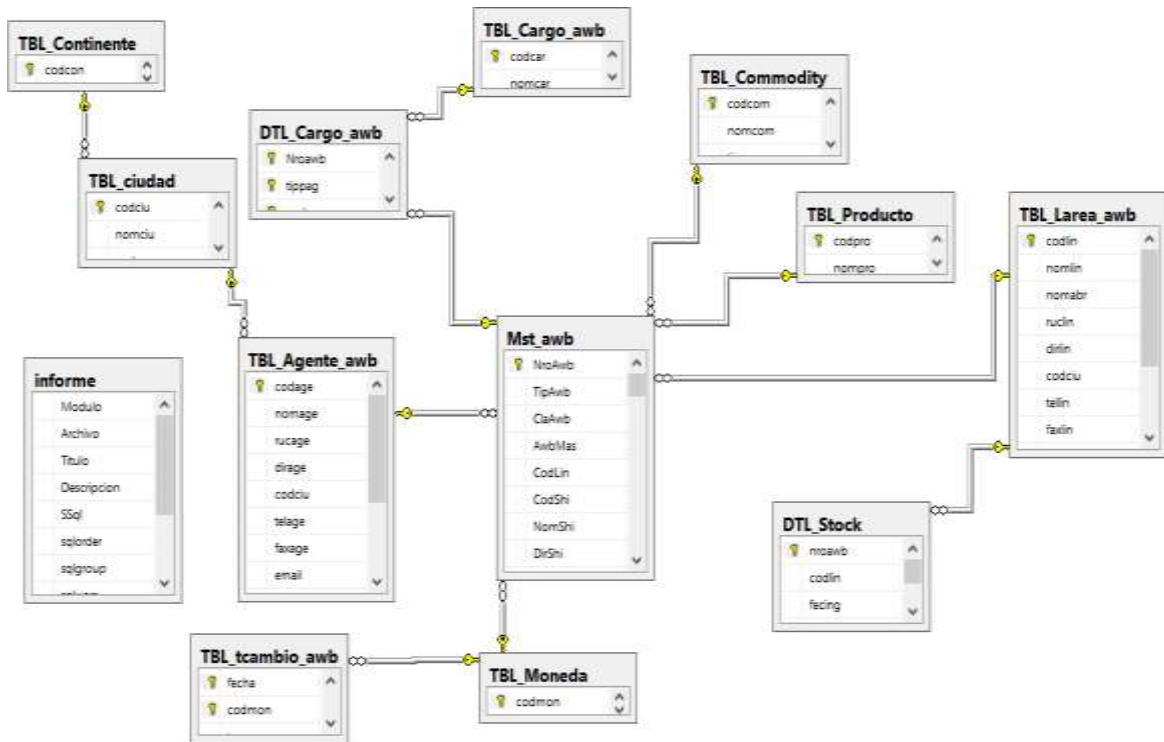


Figura 19. Diagrama con Nombre: DIAGRAM_AWB. Elaborado por el Grupo de Investigación.

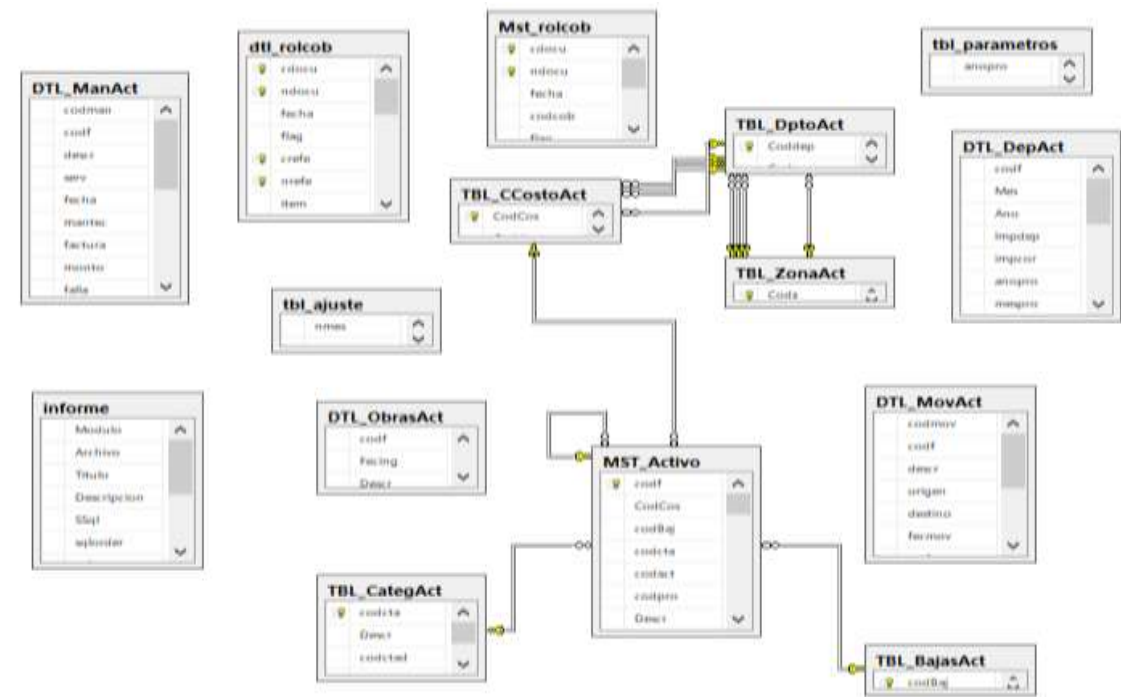


Figura 20. Diagrama con Nombre: DIAGRAM_ACTIVOSfIJOS. Elaborado por el Grupo de Investigación.

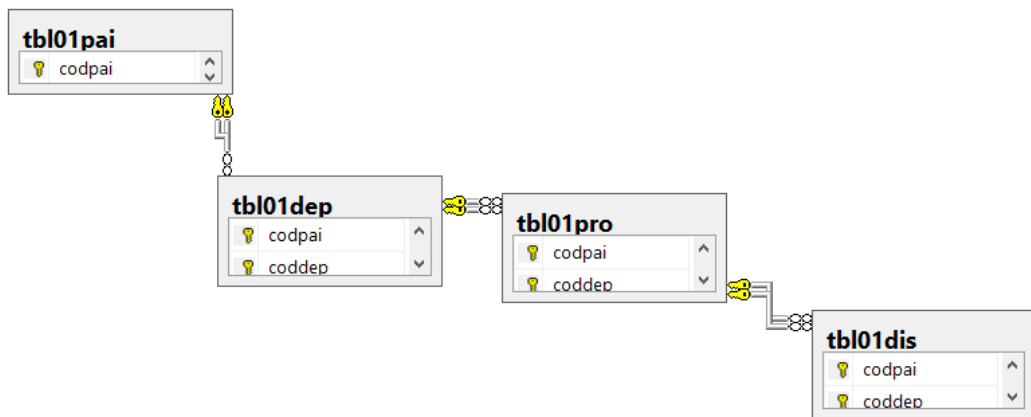


Figura 21. Diagrama con Nombre: *DIAGRAM_UBIGEO*. Elaborado por el Grupo de Investigación.

Cabe recalcar que no están presentes todas las tablas en esos diagramas, y casi todas las demás tablas tenían como nombre formatos parecidos a los de la *Figura 22*:

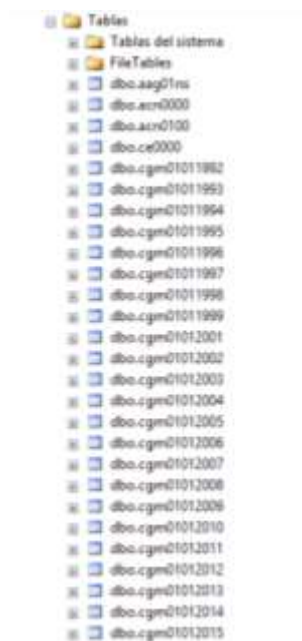


Figura 22. Nombres de algunas tablas con la que contaba la base de datos original.

Elaborado por el grupo de Investigación.

En cuanto al segundo origen de datos (hoja de cálculo), muestra información muy precisa de cada colegiado, y cuenta con los siguientes datos mostrados en la *Figura 23* (no se muestran los registros por motivos de confidencialidad):

CIP	Nombre	DNI	Dirección	Fec_Nacimiento	Capitulo	Especialidad	Fec_Ingreso	Habilidad	Deuda_Total	Categoría	Tipo Seguro
-----	--------	-----	-----------	----------------	----------	--------------	-------------	-----------	-------------	-----------	-------------

Figura 23. Columnas con la que cuenta la hoja de cálculo mencionado.
Elaborado por el Grupo de investigación.

Los campos contenidos como columnas en esta hoja de cálculos son:

CIP: es el número único que identifica a una persona como colegiado.

Nombre: Indica el nombre del Colegiado.

Dni: Documento Nacional de Identidad del Colegiado.

Dirección: Indica la dirección del Colegiado.

Fec_Nacimiento: Respresenta la Fecha de Nacimiento del Colegiado.

Capítulo: Indica al capítulo al que pertenece cada colegiado.

Especialidad: Muestra la Especialidad de Cada Colegiado.

Fec_Ingreso: Fecha de Colegiatura de cada Colegiado.

Habilidad: Indica el estado de Habilidad de Cada Colegiado.

Deuda_Total: Indica la deuda total (no detallada) que tiene el colegiado.

Tipo_Seguro: Indica con que seguro cuenta el colegiado.

Como se demostró, la información estuvo muy desorganizada, inconsistente y poco entendible, por lo que se procedió a realizar un análisis.

Comenzando a analizar la base de datos dada en Sql Server, se procedió a realizar una matriz (**ver Anexo 08**) donde se detalla cada una de las más de 1000 tablas, en las cuales se empezó a diferenciar, las tablas que contenían información relevante, como la que no tenía relevancia alguna con las variables de predicción que se necesitan para la presente investigación.

Una vez descartadas las tablas que no contenían información relevante, se hizo una segunda matriz donde se detalla cada columna de cada tabla que si contiene información clara y con datos relevantes (**ver Anexo 09**).

Después de realizar mencionado análisis, se obtuvo que la base de datos del colegio de Ingenieros CD Lambayeque está conformado por diferentes registros, tanto de colegiados como sus pagos realizados, desde 1994 en adelante.

Algunas variables más representativas que se encontraron y generaron son:

Variables Numéricas:

NroCIP: Indica el número CIP de cada colegiado, este campo es único.

NroDNI: Documento Nacional de Identidad de Cada Colegiado.

TelCli; Indica el Teléfono de cada colegiado.

Edad: Edad del Colegiado.

Nro_Hijos: Indica el Número de Hijos de cada uno de los colegiados.

Mes_Pago: Describe el mes al que se realizó cada pago (se extrajo de la fecha de apertura de cada pago).

Monto_Programado: Es el monto que se le programa a cada colegiado.

Monto_Pagado: Es el monto que se canceló respecto al monto programado.

Saldo: Es el monto que queda una vez el colegiado realiza el pago.

Días: Resultado en números de realizar la resta de días de la fecha en la que se pagó y la fecha de vencimiento respecto al pago que realizó el Colegiado.

Variables Categóricas:

NomCli: describe el nombre de cada Colegiado.

Genero: Indica el género del colegiado, ya sea masculino o femenino.

DirCli: Indica la dirección de cada Colegiado.

FecNac: Indica la fecha en que nació el Colegiado.

FecCol: Indica la fecha de Colegiatura de cada colegiado.

EstCivil: Estado Civil del Colegiado

NomCat: Categoría de Colegiado que pertenece.

Tipo_Seguro: Indica el tipo de seguro con el que cuenta actualmente.

Estado_habilidad: indica si esta hábil o no.

NomEsp: Indica el nombre de especialidad.

NomCap: Indica el Nombre de cada Capítulo.

CodVen: Indica el código del cajero que atendió dicho pago.

FechaApertura: Fecha en la que se apertura el pago para determinado mes.

FechaVencimiento: Es la fecha tope en el que se puede realizar dicho pago tomando en cuenta la Fecha de Apertura.

FechaPago: Es la fecha en la que el colegiado realizo su cuota mensual.

Familiar: Nombre de la persona, la cual es familiar de un colegiado.

NomPar: Nombre de Parentesco que tiene un familiar con un colegiado en específico.

Se procedió, una vez identificado estas variables, a desarrollar un modelo de Entidad-Relación para organizar la data que se obtuvo, quedando al final como se aprecia en la *Figura 24*:

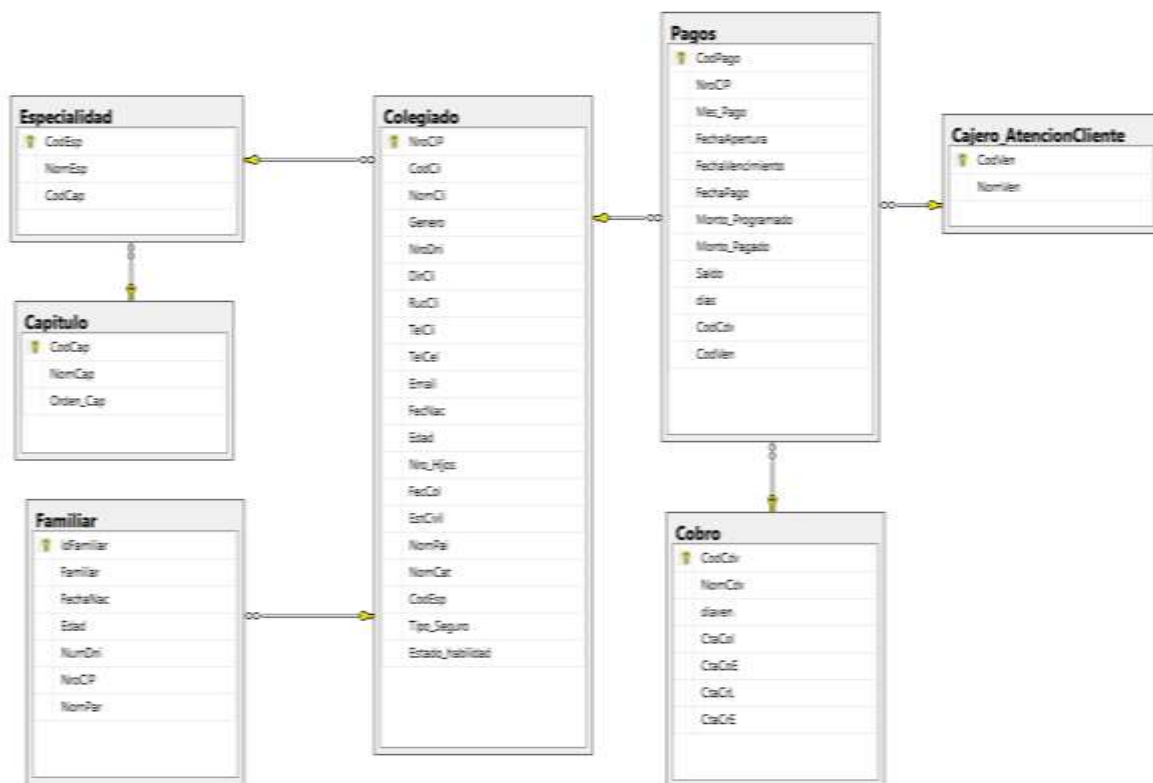


Figura 24. Diagrama Entidad – Relación (Con Datos Ordenados) con nombre: CIP.

A continuación, se describe a que hace alusión cada tabla:

Tabla 11

Descripción de tablas del modelo Entidad – Relación.

Tabla	Descripción
Especialidad	Se anotan las Especialidades pertenecientes a cada Capítulo.
Capitulo	Se guardan los nombres de los Capítulos del CIP.
Familiar	Se guardan los familiares del Colegiado al CIP.
Cobro	Tipo de cobro que realiza el colegiado.
Pagos	Cantidad de pago del Colegiado.
Cajero_AtencionCliente	Identificación de Cajeros que registran los pagos en la Tesorería del CIP.
Colegiado	Asociado o Colegiado del CIP

Elaborado por el Grupo de Investigación.

Se realizó también el diccionario de datos para cada tabla del modelo Entidad Relación elaborado por el Grupo de Investigación, el cual se profundiza a detalle en el Anexo 10. **(Ver Anexo 10).**

a) Analizando y Adquiriendo datos importantes.

Una vez teniendo la base de datos CIP, y teniendo en cuenta también el análisis realizado en el **Anexo 08** y **Anexo 09**, se procedió a pasar los datos de la base de datos de Origen (Cuyo nombre es: Colegio_Ingenieros) y de la Hoja de Cálculo de Excel, a la base de datos destino (Cuyo nombre es: CIP), tal como se detalla en la *Figura 25*:

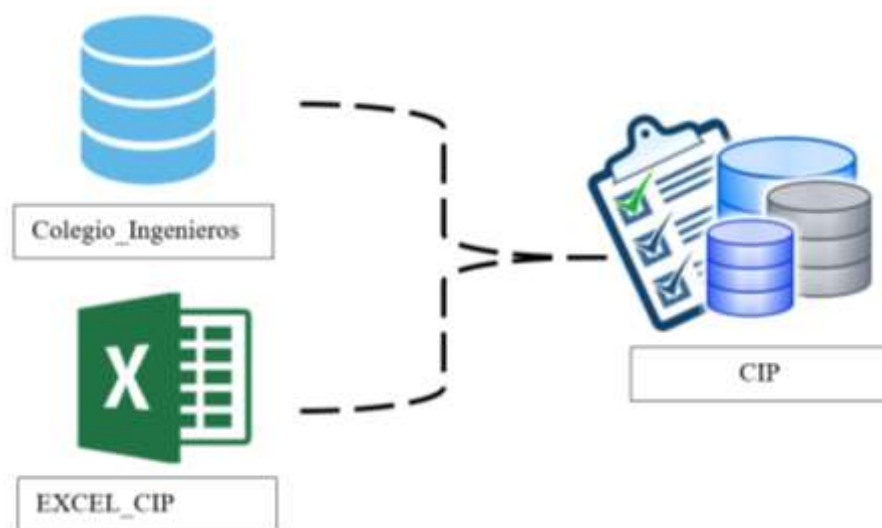


Figura 25. Pasando datos de fuentes de origen a fuente de destino.

De segundo Origen de datos (La Hoja de Cálculo) solo se considero pasar el Tipo de Seguro, ya que Las demás variables de predicción, se pasarían del primer Origen de Datos (Base de Datos SQL).

Algunas consultas realizadas para dicho procedimiento fueron:

```
-----para tabla CAPITULO

merge [CIP].[dbo].[Capitulo] as cap2
using (select * from [Colegio_Ingenieros].[dbo].[tbl_capitulo]) as cap1
on cap1.CodCap collate Modern_Spanish_CI_AS = cap2.CodCap collate
Modern_Spanish_CI_AS
when not matched then
insert(CodCap,NomCap,Orden_Cap)
values(CodCap,NomCap,orden_cap);

----- para tabla ESPECIALIDAD
```

```

merge [CIP].[dbo].[Especialidad] as esp2
using (select * from [Colegio_Ingenieros].[dbo].[tbl_Especialidad] where CodCap !=
'8' ) as esp1
on esp1.CodEsp collate Modern_Spanish_CI_AS = esp2.CodEsp collate
Modern_Spanish_CI_AS
when not matched then
insert(CodEsp,NomEsp,CodCap)
values(CodEsp,NomEsp,CodCap);

-----para tabla COLEGIADO

merge [CIP].[dbo].[Colegiado] as col2
using(select distinct
(c.nroCIP),c.codcli,c.nomcli,c.nrodni,c.ruccli,c.dircli,c.telcli,c.TelCel,c.email,c.
fecnac,c.feccol,
c.estciv,p.nompai,cat.nomcat,e.CodEsp,h.estado, DATEDIFF(YEAR,c.FecNac,GETDATE()) as
Edad from Colegio_Ingenieros.dbo.mst01cli as c
inner join Colegio_Ingenieros.dbo.tbl01pai as p on c.codpai = p.codpai
inner join Colegio_Ingenieros.dbo.tbl01cac as cat on c.codcat = cat.codcat
left join Colegio_Ingenieros.dbo.tbl_especialidad as e on c.codesp = e.CodEsp
left join Colegio_Ingenieros.dbo.habil2204201 as h on c.nroCIP = h.cip
where c.nroCIP not like '' and c.nroCIP <> '72288' and c.nroCIP <> '94001' and
c.nroCIP != '.' and c.NroCIP != 'G-001') as col1
on col1.nroCIP collate Modern_Spanish_CI_AS = col2.NroCIP collate
Modern_Spanish_CI_AS
when not matched then
insert(NroCIP,CodCli,NomCli,NroDni,DirCli,RucCli,TelCli,TelCel,Email,FecNac,Edad,Fec
Col,EstCivil,NomPai,NomCat,CodEsp,Estado_Habilidad)
values(nroCIP,codcli,nomcli,nrodni,dircli,ruccli,telcli,telcel,email,fecnac,Edad,fec
col,estciv,nompai,nomcat,codesp,estado);

```

Las Consultas completas se encuentran en el Anexo 15. **(Ver Anexo 15).**

3.3.3. Realizar el Proceso ETL para la Extracción de datos relevantes.

Una vez filtrada la información más importante en una base transaccional, se buscó como parte del proceso de Inteligencia de negocios, filtrar campos concretos que se utilizaran para pasarlos directamente a los algoritmos, esto se realizó mediante el desarrollo del ETL (Extracción, Transformación y Carga), donde se tomaron y pasaron los datos más importantes y de mayor significancia (tomando en cuenta las variables de predicción a utilizar) a un modelo dimensional.

a) Determinar el modelo de Dimensional

Para determinar el modelo dimensional más conveniente a ser utilizado, se realizó teniendo en cuenta la escala de Likert, en donde se muestra, según a los parámetros que se establecen basándose en la investigación que se desea hacer, el cuál es echo mediante una escala ordenada por lo mismo que este modelo de

instrumentos se fundamentan en una escala de ítems el cual se acompaña de una valoración que se le otorga a cada uno, y sigue una escala numerica desde el 1 hasta el 5 como se detalla en la *Tabla 12*.

Tabla 12

Leyenda de Evaluación de Likert

1	2	3	4	5
Muy Malo	Malo	Regular	Bueno	Muy Bueno

Fuente: (Likert, 1932; Edmondson, 2005).

Evaluación de Modelo Dimensional

Tabla 13

Puntuación para elegir el modelo dimensional más adecuado al caso de estudio.

N ^o	Parámetros	Estrella	Copo de Nieve	Puntaje (Estrella)	Puntaje (Copo de Nieve)
1	Facilidad de uso	Consultas menos complejas, por lo tanto, fáciles de entender.	Consultas más complejas, por lo tanto, menos fáciles de entender.	5	2
2	Rendimiento de consulta o tiempo de ejecución de consulta	Menor número de claves externas, por lo tanto, menos complejidad y menor tiempo de ejecución de consultas	Más claves foráneas, por lo tanto, más complejidad y más tiempo de ejecución de consultas	4	2
3	Normalización	Contiene tablas no normalizadas.	Contiene tablas normalizadas.	4	3
4	Número de uniones (joins)	Menor número de joins (uniones)	Mayor número de joins (uniones)	4	3
5	Tablas de dimensiones	El esquema contiene solo una tabla de dimensiones solitarias para cada dimensión	El esquema puede tener más de una tabla de dimensiones para cada dimensión	4	3

6	Mantenimiento	El esquema contiene datos redundantes y, por lo tanto, no es tan fácil de mantener	No hay datos redundantes, y dado que las tablas están normalizadas, por lo tanto, es más fácil de mantener y actualizar	2	4
7	Tipo de Data Warehouse	Data Mart	Data Warehouse	4	4
PUNTAJE				27	21

Fuente: Grupo de Investigación

El grupo de investigación realizó esta valoración para que al final, el modelo con mayor puntuación, significará que es el que mejor se adecue para esta investigación, tomando en cuenta que las características mencionadas cumplan y se adecuen con la base de datos que se tiene.

Por la puntuación obtenida, se eligió el Modelo de Estrella como modelo Dimensional a utilizarse.

Definición de Requerimientos:

En Primer lugar, para poder construir el modelo dimensional se tuvo que hacer un análisis de requerimientos (que datos se necesitaba exactamente). Los requerimientos que se presenta a continuación, están enfocados al pago de las cuotas sociales del Colegio de Ingenieros CD Lambayeque como se muestra desde la *Tabla 14* hasta la *Tabla 28*.

Tabla 14

Requisito N° 1

Identificador:	R - 1	Nombre:	Monto Pagado por cada colegiado por mes
Tipo:	Funcionales	Necesidad:	Si
Prioridad:	Alta		
Descripción:	Permitirá Conocer el monto pagado (cuota mensual) que realizó cada colegiado.		

Elaborado por el Grupo de Investigación

Tabla 15
Requisito N° 2.

Identificador:	R - 2	Nombre:	Cantidad de colegiados por Especialidad y Capítulo.
Tipo:	Funcional		
Prioridad:	Media	Necesidad:	Si
Descripción:	Nos permitirá conocer todos los colegiados que pertenecen a cada especialidad y a que capítulo pertenecen.		

Elaborado por el Grupo de Investigación

Tabla 16
Requisito N°3.

Identificador:	R - 3	Nombre:	Meses de Mayor Recaudación
Tipo:	Funcionales		
Prioridad:	Alta	Necesidad:	Si
Descripción:	Meses en los que se obtuvo el mayor ingreso de recaudación por el cobro de cuotas sociales de colegiados.		

Elaborado por el Grupo de Investigación

Tabla 17
Requisito N°4.

Identificador:	R - 4	Nombre:	Meses de Menor Recaudación
Tipo:	Funcional		
Prioridad:	Alta	Necesidad:	Si
Descripción:	Meses en los que se obtuvo el menor ingreso de recaudación por el cobro de cuotas sociales de colegiados.		

Elaborado por el Grupo de Investigación

Tabla 18
Requisito N°5.

Identificador:	R - 5	Nombre:	Pagos de Colegiados por Edad.
Tipo:	Funcionales		
Prioridad:	Alta	Necesidad:	Si

Descripción:	Permite conocer el comportamiento de pago de los colegiados por edad.
Elaborado por el Grupo de Investigación	

Tabla 19
Requisito N°6.

Identificador:	R - 6	Nombre:	Cantidad de Colegiados por estado civil.
Tipo:	Funcionales	Necesidad:	Si
Prioridad:	Alta	Descripción: Permite conocer la cantidad de colegiados pertenecientes a cada tipo de estado civil.	
Elaborado por el Grupo de Investigación			

Tabla 20
Requisito N°7.

Identificador:	R - 7	Nombre:	Cantidad de Colegiados por Tipo de Seguro
Tipo:	Funcionales	Necesidad:	Si
Prioridad:	Alta	Descripción: Permite conocer la cantidad de colegiados por el tipo de Seguro.	
Elaborado por el Grupo de Investigación			

Tabla 21
Requisito N°8.

Identificador:	R - 8	Nombre:	Cantidad de Colegiados por Categoría.
Tipo:	Funcionales	Necesidad:	Si
Prioridad:	Baja	Descripción: Permite conocer la cantidad de colegiados pertenecientes a cada Categoría (Ordinario, Vitalicio, Fallecidos, Tranferidos, Transeuntes).	
Elaborado por el Grupo de Investigación			

Tabla 22
Requisito N° 9.

Identificador:	R - 9	Nombre:	Colegiados por Género.
Tipo:	Funcionales		
Prioridad:	Baja	Necesidad:	Si
Descripción:	Ayuda a conocer Cantidad de Colegiados por Género (Masculino y Femenino).		
Elaborado por el Grupo de Investigación			

Tabla 23
Requisito N° 10.

Identificador:	R - 10	Nombre:	Cantidad de Hijos por cada Colegiados.
Tipo:	Funcionales		
Prioridad:	Media	Necesidad:	Si
Descripción:	Ayudará a conocer el número de hijos que tiene cada colegiado.		
Elaborado por el Grupo de Investigación			

Tabla 24
Requisito N°11.

Identificador:	R - 11	Nombre:	Puntualidad de cada pago de la cuota mensual.
Tipo:	Funcionales		
Prioridad:	Alta	Necesidad:	Si
Descripción:	Permitirá conocer si el colegiado pago a tiempo, pago a destiempo o no realizó el pago de la cuota mensual.		
Elaborado por el Grupo de Investigación			

Tabla 25
Requisito N° 12.

Identificador:	R - 12	Nombre:	Pagos de Colegiados por Especialidad.
Tipo:	Funcionales		
Prioridad:	Alta	Necesidad:	Si

Descripción:	Permite conocer el comportamiento de pago de los colegiados por Especialidad.
Elaborado por el Grupo de Investigación	

Tabla 26
Requisito N° 13.

Identificador:	R - 13	Nombre:	Pagos de Colegiados por Tipo de Estado Civil.
Tipo:	Funcionales		
Prioridad:	Alta	Necesidad:	Si
Descripción:	Permite conocer el comportamiento de pago de los colegiados por Tipo de Estado Civil.		
Elaborado por el Grupo de Investigación			

Tabla 27
Requisito N° 14.

Identificador:	R - 14	Nombre:	Pagos de Colegiados según su número de Hijos.
Tipo:	Funcionales		
Prioridad:	Media	Necesidad:	Si
Descripción:	Permite conocer el comportamiento de pago de los colegiados según su número de hijos.		
Elaborado por el Grupo de Investigación			

Tabla 28
Requisito N° 15.

Identificador:	R - 15	Nombre:	Saldo de Colegiados
Tipo:	Funcionales		
Prioridad:	Media	Necesidad:	Si
Descripción:	Permite conocer el saldo que le queda pendiente por pagar a cada colegiado		
Elaborado por el Grupo de Investigación.			

Elección de Dimensiones

Después de haber realizado el análisis de los requerimientos se procede a hallar las dimensiones y medidas, donde los campos que se tomaran en cuenta por su importancia y pertinencia son:

Tabla 29*Campos usados para elaborar el modelo dimensional.*

Nº	Descripción
1	Colegiado – NomCat
2	Pago - Fecha de pago
3	Colegiado – EstCivil
4	Colegiado - NomCli
5	Especialidad - NomEsp
6	Familiar – NomPar
7	Pago - Saldo
8	Pagos - Fecha_vencimiento
9	Pagos - Monto
10	Colegiado - NomCap
11	Colegiado - Género
12	Colegiado - Edad
13	Colegiado - Tipo_Seguro
14	Pago - Días_Incumplimiento

Elaborado por el Grupo de Investigación.

Después de identificar los campos importantes, se procedió a agruparlas por su relación o afinidad entre ellas, ya que generalmente cada una de estas variables serán atributos de una dimensión.

Tabla 30*Elección de dimensiones*

Cuota_Mensual	Pago - Fecha de pago
	Pagos - Fecha_vencimiento
Tiempo	Año
	Mes
	Día

		Fecha
Dimensiones	Colegiado	Colegiado - NomCli
		Colegiado - Edad
		Colegiado - EstCivil
		Colegiado - Tipo_Seguro
		Colegiado - NomCat
		Colegiado - Género
		Especialidad - NomEsp
		Capítulo - NomCap
		Colegiado - Nro_Hijos
		Medidas
	Dias Retrasados de Pago	
	Saldo	

Elaborado por el Grupo de Investigación.

b) Dimensiones Encontradas

Después del análisis respectivo se identificaron las siguientes dimensiones:

- DIM_COLEGIADO
- DIM_CUOTAMENSUAL
- DIM_TIEMPO

c) Tabla Hechos (Medidas Encontradas)

Respecto al análisis se encontraron las siguientes medidas:

- a. MONTO
- b. DIAS RETRASADOS DE PAGO
- c. SALDO

d) Modelo de datos Dimensional

El Modelo Dimensional Final es el siguiente, el cuál tiene por nombre CIP_DM:

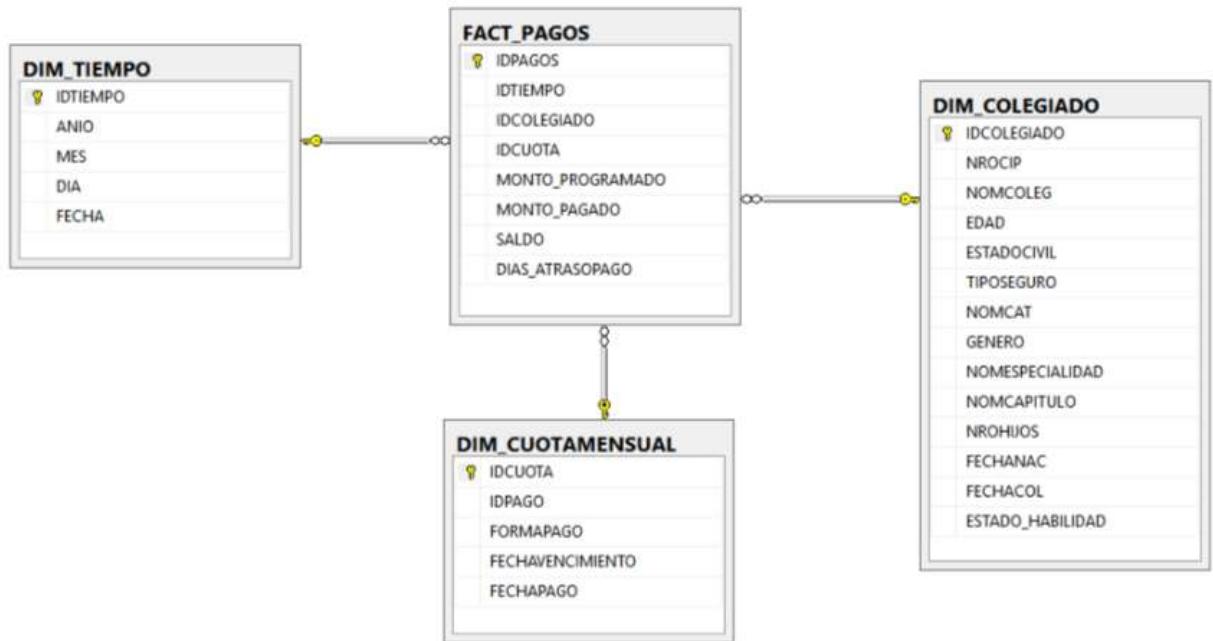


Figura 26. Modelo Físico de Modelo Dimensional CIP_DM.

e) Realizar el proceso de ETL para la extracción de datos relevantes.

Para realizar el poblamiento de las dimensiones, se toma en cuenta la *Tabla 30* mencionada anteriormente, desde donde se realizará la recolección de datos y teniendo en cuenta ello se procedió a realizar las consultas para poblamiento de Dimensiones y Tabla Hechos a la Base de Datos destino, en la *Figura 26*:

Poblamiento de Dimensión Colegiado

Para poblar la dimensión colegiado se realizó la siguiente consulta:

```

select * from (
select
c.nroCip,c.Nomcli,c.Edad,c.EstCivil,c.Tipo_Seguro,c.NomCat,c.genero,
e.NomEsp,ca.NomCap,
c.Nro_Hijos,convert(date,c.FecNac) as Fecnac,convert(date,c.FecCol) as
FecCol,
c.Estado_habilidad
from Colegiado as c
left join Especialidad as e on c.CodEsp = e.CodEsp
left join Capitulo as ca on e.CodCap = ca.CodCap) as c1
inner join EXCEL_CIP.dbo.Colegiados as c2
on c1.NroCIP = c2.CIP
where c1.Edad<86 and c1.Edad>22 and c1.NomCat not like 'FALLECIDO'
order by NomCli

```


Poblamiento de Dimensión Tiempo

Para poblar la dimensión tiempo se realizó la siguiente consulta:

```
select
distinct(FechaApertura),datepart (yyyy,FechaApertura) as Anio,
datepart(mm,FechaApertura) as Mes, datepart(dd,FechaApertura) as dias
from Pagos
where datepart(yyyy,FechaApertura) not like '7' and
datepart(yyyy,FechaApertura) not like '15' and
datepart(yyyy,FechaApertura) not like '1195'
and NroCip in (Select NroCip from CIP_DM.dbo.DIM_COLEGIADO)
order by 1
```

Poblamiento de Dimensión Cuota_Mensual

Para poblar la dimensión Cuota_Mensual se realizó la siguiente consulta:

```
select
p.CodPago,c.NomCdv,p.FechaVencimiento,p.FechaPago
from Pagos p
inner join Cobro c on p.CodCdv = c.CodCdv
where p.NroCIP in (Select NROCIP from CIP_DM.dbo.DIM_COLEGIADO)
```

Poblamiento de Tabla Hechos.

Para poblar la tabla Hechos se realizó la siguiente consulta.

```
select
t.IDTIEMPO,c.IDCOLEGIADO,cm.IDCUOTA,pa.Monto_Programado,pa.Monto_Pagado,
pa.Saldo,pa.dias
from Colegiado p
inner join CIP_DM.dbo.DIM_COLEGIADO c on p.NroCIP = c.NROCIP
inner join Pagos pa on pa.NroCIP = p.NroCIP
inner join CIP_DM.dbo.DIM_CUOTAMENSUAL cm on pa.CodPago = cm.IDPAGO
inner join CIP_DM.dbo.DIM_TIEMPO t on pa.FechaApertura = t.FECHA
where p.NroCIP in (Select NROCIP from CIP_DM.dbo.DIM_COLEGIADO)
order by 1
```

3.3.4. Aplicar las Técnicas Adaptive Boosting y Decision Tree.

3.3.4.1. Preparación de los datos.

Respecto al procesamiento de los datos, según (Trujillo 2017), menciona que “la *data quality* (calidad de los datos) es de vital importancia”, bajo esa situación puede existir datos que talvez terminan siendo innecesarios y aportan ruido en la técnica a utilizar en vez de aportar información de valor.

Por ejemplo, al principio se pensó que las variables relacionadas con el cobro de la cuota mensual del colegiado (Cajero o tesorería que cobró dicho pago y el

Tipo de pago) serían importantes para comprender los pagos realizados por los colegiados, pero después de su análisis resultaron tener un valor estático en todos los registros, se optó por su omisión.

Por otro lado, se identificó que la base de datos necesitaba tomar medidas necesarias para el tratamiento de *missings* (datos ausentes), ya que había demasiada cantidad de campos vacíos referente a los pagos de cada colegiado y otros campos.

Había pagos incluso desde 1994, pero muy pocos, a partir de los años 2000 en adelante se tenía mayor cantidad de registros de pagos de colegiados por lo que se decidió considerar los pagos históricos a partir del año 2000 hasta el año 2016. Respecto al estado civil, había números registros nulos, por lo que se reemplazaron por un valor estándar.

Así mismo, estaban presentes variables definidas como “Identificador”, es decir un código único por cada registro, había un código y un Numero de Cip, que no aportan nada ya que no son características que ayuden en la predicción, y por lo tanto solo aportan ruido y mayor carga computacional al modelo, y se decidió no tomarse en cuenta.

Se identificaron también registros que presentaban *outliers* (datos que son numéricamente muy lejano a los demás), por ejemplo, había registros donde los colegiados tenían más de 118 años de edad, por lo que no se tomaron en cuenta, y se decidió entonces, establecer un rango de edad a considerarse, y no se tomarían en cuenta ni las personas mayores a 80 años y tampoco aquellos colegiados cuya categoría sea Fallecido.

Los datos al final quedaron con una cantidad de 11103 registros, es decir, un registro por cada colegiado, donde se encuentran las variables de predicción seleccionadas (Tipo de Seguro, Genero, Número de Hijos, Historial de pagos, Historial de Puntualidad para cada pago, Especialidad, Edad, Categoría, Estado Civil y una variable llamado: PAGO), este ultima variable se extrajo del pago del año 2017, y se reemplazo de la siguiente manera: si registra pago se le consideró como: SI, caso contrario como NO; obteniendo al final dos tipos de datos. Esto servirá para corroborar con datos reales si un colegiado pagó o no y enviarle al modelo.

Otra variable que se tuvo que estandarizar fueron los días de atraso del pago, se normalizo de la siguiente manera:

- a. PAGO TIEMPO: cantidad de días menores a 0.
- b. PAGO DESTIEMPO: cantidad de días mayores a 0.
- c. NO PAGO: cantidad de días iguales a 0.

Uno de los tipos de archivos que leen los algoritmos en R son los archivos separados por comas (.csv), se tuvo que convertir los datos a dicho formato, y fueron divididos de la siguiente manera:

Tabla 31

Archivos .csv para los algoritmos.

Mes	Para Entrenamiento y Test		Datos Para predecir	
	Nombre .csv	Nº Registro s	Nombre .csv	Nº Registro s
Enero	Enero_Algoritmo	11000	Enero_Prededir	103
Febrero	Febrero_Algoritmo	11000	Febrero_Prededir	103
Marzo	Marzo_Algoritmo	11000	Marzo_Prededir	103
Abril	Abril_Algoritmo	11000	Abril_Prededir	103
Mayo	Mayo_Algoritmo	11000	Mayo_Prededir	103
Junio	Junio_Algoritmo	11000	Junio_Prededir	103
Julio	Julio_Algoritmo	11000	Julio_Prededir	103
Agosto	Agosto_Algoritmo	11000	Agosto_Prededir	103
Septiembre	Septiembre_Algoritmo	11000	Septiembre_Prededir	103
Octubre	Octubre_Algoritmo	11000	Octubre_Prededir	103

Noviembre	Noviembre_Algoritmo	11000	Noviembre_Predicir	103
Diciembre	Diciembre_Algoritmo	11000	Diciembre_Predicir	103

Elaboración Propia.

3.3.4.2. Implementación de Algoritmos

Para la implementación de los algoritmos, tanto Adaboost como Decision Tree, se usó R Studio debido a sus diversas ventajas que ofrece a comparación de otros lenguajes de programación, según (Cervan 2020) algunas de ellos son:

- a) Si bien hay muchos lenguajes de programación, R se ubica entre los mejores, esto se evidencia en un ranking de lenguajes de programación.
- b) R es más rápido de hacer implementaciones y análisis de datos porque en R ya lo tiene implementado a comparación de otros lenguajes donde se tiene que hacer todo de cero.
- c) R tiene soporte para usar análisis de Big Data y Machine Learning. Y cuenta con un sin fin de librerías.

i. Implementación de Algoritmo de Adaboost.

Características:

El algoritmo de Adaboost.M1 que se utilizó en esta investigación, presenta las siguientes características y pasos principales:

- 1) Primero asignamos iguales *pesos* a los datos de entrenamiento y elegimos un *algoritmo base* (López 2017). El clasificador "débil" utilizado en este algoritmo es CART. (Le 2013).
- 2) En cada una de las iteraciones, se aplica el *algoritmo base* al grupo de entrenamiento y se aumenta los pesos de los datos mal clasificados. (López 2017).
- 3) Iteramos *n* veces, siempre empleando el *algoritmo base* en el grupo de datos de entreno con los *pesos* actualizados modificados. (López 2017)

- 4) Al final el modelo resulta de la suma ponderada de lo que resulta de los n algoritmos base. (López 2017)
- 5) Esto implementa el algoritmo Adaboost. M1 para una tarea de clasificación binaria. La variable de destino debe ser un factor con exactamente dos niveles.(Chatterjee 2016)
- 6) El clasificador final es una combinación lineal de clasificadores de árbol de decisión débiles.(Chatterjee 2016)
- 7) La unión de AdaBoost con árboles de decisión ha demostrado ser fuertemente efectivo en diversos problemas de Aprendizaje Automático. (López 2017)

Notación Matemática y Ejemplo:

Según (Brownlee 2016), cada instancia en el grupo de data de entrenamiento es ponderada. El peso inicial se establece en:

$$weight(x_i) = \frac{1}{n}$$

Donde x_i es la i^{ma} instancia de entrenamiento y n es el número de instancias de entrenamiento.

Luego, la tasa de clasificación errónea se calcula para el modelo entrenado. Tradicionalmente, esto se calcula como:

$$error = \frac{correct - N}{N}$$

Donde $error$ es la tasa de clasificación errónea, $correct$ es el total de datos de entrenamiento pronosticadas correctamente por el modelo y N es el número total de instancias de entrenamiento.

Esto se modifica para utilizar la ponderación de las instancias de entrenamiento:

$$error = \frac{\sum_{i=1}^n (w_i * perror_i)}{\sum_{i=1}^n w}$$

Cuál es la suma ponderada de la tasa de errores de clasificación, donde w es el peso para la instancia de entrenamiento i y $perror$ es el error de predicción para la instancia de entrenamiento i que es 1 si se clasifica erróneamente y 0 si se clasifica correctamente.

Se calcula un valor de etapa (*stage*) para el modelo entrenado que proporciona una ponderación para cualquier predicción que haga el modelo. El valor de etapa para un modelo entrenado se calcula de la siguiente manera:

$$\text{iii) } \quad \textit{stage} = \ln\left(\frac{1-\textit{error}}{\textit{error}}\right)$$

$$\text{iv) } \quad \textit{stage} = \frac{1}{2} \ln\left(\frac{1-\textit{error}}{\textit{error}}\right)$$

Se puede utilizar cualquiera de los dos fórmulas, ya que lo que variaría es el coeficiente de actualización de peso (*stage*). (Alfaro, s.f.).

Donde *stage* es el valor de etapa utilizado para ponderar las predicciones del modelo, $\ln()$ es el logaritmo natural y *error* es el error de clasificación errónea del modelo. El efecto del peso de la *stage* es que los modelos más precisos tienen más peso o contribución a la predicción final.

Finalmente los pesos de entrenamiento se actualizan dando más peso a las instancias predichas incorrectamente, y menos peso a las instancias predichas correctamente. Por ejemplo, el peso de una instancia de entrenamiento (*w*) se actualiza usando:

$$w = w * e^{\textit{stage} * \textit{perror}}$$

Donde *w* es el peso para una instancia de entrenamiento específica, *e* es la constante numérica del número de Euler elevado a una potencia, *stage* es la tasa de error de clasificación para el clasificador débil y *perror* es el error que hizo el clasificador débil para predecir la variable de salida para la instancia de entrenamiento. evaluado como:

$$\textit{perror} = 0 \text{ IF } y == p$$

$$\textit{peror} = 1 \text{ IF } y \neq p$$

Donde *y* es la variable de salida para la instancia de entrenamiento y *p* es la predicción del aprendiz débil. Esto tiene el efecto de no cambiar el peso si la instancia de entrenamiento se clasificó correctamente y hacer el peso un poco más grande si el alumno débil clasificó incorrectamente la instancia.

Los modelos débiles se agregan secuencialmente, entrenados usando los datos de entrenamiento ponderados. El proceso continúa hasta que se haya creado un número preestablecido de alumnos débiles (un parámetro de usuario) o no se puedan realizar mejoras adicionales en el conjunto de datos

de capacitación. Una vez completado, te quedas con un grupo de alumnos débiles, cada uno con un valor de escenario.

Las predicciones se realizan calculando el promedio ponderado de los clasificadores débiles. Para una nueva instancia de entrada, cada aprendiz débil calcula un valor predicho como +1.0 o -1.0. Los valores predichos son ponderados por cada valor de etapa de los estudiantes débiles. La predicción para el modelo de conjunto se toma como la suma de las predicciones ponderadas. Si la suma es positiva, entonces se predice la primera clase, y si es negativa se predice la segunda clase.

Ejemplo de Adaboost:

Todo lo descrito anteriormente se aplicó con los datos reales de este estudio y se elaboró un ejemplo descrito a continuación:

En este ejemplo se usarán 10 instancias con 42 variables de entrada y una salida variable (PAGO) (Todas las instancias pertenecen a datos reales pertenecientes al presente caso de estudio). Las variables de entrada son los valores que se identificó como Variables de Predicción. La variable de salida tiene dos valores (SI y NO) que hace alusión si pagó o no, lo que hace que el problema sea una clasificación de problema binario.

Solo se considerará por el ejemplo 3 iteraciones, ésto implicará aprender 3 modelos (construido 1 modelo por cada iteración), uno tras otro, de modo que podamos observar el efecto sobre las ponderaciones de las instancias de entrenamiento y cómo se combinan las predicciones de cada uno de los dos modelos. AdaBoost usa el tocón de decisión (árboles de decisión de un nodo) como modelo interno. Se dejó por defecto el algoritmo con el que trabaja internamente Adaboost (el algoritmo CART) para elegir los puntos de división para estos árboles de decisión. Nuevamente, estos puntos de división se eligieron para crear errores de clasificación y para demostrar cómo el segundo modelo puede corregir el primer modelo, y así sucesivamente. Al final se corroborará que los resultados obtenidos en este ejemplo coincidan con los resultados que mostró el programa R al ingresarle los mismos datos y mismas iteraciones.

Decisión Modelo 1 (Iteración 1).

El primer modelo será un nodo de decisión para la división de la variable de entrada P.2016.

- SI P.2016 < 9 ENTONCES LEFT
- SI P.2016 >=9 ENTOCES RIGHT

Se procedió a comprobar las decisiones seleccionadas y a ubicarlas dentro de un Grupo: RIGHT (Si cumplen con la segunda condición) o LEFT (si cumplen con la primera condición).

Tabla 32

Dividiendo instancias en Grupo para Modelo 1.

Nº INSTANCIA	PAGO	P.2016	GRUPO
1	SI	18	RIGHT
2	SI	18	RIGHT
3	SI	18	RIGHT
4	SI	18	RIGHT
5	SI	18	RIGHT
6	SI	18	RIGHT
7	SI	0	LEFT
8	SI	18	RIGHT
9	SI	18	RIGHT
10	SI	18	RIGHT
11	SI	18	RIGHT
12	SI	18	RIGHT
13	SI	18	RIGHT
14	NO	0	LEFT
15	NO	0	LEFT
16	NO	0	LEFT
17	NO	0	LEFT
18	NO	0	LEFT

19	NO	0	LEFT
20	NO	0	LEFT

Fuente: Elaboracion Propia

Luego se procede a colocar en la *TABLA 33* el dato real (PAGO) y la predicción (Relacionada con cada grupo: RIGHT y LEFT), la columna error hace alusión a:

Si es 0, las condiciones fueron exactas,

Si es 1, las condiciones no coincidieron con los datos reales.

Tabla 33

Datos Real y Dato Arrojado por la division de Grupo (RIGHT y LEFT) y ERROR

N.º INSTANCIA	PAGO	PREDICCIÓN	ERROR
1	SI	SI	0
2	SI	SI	0
3	SI	SI	0
4	SI	SI	0
5	SI	SI	0
6	SI	SI	0
7	SI	NO	1
8	SI	SI	0
9	SI	SI	0
10	SI	SI	0
11	SI	SI	0
12	SI	SI	0
13	SI	SI	0
14	NO	NO	0
15	NO	NO	0
16	NO	NO	0
17	NO	NO	0

18	NO	NO	0
19	NO	NO	0
20	NO	NO	0

Fuente: Elaboracion Propia

Cada instancia en el grupo de datos de entrenamiento tiene el peso inicial de $1 / N$, donde N es el número de instancias de entrenamiento, en este caso 20, así que $1/20 = 0,05$.

PESOS INICIALES

$$weight = \frac{1}{20}$$

$$weight = 0,05$$

Usando este peso inicial y los errores de predicción anteriores, podemos calcular el error ponderado para cada predicción.

$$\text{WeightedError} = \text{weight} * \text{error}$$

Tabla 34
Hallando WeightedError para Modelo 1

WEIGHT	ERROR	Weighted Error
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	1	0.05
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0

0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0

Fuente: Elaboración Propia

Ahora podemos calcular la tasa de clasificación errónea como:

$$MisclassificationRate = \frac{\sum(\text{WeightedError})}{\sum(\text{Weight})}$$

$$MisclassificationRate = \frac{0.05}{1.0}$$

$$MisclassificationRate = 0.05$$

Finalmente, podemos usar la tasa de errores de clasificación para calcular la etapa (Stage) para este modelo débil. La etapa es el peso aplicado a cualquier predicción hecha por este modelo más adelante cuando la usamos para hacer predicciones.

Como estamos usando el coeflearn = 'Breiman'(es el que viene por defecto), para calcular el stage se usa la siguiente formula:

$$stage = \frac{1}{2} \ln \left(\frac{1 - error}{error} \right)$$

Por lo que La etapa se calcula como:

$$Stage = \frac{1}{2} \ln \left(\frac{1 - 0.05}{0.05} \right)$$

$$Stage = \frac{1}{2} \ln(19)$$

$$Stage = \frac{1}{2} (2.94443897917)$$

$$Stage = \mathbf{1.47221948958}$$

Actualizar pesos de instancia: Antes de que podamos preparar un segundo modelo mejorado, debemos actualizar los pesos de instancia. Este es el núcleo del impulso. Las ponderaciones se actualizan de modo que el siguiente

modelo que se crea preste más atención a las instancias de entrenamiento que los modelos anteriores se equivocaron y menos atención a las instancias que acertaron. El peso para una instancia de entrenamiento se actualiza usando:

$$w = w * e^{stage * perror}$$

Sabemos el peso actual para cada instancia de entrenamiento (0.05) y los errores cometidos en cada instancia de entrenamiento. Ahora también conocemos el escenario. Actualizar los pesos para cada instancia de entrenamiento es, por lo tanto, bastante sencillo.

El único error es el de la instancia 7 y se calcula así:

$$Weight = 0.05 * e^{(1.47221948958 * 1)}$$

$$Weight = 0.21794494717$$

A continuación, se muestran los pesos actualizados para cada instancia de entrenamiento. Se muestra que solo las ponderaciones para la instancia en la que el primer modelo se equivocó es la diferente. De hecho, es más grande para que el próximo modelo que se crea les preste más atención.

Tabla 35
Pesos Actualizados para Modelo 1.

PESOS ACTUALIZADOS	
Nº	WEIGHT
INSTANCIA	
1	0.05
2	0.05
3	0.05
4	0.05
5	0.05
6	0.05
7	0.21794494717
8	0.05
9	0.05
10	0.05
11	0.05

12	0.05
13	0.05
14	0.05
15	0.05
16	0.05
17	0.05
18	0.05
19	0.05
20	0.05

Fuente: Elaboración Propia.

Decisión Modelo 2 (Iteración 2)

Ahora podemos crear nuestro segundo modelo débil a partir de las instancias de entrenamiento ponderadas. Este segundo modelo también será un nodo de decisión, pero esta vez se hará una división en la variable T.2016. Esta variable tiene relación con la variable usada en el Modelo 1.

Se debe tener en cuenta que el algoritmo CART tomó en cuenta el peso de la instancia y se centró en dividir en los casos con un peso mayor para dar como resultado un grupo de decisión diferente. Si no, el algoritmo creará el mismo grupo de decisión en un modelo tras otro y no habrá posibilidad de corregir las predicciones hechas de modelos anteriores. Pasaremos por el mismo proceso usando este nuevo punto de división.

- SI T.2016 = NO PAGO ENTONCES LEFT
- SI T.2016 = PAGO DESTIEMPO, PAGO TIEMPO ENTONCES RIGHT

Tabla 36

Dividiendo instancias en Grupos para Modelo 2.

N.º INSTANCIA	PAGO	T.2016	GRUPO
1	SI	PAGO DISTIEMPO	RIGHT
2	SI	PAGO TIEMPO	RIGHT
3	SI	PAGO DISTIEMPO	RIGHT

4	SI	PAGO TIEMPO	RIGHT
5	SI	PAGO DISTIEMPO	RIGHT
6	SI	PAGO TIEMPO	RIGHT
7	SI	NO PAGO	LEFT
8	SI	PAGO TIEMPO	RIGHT
9	SI	PAGO DISTIEMPO	RIGHT
10	SI	PAGO DISTIEMPO	RIGHT
11	SI	PAGO DISTIEMPO	RIGHT
12	SI	PAGO TIEMPO	RIGHT
13	SI	PAGO DISTIEMPO	RIGHT
14	NO	NO PAGO	LEFT
15	NO	NO PAGO	LEFT
16	NO	NO PAGO	LEFT
17	NO	NO PAGO	LEFT
18	NO	NO PAGO	LEFT
19	NO	NO PAGO	LEFT
20	NO	NO PAGO	LEFT

Fuente: Elaboración Propia

Tabla 37

Datos Real y Dato Arrojado por la división de Grupo (RIGHT y LEFT) y ERROR para Modelo 2

Nº INSTANCIA	PAGO	PREDICCIÓN	ERROR
1	SI	SI	0
2	SI	SI	0
3	SI	SI	0
4	SI	SI	0

5	SI	SI	0
6	SI	SI	0
7	SI	NO	1
8	SI	SI	0
9	SI	SI	0
10	SI	SI	0
11	SI	SI	0
12	SI	SI	0
13	SI	SI	0
14	NO	NO	0
15	NO	NO	0
16	NO	NO	0
17	NO	NO	0
18	NO	NO	0
19	NO	NO	0
20	NO	NO	0

Fuente: Elaboración Propia

Usando este peso actualizado del modelo anterior y los errores de predicción anteriores, podemos calcular el error ponderado para cada predicción del modelo 2.

$$\text{WeightedError} = \text{weight} * \text{error}$$

Tabla 38

Hallando WeightedError para Modelo 2.

WEIGHT	ERROR	Weighted Error
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.21794494717	1	0.21794494717

0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0

Fuente: Elaboración Propia

Ahora podemos calcular la tasa de clasificación errónea como:

$$\begin{aligned}
 \text{MisclassificationRate} &= \frac{\sum(\text{WeightedError})}{\sum(\text{Weight})} \\
 \text{MisclassificationRate} &= \frac{0.21794494717}{1.16794494717} \\
 \text{MisclassificationRate} &= 0.18660549686
 \end{aligned}$$

Por lo que La etapa se calcula como:

$$\begin{aligned}
 \text{Stage} &= \frac{1}{2} \ln \left(\frac{1 - 0.1866479197}{0.1866479197} \right) \\
 \text{Stage} &= \frac{1}{2} \ln(4.35889894365) \\
 \text{Stage} &= \frac{1}{2} (1.47221948961) \\
 \text{Stage} &= \mathbf{0.7361097448}
 \end{aligned}$$

ACTUALIZAMOS LOS PESOS

El único error en este modelo también es el de la instancia 7 y se calcula así:

$$\begin{aligned}
 \text{Weight} &= 0.21794494717 * e^{(0.7361097448 * 1)} \\
 \text{Weight} &= 0.45502494416
 \end{aligned}$$

A continuación, se muestran los pesos actualizados para cada instancia de entrenamiento. Notará que solo las ponderaciones para la instancia en la que el primer modelo se equivocó es la diferente. De hecho, es más grande para que el próximo modelo que se crea les preste más atención.

Tabla 39
Pesos Actualizados para Modelo 2.

PESOS ACTUALIZADOS	
Nº	WEIGHT
INSTANCIA	
1	0.05
2	0.05
3	0.05
4	0.05
5	0.05
6	0.05
7	0.45502494416
8	0.05
9	0.05
10	0.05
11	0.05
12	0.05
13	0.05
14	0.05
15	0.05
16	0.05
17	0.05
18	0.05
19	0.05
20	0.05

Fuente: Elaboración Propia.

Decisión Modelo 3

Al no encontrar otro dato para dividir, se procede a dar como regla que todos los datos predichos sean SI, a ver qué resultado arroja dicha condición:

Tabla 40

Datos Real y Dato Arrojado por la división de Grupo (RIGHT y LEFT) y ERROR para Modelo 2.

Nº INSTANCIA	PAGO	PREDICCIÓN	ERROR
1	SI	SI	0
2	SI	SI	0
3	SI	SI	0
4	SI	SI	0
5	SI	SI	0
6	SI	SI	0
7	SI	SI	0
8	SI	SI	0
9	SI	SI	0
10	SI	SI	0
11	SI	SI	0
12	SI	SI	0
13	SI	SI	0
14	NO	SI	1
15	NO	SI	1
16	NO	SI	1
17	NO	SI	1
18	NO	SI	1
19	NO	SI	1
20	NO	SI	1

Fuente: Elaboración Propia

Usando este peso actualizado del modelo anterior y los errores de predicción anteriores, podemos calcular el error ponderado para cada predicción del modelo 2.

$$\text{WeightedError} = \text{weight} * \text{error}$$

Tabla 41

Hallando WeightedError para Modelo 3.

WEIGHT	ERROR	Weighted Error
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.45502494416	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	0	0
0.05	1	0.05
0.05	1	0.05
0.05	1	0.05
0.05	1	0.05
0.05	1	0.05
0.05	1	0.05
0.05	1	0.05
0.05	1	0.05

Fuente: Elaboración Propia

Ahora podemos calcular la tasa de clasificación errónea como:

$$\text{MisclassificationRate} = \frac{\sum(\text{WeightedError})}{\sum(\text{Weight})}$$

$$\text{MisclassificationRate} = \frac{0.35}{1.40502494416}$$

$$\text{MisclassificationRate} = 0.24910589769$$

Por lo que Stage (Etapa) se calcula como:

$$Stage = \frac{1}{2} \ln \left(\frac{1 - 0.24910589769}{0.24910589769} \right)$$

$$Stage = \frac{1}{2} \ln(3.01435698338)$$

$$Stage = \frac{1}{2}(1.10338653492)$$

$$Stage = 0.55169326745$$

PREDICCIÓN

Ahora hemos terminado con las 3 iteraciones. Veamos cómo podemos usar esta serie de modelos potenciados para hacer predicciones:

Ahora podemos hacer predicciones para el conjunto de datos de entrenamiento utilizando el modelo AdaBoost. Las predicciones para un solo clasificador son +1 o -1 y ponderadas por el valor de etapa del modelo. Por ejemplo, para este problema usaremos:

$$prediction = Stage (IF (output == 0) THEN -1 ELSE +1)$$

Usando los tres modelos anteriores, podemos hacer predicciones ponderadas dados los valores de entrada de los datos de entrenamiento. Esto podría ser fácilmente una nueva información para la cual nos gustaría hacer predicciones.

Se enumeran las instancias con el los Stage de cada modelo multiplicado +1 o -1 según se clasificó en cada caso:

Tabla 42

Instancia con los resultados predicción de los Modelos creados en cada Iteración.

Nro. INSTANCIA	MODELO 1	MODELO 2	MODELO 3
1	1.47221949	0.736109745	-0.551693267
2	1.47221949	0.736109745	-0.551693267
3	1.47221949	0.736109745	-0.551693267
4	1.47221949	0.736109745	-0.551693267
5	1.47221949	0.736109745	-0.551693267

6	1.47221949	0.736109745	-0.551693267
7	-1.47221949	-0.736109745	-0.551693267
8	1.47221949	0.736109745	-0.551693267
9	1.47221949	0.736109745	-0.551693267
10	1.47221949	0.736109745	-0.551693267
11	1.47221949	0.736109745	-0.551693267
12	1.47221949	0.736109745	-0.551693267
13	1.47221949	0.736109745	-0.551693267
14	-1.47221949	-0.736109745	0.551693267
15	-1.47221949	-0.736109745	0.551693267
16	-1.47221949	-0.736109745	0.551693267
17	-1.47221949	-0.736109745	0.551693267
18	-1.47221949	-0.736109745	0.551693267
19	-1.47221949	-0.736109745	0.551693267
20	-1.47221949	-0.736109745	0.551693267

Fuente: Elaboración propia.

Podemos sumar las predicciones para cada modelo para dar un resultado final. Si un resultado es menor que 0, entonces se predice la clase 0, y si un resultado es mayor que 0, se predice la clase 1.

Ahora podemos calcular las predicciones finales del modelo AdaBoost. Pero para eso, se debe sumar todas las predicciones de los modelos creados en cada iteración, por cada Instancia.

Tabla 43*Sumatoria de predicciones de Modelos Creados por cada Instancia.*

Nro. INSTANCIA	MODELO 1	MODELO 2	MODELO 3	SUMA
1	1.47221949	0.736109745	-0.551693267	1.65663596693
2	1.47221949	0.736109745	-0.551693267	1.65663596693
3	1.47221949	0.736109745	-0.551693267	1.65663596693
4	1.47221949	0.736109745	-0.551693267	1.65663596693
5	1.47221949	0.736109745	-0.551693267	1.65663596693
6	1.47221949	0.736109745	-0.551693267	1.65663596693
7	-1.47221949	-0.736109745	-0.551693267	-2.76002250183
8	1.47221949	0.736109745	-0.551693267	1.65663596693
9	1.47221949	0.736109745	-0.551693267	1.65663596693
10	1.47221949	0.736109745	-0.551693267	1.65663596693
11	1.47221949	0.736109745	-0.551693267	1.65663596693
12	1.47221949	0.736109745	-0.551693267	1.65663596693
13	1.47221949	0.736109745	-0.551693267	1.65663596693
14	-1.47221949	-0.736109745	0.551693267	-1.65663596693
15	-1.47221949	-0.736109745	0.551693267	-1.65663596693
16	-1.47221949	-0.736109745	0.551693267	-1.65663596693
17	-1.47221949	-0.736109745	0.551693267	-1.65663596693
18	-1.47221949	-0.736109745	0.551693267	-1.65663596693
19	-1.47221949	-0.736109745	0.551693267	-1.65663596693
20	-1.47221949	-0.736109745	0.551693267	-1.65663596693

Fuente: Elaboración Propia.

Podemos ver que las predicciones coinciden perfectamente con los valores de Y esperados, excepto 1 instancia. O, dicho de otra manera, el modelo AdaBoost logró una precisión del 95% en los datos de entrenamiento como se comprueba.

Tabla 44

Tabla Final de Error comparando la salida Original, con la salida Predicha por los modelos.

SUMA	PREDICCIÓN	PAGO (VARIABLE REAL)	ERROR
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
-2.76002250183	NO	SI	1
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
1.65663596693	SI	SI	0
-1.65663596693	NO	NO	0
-1.65663596693	NO	NO	0

-1.65663596693	NO	NO	0
-1.65663596693	NO	NO	0
-1.65663596693	NO	NO	0
-1.65663596693	NO	NO	0
-1.65663596693	NO	NO	0

Fuente: Elaboración propia

Matriz de Confusión:

Tabla 45
Matriz de Confusión Final.

	NO	SI
NO	7	1
SI	0	12

Fuente: Elaboración propia.

Donde:

$$Accuracy = \frac{12 + 7}{20}$$

$$Accuracy = 0.95$$

$$Accuracy = 95\%$$

Los resultados, tanto de los Stage como de Accuracy, se comprueba cargando la misma data en R y programándole 3 iteraciones, se puede apreciar que los valores coinciden perfectamente.


```

> modelito$weights #pesos para todos los arboles
[1] 1.4722195 0.7361097 0.5516933
> predicciones <- predict(object = modelito, newdata=colegiados.entreno, type = "class")
> # Matriz de confusión
> confusionMatrix(predicciones$confusion)
Confusion Matrix and Statistics

              observed Class
Predicted Class NO SI
              NO  7  1
              SI  0 12

              Accuracy : 0.95
              95% CI   : (0.7513, 0.9987)
              No Information Rate : 0.65
              P-value [Acc > NIR] : 0.002133

              Kappa : 0.8936
              Mcnemar's Test P-value : 1.000000

              Sensitivity : 1.0000
              Specificity : 0.9231
              Pos Pred Value : 0.8750
              Neg Pred Value : 1.0000
              Prevalence : 0.3500
              Detection Rate : 0.3500
              Detection Prevalence : 0.4000
              Balanced Accuracy : 0.9615

              'Positive' Class : NO

```

Figura 27. Comparación con Resultados arrojados en R bajo los mismos datos y argumentos

Como se demostró, el ejemplo mostrado en la *Figura 27*, se puede automatizar en la herramienta R. Viendo mas a detalle las sintáxis de R, según la documentación oficial mas reciente del lenguaje R (Alfaro, Gamez, and Garcia 2018), mencionan que la sintaxis principal de la técnica es el siguiente:

boosting (formula, data, boos = TRUE, mfinal = 100, coeflearn = 'Breiman', control...)

Y sus argumentos son los siguientes:

formula: una fórmula, (hace alusión generalmente a la variable independiente o un conjunto de variables excluidas para el entrenamiento).

data: un marco de datos en el que interpretar las variables nombradas en la fórmula.

boos: si es VERDADERO (de forma predeterminada), se dibuja una muestra de arranque del conjunto de entrenamiento utilizando los pesos para cada observación en esa iteración. Si es FALSO, cada observación se utiliza con sus pesos.

mfinal: un número entero, el número de iteraciones para las que se ejecuta el impulso o el número de árboles a utilizar. Por defecto a mfinal = 100 iteraciones.

Coeflearn: si se usa 'Breiman' (por defecto), se usa $\alpha = 1 / 2 \ln ((1-\text{err}) / \text{err})$. Si se usa 'Freund' $\alpha = \ln ((1-\text{err}) / \text{err})$. En ambos casos, se utiliza el algoritmo AdaBoost.M1 y α es el coeficiente de actualización de peso. Por otro lado, si coeflearn es 'Zhu', el algoritmo SAMME se implementa con $\alpha = \ln ((1-\text{err}) / \text{err}) + \ln (\text{nclasses}-1)$.

control: opciones que controlan los detalles del algoritmo rpart.

...: otros argumentos pasaron a otros métodos.

Implementación:

Para elegir bien los argumentos que se utilizaran en este caso de estudio, se tomo en cuenta el estudio realizado según (Alfaro, Gámez, and García 2013), donde se realizaron diversas pruebas con argumentos principales que al combinarlos influían en los resultados, estos argumentos eran *mfinal*, *coeflearn* y *boos*.

Ellos realizaron 400 iteraciones(*mfinal*), y los argumentos de *coeflearn* y *boos* los fueron combinando hasta obtener el menor error posible, obteniendo los siguientes resultados:

Tabla 46
Error de prueba en porcentaje.

Options	400 Iterations	Minimum (Iterations)
"Breiman"/FALSE	15.36	15.15 (374)
"Breiman"/TRUE	19.28	17.20 (253)
"Freund"/FALSE	11.42	11.42 (398)
"Freund"/TRUE	16.84	16.05 (325)

Fuente: (Alfaro et al. 2013)

Por lo tanto, la combinación ganadora en términos de accuracy está aumentando con *coeflearn* = "Freund" y *boos* = FALSE.

Teniendo en cuenta esa comparación, antes de aplicar el modelado a todos los meses para esta investigación, se hizo una comparación previa para saber que combinación de argumentos presenta mejores resultados para este caso de estudio, tomando un mes para las pruebas, y se realizan diferentes iteraciones (10, 50, 100 y 400), esto se detalla en la Tabla 47.

Tabla 47

Identificando el mejor ajuste del modelo de Adaboost al caso de estudio (evaluando accuracy).

coflearn	boos	ACCURACY			
		ITERACIONES			
		10	50	100	400
Breiman	TRUE	0.8621	0.8624	0.8618	0.8603
Breiman	FALSE	0.8624	0.8679	0.8679	0.8679
Freund	TRUE	0.8542	0.8582	0.8548	0.8548
Freund	FALSE	0.8545	0.8582	0.8612	0.8621

Fuente: Elaboracion Propia

Observando la *Tabla 47* el mejor ajuste es utilizando los argumentos de *coflearn* = "Breiman" y *boos* = FALSE, ya que iterando 50, 100 y 400 veces, el accuracy se mantiene en 0.8679, y es éste el mejor entre los demás resultados, por lo que se usará estos argumentos en esta investigación. En cuanto a las iteraciones, se elegirá 100 iteraciones (por defecto) porque no tardo mucho al construir el modelo, aprox. 1 min, a comparación de 400 iteraciones que demoro aprox. 9 min.

Parte del código utilizado es el siguiente:

```

colegiados <- read.csv("Enero_Algoritmo.csv",dec = ".")
set.seed(101)
tamano.total <- nrow(colegiados)
tamano.entreno <- round(tamano.total*0.7)
colegiados.indices <- sample(1:tamano.total , size=tamano.entreno)
colegiados.entreno <- colegiados[colegiados.indices,]
colegiados.test <- colegiados[-colegiados.indices,]

# Ejecucion del modelo
summary(modelito)
# test de prediccion
prediccion

# Matriz de Confusion
confusion

# Correccion de errores
100 * (sum(diag(predicciones$confusion)) / sum(predicciones$confusion))

```

Figura 28. Cargando Datos y separando en Entrenamiento (70%) y Test (30%).

Figura 29. Creando el Modelo de Adaboost, Evaluandolo con el Test y Generando Matriz de Confusion.

```

colAUC(as.numeric(as.factor(predicciones$class)),as.factor(colegiados.test$PAGO),plotROC = TRUE ,
      alg = c("wilcoxon", "ROC"))

abline(a=0, b=1)

```

Figura 30. Aplicando AUC - Curva Roc

El código completo de Adaboost utilizado se encuentra en el **Anexo 8**.

ii. Implementación de Algoritmo de Decisión Tree.

Características:

El Árbol de Decisión C5.0 que se usará en el presente trabajo tiene las siguientes características y mejoras que su predecesor (C4.5) según (Pandya et al. 2015):

- 1) C5.0 es más rápido que C 4.5.
- 2) Usa la memoria más eficientemente C5.0 que C4.5.
- 3) C5.0 logra árboles de decisión mucho más diminutos comparándolo con C4.5.
- 4) Los grupos de reglas C5.0 tienen menos errores en casos no vistos.
- 5) C5.0 permite automáticamente la eliminación de atributos inútiles.
- 6) El árbol de decisión grande se puede ver como un agrupamiento de reglas que son muy comprensibles y fáciles de entender.
- 7) El árbol de decisión C5.0 otorga el reconocimiento en ruido y datos que falta.
- 8) Se reesuelve el problema de sobreinstalación y la poda de errores en el algoritmo C5.0.

- 9) En cuanto a la técnica usada para clasificar, el algoritmo C5.0 puede prever cuáles atributos son más importantes y cuáles no lo son para realizar la clasificación.
- 10) El método C5.0 es una extensión adicional de C4.5 y el pináculo de esa línea de métodos. Fue propietario durante mucho tiempo, aunque el código fue publicado y está disponible en el paquete C5.0 (Brownlee 2014).
- 11) El modelo cuenta con algunas características y opciones en las que se resalta: creación de un único árbol de clasificación, el procedimiento de refuerzo de C5.0 (boosting) entre otras características diversas del algoritmo. (Kuhn and Johnson 2013).

Procedimiento de Notación Matemática de creación de un único árbol de decisión C5.0:

C5.0 elabora árboles de decisión basándose en un conjunto de datos de entrenamiento, empleando la idea de entropía de información. (Luis 2019).

Al principio, tomamos el conjunto de muestras dado como la raíz del árbol de decisión. En segundo lugar, calculamos la relación de ganancia de información de cada atributo de las muestras de entrenamiento y seleccionamos el dato con la relación de ganancia de información más alta y se usa como el atributo dividido. Luego, creamos un nodo para el atributo dividido y usamos el atributo para hacer un indicador para el nodo. Tercero, creamos una rama para cada valor del atributo dividido y de acuerdo con esto, dividimos el conjunto de datos en varios subconjuntos. Obviamente, el número de subconjuntos es igual al número de los valores del atributo de prueba. (Pang and Gong 2009).

Paso 1:

(Entropía de información): suponga que un conjunto dado S consta de n muestras. El atributo de clase C tiene m valores diferentes C_i ($i = 1, 2, \dots, m$). Cada muestra pertenece a una de las m clases. Sea n_i el número de muestras de clase C_i . Luego, la entropía de información $E(S)$ necesaria para clasificar S se define de la siguiente manera:

$$E(S) = - \sum_{i=1}^m p_i \log_2 p_i$$

Donde p_i denota la proporción del número de muestras de clase C_i al número de todas las muestras en el conjunto de entrenamiento.

Por lo general, podemos calcularlo por $p_i = n_i / |S|$. Donde $|S|$ denota el número de muestras en el conjunto S , $|S| = n$.

Paso 2:

(Entropía de información condicional): suponga que el atributo A tiene v valores diferentes (a_1, a_2, \dots, a_v), que dividen el conjunto S en v subconjuntos (S_1, S_2, \dots, S_v). Sea n_{ij} el número de muestras de clase C_i en el subconjunto S_j . Entonces la entropía condicional del atributo $E(S|A)$ se define de la siguiente manera:

$$E(S|A) = - \sum_{j=1}^v p'_j \sum_{i=1}^m p_{ij} \log_2(p_{ij})$$

Donde p'_j denota la proporción del número de muestras cuyos valores del atributo A son a_j al número de todas las muestras en el conjunto de entrenamiento,

$$p'_j = |S_j| / |S| = \sum_{i=1}^m n_{ij} / n.$$

p_{ij} es la probabilidad condicional de que la muestra cuyo valor del atributo A es a_j pertenezca a la clase C_i , $p_{ij} = n_{ij} / |S_j|$.

$|S_j|$ es el total de muestras cuyos datos del atributo A son a_j , $|S_j| = \sum_{i=1}^m n_{ij}$.

Paso 3:

(Ratio de ganancia de información): En cuanto a la ganancia de información del atributo A se define como:

$$Gain(A) = E(S) - E(S|A)$$

Una vez hallado la ganancia del atributo A , se procede hacer lo mismo con los demás atributos. Luego se escoge como nodo raíz el atributo con mas ganancia. Después se vuelve a repetir el proceso para hallar a los nodos hijos y así

sucesivamente, hasta que todos los subconjuntos puedan satisfacer una de las siguientes tres condiciones:

- (1) Todas las muestras en el subconjunto pertenecen a una clase. Al mismo tiempo, creamos hojas para subconjuntos.
- (2) Todos los atributos del subconjunto se han transaccionado y no quedan atributos que puedan usarse para dividir el conjunto de datos. En esta condición, clasificamos este subconjunto como la clase a la que pertenecen la mayoría de las muestras y creamos un permiso para el subconjunto.
- (3) Todos los atributos de prueba que quedan de las muestras en el subconjunto tienen el mismo valor, pero las muestras aún pertenecen a diferentes clases. En esta condición, creamos un permiso a cuya clase pertenecen la mayoría de las muestras del subconjunto.

Procedimiento de Notación Matemática de creación de C5.0 con Boosting:

Boosting es una de las mejoras más importantes en C5.0 en comparación con C4.5. El algoritmo de refuerzo establece el peso de cada muestra, lo que presenta su importancia. Cuanto mayor es el peso, mayor es la influencia de la muestra en el árbol de decisión. Inicialmente, cada muestra tiene el mismo peso. En cada ensayo, se construye un nuevo árbol de decisión. Se ajusta el peso de cada muestra, de modo que el aprendiz se centre en las muestras que están mal clasificadas por el árbol de decisión construido en la última prueba, lo que significa que estas muestras tendrán un mayor peso. (Pang and Gong 2009).

El procedimiento de refuerzo de C5.0 es similar al algoritmo AdaBoost descrito anteriormente en el sentido básico: los modelos se ajustan secuencialmente y cada iteración ajusta los pesos de los casos en función de la precisión de la predicción de una muestra. (Kuhn and Johnson 2013). Sin embargo, hay algunas diferencias:

C5.0 intenta mantener un tamaño de árbol similar al inicial. Esto se correlaciona con la cantidad de nodos terminales, que aumentan en número a medida que crece el árbol. (Jansson 2016).

C5.0 calcula las probabilidades de clase para todos los modelos potenciados y dentro de estos modelos, se calculan los promedios ponderados. Luego, a partir de estos modelos, C5.0 elige la clase que tiene la probabilidad máxima dentro del grupo.(Jansson 2016).

El modelo dejará de aumentar automáticamente si el modelo es muy efectivo (es decir, la sumatoria de los pesos para los datos mal clasificados es menor que 0.10) o si es altamente ineficaz (por ejemplo, el peso promedio de las muestras incorrectas es mayor al 50%) .(Kuhn and Johnson 2013).

Pseudocódigo que representa el procedimiento de ponderación en C5.0(Jansson 2016):

Procedimiento

N = tamaño del conjunto de entrenamiento.

N_- = número de muestras clasificadas incorrectamente.

T = número de iteraciones de boosting.

$W_{i,t}$ = Peso de la i -ésima muestra durante la t -ésima ronda de refuerzo.

S_+ = suma de pesos para muestras clasificadas correctamente.

S_- = suma de pesos para muestras clasificadas incorrectamente.

for $t \leftarrow 1$ hasta T **hacer:**

 Construye un árbol de decisiones.

for $i \leftarrow 1$ hasta N **hacer:**

$$midpoint \text{ (punto medio)} = \frac{1}{2} \left[\frac{1}{2} (S_+ + S_-) - S_- \right]$$

$$W_{i,t} = W_{i,t-1} \frac{S_+ - midpoint}{S_+}$$

\leftarrow peso si se clasifica correctamente

$$W_{i,t} = W_{i,t-1} + \frac{midpoint}{N_-}$$

\leftarrow peso si está mal clasificado

fin

fin

fin Procedimiento

Ejemplo de Árbol de decisión C5.0

El proceso de construir un árbol de decisión es bastante largo, por lo que solo se empleó 14 instancias y algunas variables de entrada, no todas, para este ejemplo, ya que la finalidad es demostrar el funcionamiento de C5.0. Las variables que se usaran son: Genero, Estado Civil, Edad, P.2015, P.2016, T.2015, T.2016 y una variable de salida (PAGO). La variable de salida tiene dos valores (SI y NO). Todo lo descrito anteriormente se aplicó con los datos reales de este estudio.

Respecto a los datos de la variable de salida, el ejemplo cuenta con 5 instancias con valor de NO y 9 instancias con valor de SI.

Los datos se detallan a continuación en la siguiente tabla:

Tabla 48
Datos empleados para ejemplo de C5.0.

Genero	Estado_Civil	Edad	P.2015	P.2016	T.2015	T.2016	PAGO
M	D	27	0	0	NO PAGO	NO PAGO	NO
M	D	28	18	18	PAGO	PAGO	SI
M	D	28	18	18	TIEMPO PAGO TIEMPO	TIEMPO PAGO DESTIEMPO	SI
M	D	29	0	18	NO PAGO	PAGO TIEMPO	SI
F	S	30	0	0	NO PAGO	NO PAGO	SI
M	D	30	0	0	NO PAGO	NO PAGO	SI
F	D	31	18	18	PAGO TIEMPO	PAGO TIEMPO	SI
M	S	31	0	18	NO PAGO	PAGO TIEMPO	NO
M	S	31	18	18	PAGO TIEMPO	PAGO DESTIEMPO	SI
M	D	32	18	0	PAGO TIEMPO	NO PAGO	SI
M	D	34	18	0	PAGO DESTIEMPO	NO PAGO	NO

F	D	34	18	18	PAGO DESTIEMP O	PAGO TIEMPO	NO
M	S	35	0	0	NO PAGO	NO PAGO	NO
M	S	35	18	18	PAGO TIEMPO	PAGO DESTIEMP O	SI

Fuente: Elaboracion Propia

Hallando al Nodo Padre.

Como se describió anteriormente el proceso empieza calculando la entropía general, para ello primero necesitamos saber cuantas instancias hay de la clase SI y NO de la variable de salida PAGO.

Tabla 49.

Cantidad de datos de la clase SI y NO para ejemplo C50

Clase NO:	5
Clase SI:	9
TOTAL:	14

Fuente: Elaboracion Propia

Ahora sí procedemos a calcular la entropía de la información general, dada por:

$$E(S) = - \sum_{i=1}^m p_i \log_2 p_i$$

Entonces:

$$E(S) = - \left(\frac{5}{14}\right) \log_2 \left(\frac{5}{14}\right) - \left(\frac{9}{14}\right) \log_2 \left(\frac{9}{14}\right)$$

$$E(S) = 0.940285959$$

Como segundo paso se halla la entropía de la información y la ganancia de cada atributo o variable predictora, cuya notación matemática es la siguiente respectivamente:

$$E(S|A) = - \sum_{j=1}^v p'_i \sum_{i=1}^m p_{ij} \log_2(p_{ij})$$

$$Gain(A) = E(S) - E(S|A)$$

Variable Genero:

Tabla 50.

Cantidad de datos de la clase SI y NO para ejemplo C50 de la variable Genero.

	M	F	Total
NO	4	1	5
SI	7	2	9
Total	11	3	14

Fuente: Elaboracion Propia

$$E_{(Genero)} = \left(\frac{11}{14}\right) \left(- \left(\frac{4}{11}\right) \log_2 \left(\frac{4}{11}\right) - \left(\frac{7}{11}\right) \log_2 \left(\frac{7}{11}\right) \right) + \left(\frac{3}{14}\right) \left(- \left(\frac{1}{3}\right) \log_2 \left(\frac{1}{3}\right) - \left(\frac{2}{3}\right) \log_2 \left(\frac{2}{3}\right) \right)$$

$$E_{(Genero)} = 0.939796489$$

$$Gain_{(Genero)} = 0.940285959 - 0.939796489$$

$$Gain_{(Genero)} = 0.000489469$$

Lo mismo se hace con las demás variables dando como resultado:

Variable Estado Civil:

$$E_{(Estado Civil)} = 0.937101106$$

$$Gain_{(Estado Civil)} = 0.003184853$$

Variable Edad:

En la variable edad, por ser un atributo continuo, es decir, hay muchos valores distintos, se calcula el umbral que no es mas que una división del conjunto de ejemplos en dos subconjuntos. Al primero pertenecen los ejemplos con valor del atributo menor o igual que el umbral y al segundo los ejemplos con valor mayor.(Christopher et al. 2003).

Y esta definido por:

$$U = \frac{a_i + a_{i+1}}{2}$$

Osea, se realiza la sumatoria del elemento a con su valor sucesivo y se divide entre 2 basándose en las reglas siguientes:

- a) Si el siguiente ejemplo está en la misma categoría que el ejemplo actual, el intervalo no se divide.
- b) Si el siguiente ejemplo tiene el mismo valor que el ejemplo actual, el intervalo no se divide

Entonces teniendo en cuenta eso procedemos a calcular el umbral con los siguientes datos de edad:

Tabla 51.
Datos de la variable Edad.

Edad	PAGO
27	NO
28	SI
28	SI
29	SI
30	SI
30	SI
31	SI
31	NO
31	SI
32	SI
34	NO
34	NO
35	NO
35	SI

Fuente: Elaboracion Propia

Se observa que los datos que cuya condición si cumplen son dos, 27-28 (cuyo umbral sería $(27 + 28)/2 = 27,5$) y 32-34 (cuyo umbral sería $(32 + 34)/2 = 33$). Entonces se procede a realizar el cálculo de la entropía y ganancia de ambos umbrales:

$$E_{(U_{27.5})} = 0.826885094$$

$$Gain_{(U_{27.5})} = 0.113400864$$

$$E_{(U_{33})} = 0.747456675$$

$$Gain_{(U_{33})} = 0.192829284$$

Ahora bien, se deberá escoger como representante de la Variable edad el Umbral con la mayor ganancia, la cual es el umbral de 33. Entonces:

$$E_{(Edad)} = 0.747456675$$

$$Gain_{(Edad)} = 0.192829284$$

Variable P.2015:

$$E_{(P.2015)} = 0.892158928$$

$$Gain_{(P.2015)} = 0.04812703$$

Variable P.2016:

$$E_{(P.2016)} = 0.892158928$$

$$Gain_{(P.2016)} = 0.04812703$$

Variable T.2015:

$$E_{(T.2015)} = 0.428571429$$

$$Gain_{(T.2015)} = 0.51171453$$

Variable T.2016:

$$E_{(T.2016)} = 0.775339498$$

$$Gain_{(T.2016)} = 0.164946461$$

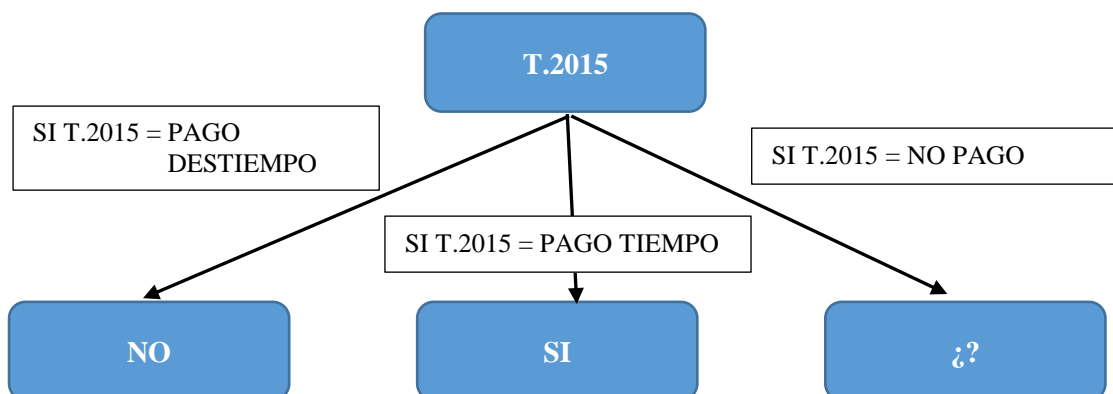
Una vez teniendo la ganancia de todas las variables:

Tabla 52.
Variables con sus ganancias respectivas.

VARIABLE	GANANCIA
Genero	0.000489469
Estado_Civil	0.003184853
Edad	0.192829284
P.2015	0.04812703
P.2016	0.048127030
T.2015	0.51171453
T.2016	0.164946461

Fuente: Elaboracion Propia

Se elige la que tiene una mayor ganancia como nodo padre, la cual es la variable T.2015, y teniendo en cuenta que tiene 3 niveles (PAGO DESTIEMPO, PAGO TIEMPO, NO PAGO) queda de la siguiente forma:



En la primera decisión se observa que los datos cuya variable T.2015 es igual a PAGO DESTIEMPO todos sus registros indican que NO pagará. Caso contrario en los datos donde T.2015 son igual PAGO TIEMPO todos los registros indican que SI pago. Como en ambos casos solo existe una clase en cada uno (NO y SI respectivamente), ya no se puede calcular la entropía y ganancia. Pero en el caso de los datos donde la Variable T.2015 es igual a NO PAGO, si existen dos clases, entonces en ese grupo de datos si se seguirá calculando para encontrar los nodos hijos.

Hallando al Nodo Hijo.

Se tienen como nuevos datos los siguientes:

Tabla 53.

Nuevos datos para calcular la entropía y la ganancia.

Genero	Estado_Civil	Edad	P-2015	P-2016	T-2016	PAGO
M	D	27	0	0	NO PAGO	NO
M	D	29	0	18	PAGO	SI
F	S	30	0	0	NO PAGO	SI
M	D	30	0	0	NO PAGO	SI
M	S	31	0	18	PAGO	NO
M	S	35	0	0	NO PAGO	NO

Fuente: Elaboracion Propia

Sabiendo que se tienen 3 instancias con valores de NO y 3 instancias con valores de SI en el atributo de salida PAGO, se calcula la entropía general del nuevo conjunto de datos:

$$E(S) = 1$$

Luego se calcula la entropía y la ganancia en cada variable:

Variable Género:

$$E_{(Genero)} = 0.809125495$$

$$Gain_{(Genero)} = 0.190874505$$

Variable Estado Civil:

$$E_{(Estado Civil)} = 0.918295834$$

$$Gain_{(Estado Civil)} = 0.081704166$$

Variable Edad:

Al igual que en el caso anterior, para la variable edad también se calculó los posibles Umbrales y se eligió el que tiene mayor ganancia, el cual fue el Umbral de 30.

$$E_{(Edad_{U_{30}})} = 0.54085208$$

$$Gain_{(Edad_{U_{30}})} = 0.459147917$$

Variable P.2015:

$$E_{(P.2015)} = 1$$

$$Gain_{(P.2015)} = 0$$

Variable P.2016:

$$E_{(P.2016)} = 1$$

$$Gain_{(P.2016)} = 0$$

Variable T.2016:

$$E_{(T.2016)} = 1$$

$$Gain_{(T.2016)} = 0$$

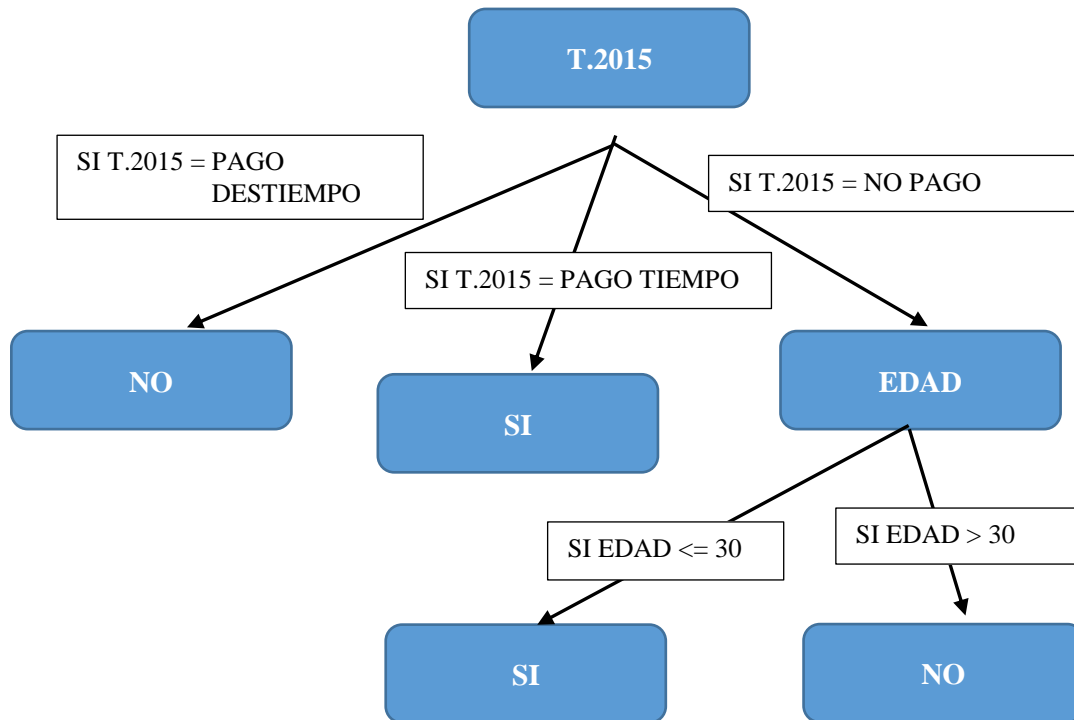
Una vez teniendo el resultado de las ganancias de cada variable, las cuales se muestran a continuación:

Tabla 54.
Nueva información de variables y sus ganancias.

VARIABLES	GANANCIAS
Genero	0.190874505
Estado_Civil	0.081704166
Edad	0.459147917
P.2015	0
P.2016	0
T.2016	0

Fuente: Elaboracion Propia

La variable Edad es la que tiene mayor ganancia por lo que se elige para la próxima división del árbol, quedando de la siguiente manera:



Prediccion:

Ahora para calcular que tan bien o mal esta entrenado el modelo, se procede a realizar un test con 10 instancias y se obtuvo los siguientes resultados:

Tabla 55.*Datos para test de predicción del árbol C5.0 y sus resultados (ESTADO DE PREDICCIÓN).*

Genero	Estado Civil	Edad	P.2015	P.2016	T.2015	T-2016	PAGO	REGLAS DEL ÁRBOL DE DECISION	PREDICCION	ESTADO PREDICCION
M	S	37	18	18	PAGO TIEMPO	PAGO TIEMPO	SI	T.2015 es igual a PAGO TIEMPO entonces: SI	SI	CORRECTO
F	S	29	18	18	PAGO DESTIEMPO	PAGO DESTIEMPO	SI	T.2015 es igual a PAGO DESTIEMPO entonces: NO	NO	INCORRECTO
M	S	49	0	0	NO PAGO	NO PAGO	NO	T.2015 es igual a NO PAGO entonces como EDAD es mayor que 30 entonces: NO	NO	CORRECTO
M	S	55	0	0	NO PAGO	NO PAGO	NO	T.2015 es igual a NO PAGO entonces como EDAD es mayor que	NO	CORRECTO

F	D	34	18	18	PAGO DESTIEMPO	PAGO DESTIEMPO	SI	30 entonces: NO T.2015 es igual a PAGO DESTIEMPO entonces: NO	NO	INCORRECTO
M	D	29	18	0	PAGO DESTIEMPO	NO PAGO	NO	T.2015 es igual a PAGO DESTIEMPO entonces: NO	NO	CORRECTO
F	S	36	0	0	NO PAGO	NO PAGO	NO	T.2015 es igual a NO PAGO entonces como EDAD es mayor que 30 entonces: NO	NO	CORRECTO
M	S	76	0	0	NO PAGO	NO PAGO	NO	T.2015 es igual a NO PAGO entonces como EDAD es mayor que 30 entonces: NO	NO	CORRECTO
M	S	57	18	18	PAGO DESTIEMPO	PAGO DESTIEMPO	SI	T.2015 es igual a PAGO DESTIEMPO	NO	INCORRECTO

M	D	50	18	18	PAGO DESTIEMPO	PAGO TIEMPO	SI	entonces: NO T.2015 es igual a PAGO DESTIEMPO entonces: NO	NO	INCORRECTO
---	---	----	----	----	-------------------	----------------	----	--	----	------------

Fuente: Elaboracion Propia

Como podemos observar hay 4 instancias clasificadas incorrectamente, para apreciarlo mejor se presenta en una matriz de confusión:

Tabla 56.

Matriz de confusión de resultados de Predicción de ejemplo de C5.0

	NO	SI
NO	5	4
SI	0	1

Fuente: Elaboracion Propia

Para comprobar que tan efectivo fue, se calcula el Accuracy:

$$Accuracy = \frac{5 + 1}{5 + 4 + 1}$$

$$Accuracy = 0.6$$

Los resultados de los nodos para el árbol de decisión y el accuracy, se comprueba cargando la misma data en R, y se puede apreciar que los valores coinciden perfectamente.

```
Decision tree:
```

```
T.2015 = PAGO DESTIEMPO: NO (2)
T.2015 = PAGO TIEMPO: SI (6)
T.2015 = NO PAGO:
: ...Edad <= 30: SI (4/1)
    Edad > 30: NO (2)
```

Figura 31. Resultado del árbol de decisión generado por R.

```
prediccion1 NO SI
NO 5 4
SI 0 1
```

```
Accuracy : 0.6
```

Figura 32. Resultado de matriz de confusión y el Accuracy generado por R.

Implementación:

Como se detalla en la *Figura 31* y la *Figura 32*, puede simplificar el proceso empleando la herramienta R, donde, según la documentación oficial de R (Max, Weston, and Culp 2018), menciona que se usa de la siguiente manera:

```
# method for class 'formula'  
C5.0(formula, data, weights, subset, na.action = na.pass,...)
```

Y sus argumentos son los siguientes:

Trials: un número entero que representa el número de repetición de refuerzo. Por ejemplo, el valor de “1” indica que se construirá un solo modelo.

Rules: Una lógica: ¿debería descomponerse el árbol en un modelo basado en reglas?

Weights: un vector numérico opcional de pesos de casos. Tenga en cuenta que los datos utilizados para los pesos de casos no se utilizarán como una variable de división en el modelo.

Control: una lista de parámetros de control.

Costs: Una matriz de costes asociada a los posibles errores. La matriz debe tener columnas y filas C, donde C es el número de niveles de clase.

...: otras opciones para pasar a la función (no se utiliza actualmente con el método predeterminado)

Formula: una fórmula, con una respuesta y al menos un predictor.

Data: un marco de datos opcional en el que interpretar las variables nombradas en la fórmula.

SubSet: expresión opcional que dice que solo se debe usar un subconjunto de las filas de los datos en el ajuste.

na.acción: una función que indica lo que debería ocurrir cuando los datos contienen NA. El valor predeterminado es incluir valores faltantes ya que el modelo puede acomodarlos.

Al igual que en el anterior algoritmo, se revisó publicaciones para saber cuál de todos los argumentos, u opciones al construir el modelo, será mejor para el presente caso de estudio. Así, según (Rodrigo 2017), en su publicación menciona que “C5.0 incorpora un método propio de boosting (semejante a AdaBoost) para generar modelos de tipo ensemble basados en árboles (...) Los resultados muestran que, mediante boosting, se mejora considerablemente la capacidad predictiva del modelo”. Esto se realiza a través del argumento *trials* y el considero 100 iteraciones.

En cambio, (Nestor 2016) en su estudio hizo una comparación de un árbol de clasificación C5.0 sin ninguna modificación, un árbol de clasificación C5.0 con boosting y un árbol de clasificación C5.0 con poda obteniendo así como buenos resultados el árbol de clasificación con poda con 94,25%, frente a un 94,1% del árbol de clasificación sin modificaciones y de un 94,1% del árbol de clasificación con poda. El autor menciona que en el caso de la Poda, Se utilizó el comando **C5.0Control(noGlobalPruning = TRUE)**, para apagar la poda global que puede ser beneficioso en algunas aplicaciones, particularmente cuando se generan conjuntos de reglas. Otra opción es el factor de confianza **CF** que afecta a la forma en que se estima que las tasas de error y por lo tanto la severidad de la poda; los valores menores que el valor predeterminado (25%) causa más del árbol inicial ser podados, mientras que los valores más grandes en menos de poda, para su caso se usó un CF=0.20.

Teniendo en cuenta los resultados obtenidos por los autores antes mencionados en sus estudios, se decidió, antes de aplicar el modelo a todos los meses para esta investigación, realizar también una comparación previa para saber si implementando boosting en C5.0 (Teniendo en cuenta los resultados de los estudios anteriores, se empleará para esta comparación previa dos tipos de iteraciones, una de 10 y otra 100) o implementando C5.0 con poda (Teniendo en cuenta los resultados obtenidos del segundo autor se considera *noGlobalPruning = TRUE* y *CF=0.20*) presenta mejores resultados

para este caso de estudio, tomando un mes(enero) para las pruebas, también se realizó una comparación de C5.0 uniendo Boosting con la Poda, obteniendo los siguientes resultados:

Tabla 57

Comparando C5.0 sin modificaciones, C5.0 con boosting (10 y 100 iteraciones) y C5.0 con Poda.

		Accuracy
C5.0 Sin Modificaciones		0.8648
C5.0 Con Boosting	trial =10	0.8648
	trial =100	0.8655
C5.0 Con Poda		0.8648
5.0 Con Boosting y Poda	trial =10	0.8655
	trial =100	0.8682

Fuente: Elaboracion Propia

Se observa en la Tabla 57, que el mejor Accuracy es cuando se utiliza Boosting (100 iteraciones) y Poda al mismo tiempo, por lo que se concluye que la mejor opción es utilizar C5.0 con Boosting y Poda, ya que presenta una leve mejora en la Exactitud.

Implementación:

```
library(C50)
datos <- read.csv("Enero_Algoritmo.csv",dec = ".")
set.seed(101)
tamano.total <- nrow(datos)
tamano.entreno <- round(tamano.total*0.7)
datos.indices <- sample(1:tamano.total , size=tamano.entreno)
datos.entreno <- datos[datos.indices,] #entrenamiento (el 70%)
datos.test <- datos[-datos.indices,] #test(el restante, osea el 30%)
```

Figura 33. *Cargando Datos y separando en Entrenamiento (70%) y Test (30%).*


```

# Ejecución de la técnica de Arbol de Decisión C5.0
modelo1 <- C5.0(PAGO ~ ., data = datos.entreno,
               trial = 100,
               control = C5.0Control(noGlobalPruning = TRUE,CF=0.20))

# test (Calculando Tiempo en segundos)
precision1 <- predict(modelo1,newdata=datos.test)

# Matriz de confusión
library(caret)
tabla1 <- table(precision1, datos.test$PAGO)
confusionMatrix(tabla1)

# % correctamente clasificados
100 * sum(diag(tabla1)) / sum(tabla1)

```

Figura 34. Creando el Modelo de C5.0, Evaluandolo con el Test y Generando Matriz de Confusion

```

#-----AUC - CURVA ROC
library(caTools)
colAUC(as.numeric(as.factor(precision1)),as.factor(datos.test$PAGO),plotROC = TRUE)

```

Figura 35. Aplicando AUC - Curva Roc

El código completo de Decision Tree utilizado se encuentra en el **Anexo 9**.

3.3.5. Evaluar los resultados Obtenidos.

En esta fase se realizó las pruebas para evaluar la eficiencia de las técnicas Adaboost y Decision Tree, los algoritmos se le evaluó en base a 5 indicadores que fueron Precision (Accuracy en inglés), Exactitud (Precision en inglés), sensibilidad, especificidad y el AUC (Área Bajo la Curva).

Rendimiento y Eficiencia

Para evaluar la eficiencia de las técnicas se utilizó la herramienta llamada Matriz de Confusión que “contiene información sobre clasificaciones reales y pronosticadas realizadas por un sistema de clasificación. El rendimiento de tales sistemas se evalúa comúnmente utilizando los datos de la matriz.” (Kohavi y Provost, 1998).

La matriz de confusión es representada de la siguiente forma:

Tabla 58*Representación de una matriz de confusión según Kohavi & Provost.*

	SI	No
SI	Verdadero Positivo (VP)	Falso Negativo (FN)
No	Falso Positivo (FP)	Verdadero Negativo (VN)

Fuente: (Kohavi & Provost, 1998)

Las entradas en la matriz de confusión, tienen el siguiente significado:

Tabla 59*Interpretación de los valores generado por la matriz de confusión.*

SIGLA Y DESCRIPCION
VP es la cantidad de <i>positivos</i> que fueron <i>clasificados correctamente</i> como positivos por el modelo.
VN es la cantidad de <i>negativos</i> que fueron <i>clasificados correctamente</i> como negativos por el modelo.
FN es la cantidad de <i>positivos</i> que fueron <i>clasificados incorrectamente</i> como negativos.
FP es la cantidad de <i>negativos</i> que fueron <i>clasificados incorrectamente</i> como positivos.

Fuente: Según (Zelada 2017).

Basado en la *Tabla 59* se consideraron diferentes medidas a ser evaluadas para conocer su eficiencia en el desempeño de ambas técnicas, estas son:

Precisión

En inglés Accuracy (Recuero 2018), es la cantidad general de data clasificada correctamente. Y se representa bajo la formula:

$$Exactitud = \frac{VP + VN}{Total}$$

Exactitud.

En inglés Precision (Recuero 2018), representa la cantidad de data que clasifica correctamente cuando clasifica positivos. Se representa bajo la siguiente fórmula:

$$\text{Precisión} = \frac{VP}{\text{Total clasificados positivos}}$$

Se debe tener en cuenta que en inglés se emplea el término “**precision**” para hacer referencia a la **exactitud** y el término en español **precisión** hace referencia a “**accuracy**”. Esto puede ser algo confuso y llevar a un error de interpretación. (Recuero 2018).

Sensibilidad.

Tabién llamada Recall o Exhaustividad. Responde a la pregunta: Cuando es positiva la clase que se analiza, ¿cuál es el porcentaje logra clasificar? Se representa bajo la siguiente formula:

$$\text{Sensibilidad} = \frac{VP}{\text{Total Positivos}}$$

Especificidad.

Responde a la pregunta: Cuando la clase que se analiza es negativa, ¿cuál es el porcentaje que logra clasificar? Se representa bajo la siguiente formula:

$$\text{Especificidad} = \frac{VN}{\text{Total Negativos}}$$

AUC.

El Área bajo la curva ROC (por sus siglas en inglés: AUC) representa los pares precisión-cobertura, codificados como ratios de verdaderos positivos y falsos positivos, según vamos variando la configuración. Una curva más alta indicará mejores valores de precisión a un mismo valor de cobertura.(Huerta 2015). Básicamente su formula es:

$$AUC = \frac{1}{2}(\text{Especificidad} + \text{Sensibilidad})$$

El AUC esta comprendido dentro del valor de 0 a 1. Un modelo con una tasa de error de predicción del 100% tiene un AUC de 0.0; otras predicciones con un AUC correcto del 100% es 1.0.

La herramienta R ofrece tanto el Tiempo de cada proceso, para medir la variable Dependiente (Tiempo para generar estimación de morosidad), como la matriz de confusión, para medir la variable Independiente (Precisión, Exactitud, Sensibilidad, Especificidad, AUC).

Estos resultados que muestra la herramienta R, es decir, los indicadores ya calculados, se tomaron en cuenta para guardar dicha información en Fichas de Observación como la de la *Figura 36*.

Ficha de observación para Adaboost	
Numero de Análisis:	1
Mes Analizado:	Enero
Tiempo de Construcción del Modelo (seg):	43.68
Tiempo de realización de la predicción (seg):	2.48

Matriz de confusión		
	NO	SI
NO	1046	185
SI	251	1818

Evaluación Precisión (Accuracy) (%)	
Porcentaje Dividido	86.79

Otras Evaluaciones (%)	
Exactitud (Precision)	87.87
Sensibilidad	80.65
Especificidad	90.76
AUC	0.857

Figura 36. Ficha de observación del rendimiento del algoritmo de Adaboost del mes de enero. Fuente: Elaboración propia

La *Figura 36* muestra el análisis realizado al algoritmo Adaboost en el mes de enero, ese análisis se realizó con un tipo de prueba denominado Percentage Split (Porcentaje Dividido), esta prueba consiste en separar el conjunto de datos para entrenamiento y dividirlo en dos partes, la primera para construir el

clasificador y la segunda para hacer la prueba. El porcentaje para entrenamiento es 70% y el de test es 30%.

Las demás Fichas de Observación se encuentran en el Anexo 3.2. **(Ver Anexo 3.2).**

Según el Algoritmo, y aplicándolo a los diferentes meses, el resultado del análisis de los algoritmos presentó pequeñas variaciones, éstas se detallaron con gráficos y tablas en el punto 3.1. del presente Capítulo III.

IV. CONCLUSIONES Y RECOMENDACIONES

4.1. Conclusiones

- a) Las técnicas de Decision Tree y Adaboost son algunas de las principales técnicas de predicción más utilizadas, con más documentación y con resultados muy buenos, esto se afirma después de revisar la literatura.
- b) Dentro de la literatura revisada de estudios parecidos al presente caso de investigación, se determinó que el Arbol de Decisión más adecuado es el árbol c5.0, por ser más actual y mejorado y es el que se usará en esta investigación.
- c) En cuanto a la variante de Adaboost, en lo que se revisó, el más adecuado para este estudio es Adaboost.M1 ya que es muy bueno prediciendo variables binarias como salida, y cumple con este caso de estudio en particular ya que se desea predecir si determinados colegiados pagaran o no (binario).
- d) Según los artículos científicos revisados, y otras literaturas, las variables de predicción más importantes, es decir indispensables, para este tipo de trabajos de investigación son Edad, Monto Pagado, Tiempo en que se efectuó el pago, Especialidad, son las variables más representativas y se tiene que contar con esas éstas para poder realizar predicción de morosidad.
- e) El proceso de ETL (extracción, transformación y carga) hace posible extraer datos necesarios muy importantes para las variables predictoras

identificadas como son la Edad, Género, Estado Civil, Categoría, Especialidad, Pago por meses y Tiempo en que se efectuó el pago. Se concluye que este proceso da como resultado datos ordenados, coherentes y concisos, solo si el proceso de ETL se realiza correctamente.

- f) El modelo Dimensional mas Adecuado es el Modelo de Estrella, se afirma esto ya que se realizo una ponderación bajo la escala de Likert y ganó este Modelo, se tuvo en cuenta factores como, por ejemplo: facilidad de uso, rendimiento, tabla de dimensiones, número de join, etc.
- g) La mejor combinación de argumentos en R para AdaBoost según la prueba realizada es *coflearn* = "Freund" y *boos* = FALSE ya que obtiene mejores resultados de Accuracy.
- h) En Decisión Tree, específicamente C5.0, posee un método de boosting semejante a Adaboost, por lo que, al realizar las pruebas previas correspondientes, empleando boosting se obtuvieron mejores resultados, y también posee un método de poda, que al combinarlo hay cierta mejora en crear el árbol de decisión, y por lo tanto ambos métodos se empleará para este caso de estudio.
- i) Haciendo uso de la herramienta R donde fueron Implementados ambas técnicas, se llegó a la conclusión de que Decisión Tree ganó por poco a Adaboost con 86.73% contra 86.71% en Accuracy.

4.2. Recomendaciones:

- a) Se recomienda no asumir que una técnica es mejor que otra en todos los casos, ya que como se demostró, va a depender mucho de las variables predictivas, los tipos de datos, la cantidad de datos y los argumentos que utilicen dichas técnicas.
- b) Se recomienda emplear el árbol de decisión C5.0 ya que utilizando el método de boosting y la poda incorporada, mejora de gran medida su predicción.
- c) Se recomienda probar las mismas técnicas de este estudio en otra herramienta diferente a R, por ejemplo, Python o Weka, esto con la finalidad de comparar cuál herramienta muestra mejores resultados para el presente caso de estudio.
- d) Se recomienda usar el historial de pago de colegiados desde años donde presente mayor cantidad de información y no donde esten presente gran cantidad de campos vacíos.

V. REFERENCIAS

- Alaka, Hafiz A., Lukumon O. Oyedele, Hakeem A. Owolabi, Vikas Kumar, Saheed O. Ajayi, Olugbenga O. Akinade, and Muhammad Bilal. 2018. "Systematic Review of Bankruptcy Prediction Models: Towards a Framework for Tool Selection." *Systematic Review of Bankruptcy Prediction Models Towards a Framework for Tool Selection* 94:164–84.
- Alfaro, Esteban. n.d. "Boosting Function | R Documentation." Retrieved December 11, 2018 (<https://www.rdocumentation.org/packages/adabag/versions/4.2/topics/boosting>).
- Alfaro, Esteban, and David Elizondo. 2008. "Bankruptcy Forecasting : An Empirical Comparison of AdaBoost and Neural Networks." 45:110–22.
- Alfaro, Esteban, Matias Gamez, and Noelia Garcia. 2018. "Package ' Adabag .'"
- Alfaro, Esteban, Matías Gámez, and Noelia García. 2013. "**Adabag** : An R Package for Classification with Boosting and Bagging." *Journal of Statistical Software* 54(2).
- Aziz, M. Adnan, and Humayon A. Dar. 2006. "Predicting Corporate Bankruptcy: Where We Stand?" *Corporate Governance* 6(1):18–33.
- Bae, Jae Kwon. 2012. "Predicting Financial Distress of the South Korean Manufacturing Industries." *Expert Systems with Applications* 39(10):9159–65.
- Barboza, Flavio, Herbert Kimura, and Edward Altman. 2017. "Machine Learning Models and Bankruptcy Prediction." *Expert Systems with Applications* 83:405–17.
- Betancour, Gustavo. 2005. "Las Máquinas de Soporte Vectorial (SVMs)." *Scientia Et Technica* (27):67–72.
- Borisov, Alexander, Igor Chikalov, Victor Eruhimov, and Eugene Tuv. 2005. "Performance and Scalability Analysis of Tree-Based Models in Large-Scale Data-Mining Problems." *Understanding the Platform Requirements of Emerging Enterprise Solutions* 9(2):143–51.
- Brachfield, Pere. 2013. "El Problema de La Morosidad Empresarial - BRACHFIELD Credit & Risk Consultants." Retrieved April 25, 2018 (<http://perebrachfield.com/blog/morosidad/el-problema-de-la-morosidad->

- empresarial/).
- Brownlee, Jason. 2016. *Master Machine Learning Algorithms Discover How They Work and Implement Them From Scratch*. v1.1.
- Brownlee, Jason. 2013. "A Tour of Machine Learning Algorithms." *Knowledge and Information Systems*. Retrieved October 12, 2018 (<http://link.springer.com/10.1007/s10115-007-0114-2>).
- Brownlee, Jason. 2014. "Non-Linear Classification in R with Decision Trees." Retrieved November 15, 2018 (<https://machinelearningmastery.com/non-linear-classification-in-r-with-decision-trees/>).
- Calvo, Diego. 2016a. "Árboles de Clasificación En R - Diego Calvo." Retrieved November 15, 2018 (<http://www.diegocalvo.es/arboles-de-clasificacion-en-r/>).
- Calvo, Diego. 2016b. "Cálculo de La Ganancia de Información En R." Retrieved July 27, 2020 (<https://www.diegocalvo.es/calculo-de-la-ganancia-de-informacion-en-r/>).
- Campos y Cobarrubias, G., and Nallely Emma Lule Martinez. 2012. "La Observación, Un Método Para El Estudio de La Realidad." *Revista Xihmai*, 45–60.
- Castro, Teresa de Jesús. 2015. "Inteligencia de Negocios, Una Introducción - GestioPolis." Retrieved April 25, 2018 (<https://www.gestipolis.com/inteligencia-de-negocios-una-introduccion/>).
- Cervan, Dheybi. 2020. "Ventajas de Usar El Software R." Retrieved July 27, 2020 (<https://dheybicervan.com/ventajas-de-usar-el-software-r/>).
- Chatterjee, Sourav. 2016. "Package 'FastAdaboost'."
- Christopher, Exposito, Exposito Airam, Lopez Israel, Melian Belen, and Marcos Moreno. 2003. "Arboles de Decision." 121–33.
- Colebatch, H. J. H., and L. A. Engel. 1974. "An Experimental Comparison of Classification Techniques in Debt Recoveries Scoring: Evidence from South Africa's Unsecured Lending Market." *Journal of Applied Physiology* 37(6):798–805.
- Delen, Dursun, Gregory Moscato, and Inga Linda Toma. 2018. "The Impact of Real-Time Business Intelligence and Advanced Analytics on the Behaviour of Business Decision Makers." 49–53.
- Díaz, Ángel Arnulfo. 2016. "TÉCNICAS DE MINERÍA DE DATOS PARA

PREDICCIÓN DEL DIAGNÓSTICO DE HIPERTENSIÓN ARTERIAL.”

- DIR&GE. 2017. “Del Análisis Del Pasado a La Predicción Del Futuro: 3 Rasgos de La Inteligencia de Negocio Predictiva.” Retrieved April 28, 2018 (<https://directivosygerentes.es/innovacion/transformacion-digital/td-innovacion/del-analisis-del-pasado-la-prediccion-del-futuro-4-rasgos-la-inteligencia-negocio-predictiva>).
- Gartner. 2018. “Mastering the New Business Executive Job of the CIO.”
- Garzón, María Jesús. 2015. “La Importancia de La Información En La Empresa | La Nueva Ruta Del Empleo.” Retrieved April 28, 2018 (<http://www.lanuevarutadelempleo.com/node/6497>).
- Gordini, Niccolò. 2014. “A Genetic Algorithm Approach for SMEs Bankruptcy Prediction: Empirical Evidence from Italy.” *Expert Systems with Applications* 41(14):6433–45.
- Heo, Junyoung, and Jin Yong Yang. 2014. “AdaBoost Based Bankruptcy Forecasting of Korean Construction Companies.” *Applied Soft Computing Journal* 24:494–99.
- Huerta, Sergio. 2015. “Medidas de Calidad de Algoritmos de Clasificación - S3lab.” Retrieved December 7, 2018 (<http://s3lab.deusto.es/medidas-calidad-algoritmos-clasificacion/>).
- Jansson, Johan. 2016. “Decision Tree Classification Od Products Using C5.0 and Prediction of Workload Using Time Series Analysis.” (2016:098).
- Jara, Jefferson, Diego Giral, and Fernando Martínez. 2016. “Implementation of Algorithms Based on Support Vector Machine (SVM) for Electric Systems: Topic Review.” *Tecnura* 20(48):149–70.
- Jiménez, Luis Alfredo Fernández. 2007. *Modelos Avanzados Para La Predicción a Corto Plazo de La Producción Eléctrica En Parques Eólicos*.
- Jones, Stewart, David Johnstone, and Roy Wilson. 2015. “An Empirical Evaluation of the Performance of Binary Classifiers in the Prediction of Credit Ratings Changes.” *Journal of Banking and Finance* 56:72–85.
- Kaur, Arvinder, and Inderpreet Kaur. 2016. “An Empirical Evaluation of Classification Algorithms for Fault Prediction in Open Source Projects.” *JOURNAL OF KING SAUD UNIVERSITY - COMPUTER AND INFORMATION SCIENCES*.

- Kim, Sung Kean, Tae Keun Yoo, Ein Oh, and Deok Won Kim. 2013. "Osteoporosis Risk Prediction Using Machine Learning and Conventional Methods." 188–91.
- Kuhn, Max, and Kjell Johnson. 2013. *Applied Predictive Modeling with Applications in R*.
- Le, Ph. D. 2013. "Un Breve Ejemplo Para Adaboost | Big Data Knowledge Sharing." Retrieved November 15, 2018 (<https://qizeresearch.wordpress.com/2013/12/05/short-example-for-adaboost/>).
- Lee, Yong-ju, and Okgee Min. 2017. "Comparative Analysis of Machine Learning Algorithms to Urban Traffic Prediction." 1034–36.
- Li, Yachao, and Yufa Wang. 2018. "Machine Learning Methods of Bankruptcy Prediction Using Accounting Ratios." *Open Journal of Business and Management* 06(01):1–20.
- López, Raúl. 2017. "Boosting En Machine Learning Con Python." Retrieved November 15, 2018 (<https://relopezbriega.github.io/blog/2017/06/10/boosting-en-machine-learning-con-python/>).
- Luis, Orellana. 2019. "Análisis de La Base de Datos ZOO Por Medio de Random Forest." 0–44.
- Marqués, A. I., V. García, and J. S. Sánchez. 2012. "Exploring the Behaviour of Base Classifiers in Credit Scoring Ensembles." *Expert Systems with Applications* 39(11):10244–50.
- Mathew, Vimala, Tom Toby, Vikram Singh, B. Maheswara Rao, and M. Goutham Kumar. 2017. "Prediction of Remaining Useful Lifetime (RUL) of Turbofan Engine Using Machine Learning." (Iccs):306–11.
- MAThWorks. 2018. "AdaBoost - MATLAB y Simulink." Retrieved November 9, 2018 (<https://www.mathworks.com/discovery/adaboost.html>).
- Max, Author, Steve Weston, and Mark Culp. 2018. "Package ' C50 .'." 1–13.
- Montalvo, Irene. 2016. "Análisis Comparativo De Técnicas De Minería De Datos Para La Predicción De Ventas." 133.
- Mselmi, Nada, Amine Lahiani, and Taher Hamza. 2017. "Financial Distress Prediction: The Case of French Small and Medium-Sized Firms." *International Review of Financial Analysis* 50:67–80.
- Nagalla, Rakshita, Prasanna Pothuganti, and Digvijay S. Pawar. 2017. "Analyzing

- Gap Acceptance Behavior at Unsignalized Intersections Using Support Vector Machines, Decision Tree and Random Forests.” *Procedia Computer Science* 109(2016):474–81.
- Nagaraj, Kalyan, and Amulyashree Sridhar. 2015. “A PREDICTIVE SYSTEM FOR DETECTION OF BANKRUPTCY USING MACHINE LEARNING.”
- Nestor, Val. 2016. “Clasificación Del Nivel de Pobreza En La Encuesta ENAHO Aplicando Regresión Logística y Árboles de Clasificación Con El Algoritmo C5.0.” *Journal of Petrology* 369(1):1689–99.
- Pandya, Rutvija, Jayati Pandya, K. P. Dholakiya, and Infotech Amreli. 2015. “C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning.” *International Journal of Computer Applications* 117(16):975–8887.
- Pang, Su-lin, and Ji-zhang Gong. 2009. “C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks.” *Systems Engineering - Theory & Practice* 29(12):94–104.
- Portugal, Ivens, Paulo Alencar, and Donald Cowan. 2017. “The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review.” *Expert Systems with Applications* 97:31.
- Quantic. 2014. “Modelos Predictivos: De La Videncia Al Business Intelligence - Quantic | Blog.” Retrieved April 28, 2018 (<https://www.quanticsolutions.es/blog/modelos-predictivos-videncia-business-intelligence/>).
- Rathor, Abhinav. 2017. “A Review at Machine Learning Algorithms Targeting Big Data Challenges.” 753–59.
- Recuero, Paloma. 2018. “Machine Learning a Tu Alcance: La Matriz de Confusión.” Retrieved October 12, 2018 (<https://data-speaks.luca-d3.com/2018/01/ML-a-tu-alcance-matriz-confusion.html>).
- Reyes, Laura. 2017. “APLICACIÓN DEL ALGORITMO ADABOOST.RT PARA LA PREDICCIÓN DEL INDICE COLCAP Y EL DISEÑO DE UN CONTROLADOR NO LINEAL.” Universidad Distrital Francisco José de Caldas.
- Rodrigo, Joaquín Amat. 2017. “Árboles de Predicción : Bagging , Random Forest , Boosting y C5 . 0.”
- Sánchez, Ricardo. 2003. “Inteligencia de Negocios o Business Intelligence BI.” Retrieved April 28, 2018 (<https://www.gestiopolis.com/inteligencia-de->

- negocios-o-business-intelligence-bi/).
- Soltero, Francisco, and Diego Bodas. 2005. "Clasificadores Inductivos Para El Posicionamiento Web. (Spanish)." *El Profesional de La Información* 14(1):4–13.
- Tang, Lingxiao, Fei Cai, and Yao Ouyang. 2018. "Applying a Nonparametric Random Forest Algorithm to Assess the Credit Risk of the Energy Industry in China." *Technological Forecasting and Social Change* (March):0–1.
- The Data Warehouse Institute. 2016. "Inteligencia de Negocio." Retrieved April 28, 2018 (<http://sci2s.ugr.es/graduateCourses/in>).
- Trujillo, D. 2017. "Aplicación de Metodologías Machine Learning En La Gestión de Riesgo de Crédito."
- Twala, Bhekisipho. 2010. "Multiple Classifier Application to Credit Risk Assessment." *Expert Systems with Applications* 37(4):3326–36.
- Vicentin, Jimena. 2016. "PREDICCIÓN DE PRECIOS MEDIANTE MODELIZACIÓN MULTIVARIADA DE SERIES DE TIEMPO. UNA APLICACIÓN AL SECTOR LÁCTEO ARGENTINO."
- Villanueva, Leslie. 2015. "La Toma de Decisiones En La Organización y El Gran Valor Del Profesional de La Información En Su Desarrollo – IFT." Retrieved April 28, 2018 (<http://www.infotecarios.com/la-toma-de-decisiones-en-la-organizacion-y-el-gran-valor-del-profesional-de-la-informacion-en-su-desarrollo/>).
- Wei, Yixuan, Xingxing Zhang, Yong Shi, Liang Xia, Song Pan, Jinshun Wu, and Mengjie Han. 2018. "A Review of Data-Driven Approaches for Prediction and Classification of Building Energy Consumption." *Renewable and Sustainable Energy Reviews* 82(September 2017):1027–47.
- Yeh, I. Cheng, and Che hui Lien. 2009. "The Comparisons of Data Mining Techniques for the Predictive Accuracy of Probability of Default of Credit Card Clients." *Expert Systems with Applications* 36(2 PART 1):2473–80.
- Yoon, Jong Sik, and Young S. Kwon. 2010. "A Practical Approach to Bankruptcy Prediction for Small Businesses: Substituting the Unavailable Financial Data for Credit Card Sales Information." *Expert Systems with Applications* 37(5):3624–29.
- Zelada, Carlos. 2017. "RPubs - Matriz de Confusión - Evaluación de Modelos de

- Predicción.” Retrieved December 7, 2018 (<https://rpubs.com/chzelada/275494>).
- Zhang, Haoting, Hongliang He, and Wenyu Zhang. 2018. “Classifier Selection and Clustering with Fuzzy Assignment in Ensemble Model for Credit Scoring.” *Neurocomputing* 316:210–21.
- Stackowiak, R., Rayman, J., Greenwald, R. (2007). Oracle Data Warehousing and Business Intelligence Solutions. Recuperado de: <http://download1.libgen.io/ads.php?md5=E9A7C68335CD256BF91741536B061850>
- Data Warehouse Institute. (2017). Oracle Data Warehousing and Business Intelligence Solutions. Recuperado de: <https://tdwi.org/events/onsite-education/onsite/sessions/business-intelligence/ppm-all-tdwi-business-intelligence-and-analytics-principles-and-practices.aspx>
- Yglesias, R. (2008). Inteligencia de negocio operacional. Recuperado de: <http://www.oracle.com/technetwork/es/documentation/317539-esa.pdf>
- Mende, T., Koschke, R., Peleska, J. (2011). On the Utility of a Defect Prediction Model during HW/SW Integration Testing: A Retrospective Case Study. Recuperado de: <http://sci-hub.tw/https://ieeexplore.ieee.org/document/5741333/>
- Solano, A. (2003). Toma de decisiones gerenciales. Recuperado de: <https://dialnet.unirioja.es/descarga/articulo/4835719.pdf>
- Gutiérrez, A. (2004). Toma de decisiones. Recuperado de: <http://cursos.aiu.edu/Toma%20de%20Decisiones/PDF/Tema%201.pdf>

- More, G. (2013). Métodos y técnicas de la investigación científica. Recuperado de:
http://biblioteca.ucv.cl/site/servicios/documentos/metodologias_investigacion.pdf
- Latorre, M., Seco del pozo, C. (2013). Estrategias y técnicas metodológicas. Recuperado de: <http://www.umch.edu.pe/arch/hnomarino/metodo.pdf>
- Astray, G., Castillo, J., Ferreiro, J., Gálvez y Mejuto. J. (2010). Artificial neural networks: a promising tool to evaluate the authenticity of wine. Recuperado de:
<https://www.tandfonline.com/doi/abs/10.1080/19476330903335277#aHR0cHM6Ly93d3cudGFuZGZvbmxpbmUuY29tL2RvaS9wZGYvMTAuMTA4MC8xOTQ3NjMzMzNTI3Nz9uZWVhbnRlZXRyZdWVAQEAw>
- Martinez, P. (2016). Predicción científica y prescripción en mejora genética vegetal en cuanto ciencia aplicada de diseño: El caso de la mejora de frutales del género Prunus. Recuperado de:
https://revistas.unal.edu.co/index.php/acta_agronomica/article/view/52329/61221
- Gonzales, M. (2009). Técnicas de predicción económica. Recuperado de:
<https://addi.ehu.es/bitstream/handle/10810/12493/05-09pil.pdf?sequence=1>
- Gonzales, J., Gómez, R. (2014). La morosidad; un acuciante problema financiero de nuestros días. Recuperado de:
http://www.aeca1.org/pub/on_line/comunicaciones_xvcongresoaecca/cd/111b.pdf
- Vallcorba, M., Delgado, J. (2007). Determinantes de la morosidad bancaria en una economía dolarizada. el caso uruguayo. Recuperado de:
<https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/07/Fic/dt0722.pdf>

- Arredondo, T. (2009). Introducción a los Técnicas Genéticos. Recuperado de:
<http://profesores.elo.utfsm.cl/~tarredondo/info/softcomp/Introduccion%20a%20los%20Técnicas%20geneticos.pdf>
- Ferrero, R. (2018). ¿CÓMO SELECCIONAR LAS VARIABLES ADECUADAS PARA TU MODELO? España: Data Science. Recuperado de:
<https://www.maximaformacion.es/blog-dat/como-seleccionar-las-variables-adequadas-para-tu-modelo/>
- Calvo, D. (2016). Análisis Adaboost en R. España. Recuperado de:
<http://www.diegocalvo.es/analisis-adaboost-en-r/>
- Aenor (1994). Norma UNE 50-121-91. Documentación. Métodos para el análisis de documentos, determinación de su contenido y selección de los términos de indización. En: Documentación. Recopilación de normas UNE.
- Bueno, E. (2003). La Investigación científica: Teoría y Metodología. Mexico. Consultado en: <http://www.postgradoune.edu.pe/pdf/documentos-academicos/ciencias-de-la-educacion/13.pdf>

VI. ANEXOS

ANEXO 01: Resolución de Aprobación del trabajo de Investigación.

FACULTAD DE INGENIERÍA, ARQUITECTURA Y URBANISMO
RESOLUCIÓN N° 0067-2019/FIAU-USS
Pimentel, 3 de agosto de 2020

VISTO:

El Acta de reunión N°1207-2020, de fecha 12 de julio de 2020 del Comité de investigación de la Escuela profesional de INGENIERÍA DE SISTEMAS, para la ejecución de la Tesis: "**ANÁLISIS COMPARATIVO DE ADAPTIVE BOOSTING Y DECISION TREE PARA PREDECIR LA MOROSIDAD DE CUOTAS SOCIALES DEL COLEGIO DE INGENIEROS DEL PERÚ CONSEJO DEPARTAMENTAL LAMBAYEQUE**", presentado por **SILVA PARRAGUEZ MAXIMO GABRIEL**, del Programa de estudios INGENIERÍA DE SISTEMAS, y;

CONSIDERANDO:

Que, de conformidad con la Ley Universitaria N° 30220 en su artículo 48º que a letra dice: "La investigación constituye una función esencial y obligatoria de la universidad, que la fomenta y realiza, respondiendo a través de la producción de conocimiento y desarrollo de tecnologías a las necesidades de la sociedad, con especial énfasis en la realidad nacional. Los docentes, estudiantes y graduados participan en la actividad investigadora en su propia institución o en redes de investigación nacional o internacional, creadas por las instituciones universitarias públicas o privadas.";

Que, de conformidad con el Reglamento de grados y títulos en su artículo 21º señala: "Los temas de trabajo de investigación, trabajo académico y tesis son aprobados por el Comité de Investigación y derivados a la facultad o Escuela de Posgrado, según corresponda, para la emisión de la resolución respectiva. El periodo de vigencia de los mismos será de dos años, a partir de su aprobación. En caso un tema perdiera vigencia, el Comité de Investigación evaluará la ampliación de la misma."

Que, de conformidad con el Reglamento de grados y títulos en su artículo 24º señala: La tesis es un estudio que debe denotar rigurosidad metodológica, originalidad, relevancia social, utilidad teórica y/o práctica en el ámbito de la escuela profesional. Para el grado de doctor se requiere una tesis de máxima rigurosidad académica y de carácter original. Es individual para la obtención de un grado; es individual o en pares para obtener un título profesional. Asimismo, en su artículo 25º señala: "El tema debe responder a alguna de las líneas de investigación institucionales de la USS S.A.C."

Que, en el Acta de reunión N°1207-2020 de fecha 12 de julio de 2020, del Comité de investigación de la Escuela profesional de INGENIERÍA DE SISTEMAS, se indica entre los acuerdos la aprobación del tema de la Tesis denominado "**ANÁLISIS COMPARATIVO DE ADAPTIVE BOOSTING Y DECISION TREE PARA PREDECIR LA MOROSIDAD DE CUOTAS SOCIALES DEL COLEGIO DE INGENIEROS DEL PERÚ CONSEJO DEPARTAMENTAL LAMBAYEQUE**" de la línea de investigación de INFRAESTRUCTURA, TECNOLOGÍA Y MEDIO AMBIENTE, a cargo de **SILVA PARRAGUEZ MAXIMO GABRIEL** en condición de egresado, del Programa de estudios INGENIERÍA DE SISTEMAS.

Estando a lo expuesto, y en uso de las atribuciones conferidas y de conformidad con las normas y reglamentos vigentes;

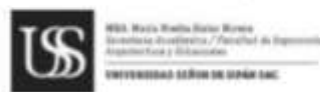
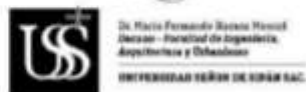
SE RESUELVE:

ARTÍCULO 1°: APROBAR, el tema del Tesis denominado "**ANÁLISIS COMPARATIVO DE ADAPTIVE BOOSTING Y DECISION TREE PARA PREDECIR LA MOROSIDAD DE CUOTAS SOCIALES DEL COLEGIO DE INGENIEROS DEL PERÚ CONSEJO DEPARTAMENTAL LAMBAYEQUE**", perteneciente a la línea de investigación de INFRAESTRUCTURA, TECNOLOGÍA Y MEDIO AMBIENTE, a cargo de **SILVA PARRAGUEZ MAXIMO GABRIEL**, del Programa de estudios INGENIERÍA DE SISTEMAS.

ARTÍCULO 2°: ESTABLECER, que la inscripción del Título del Tesis se realice a partir de emitida la presente resolución y tendrá una vigencia de dos (02) años.

ARTÍCULO 3°: DEJAR SIN EFECTO, toda Resolución emitida por la Facultad que se oponga a la presente Resolución.

REGÍSTRESE, COMUNÍQUESE Y ARCHÍVESE



Cc: Interesado, Archivo

ANEXO 02: Carta de aceptación de la institución para la recolección de datos.

AUTORIZACIÓN PARA EL RECOJO DE INFORMACIÓN

Chiclayo, abril del 2018

Quien suscribe:

Sr.

ING. DENNY JOHN FUENTES ADRIANZEN

AUTORIZA: Permiso para recojo de información pertinente en función del proyecto de investigación, denominado: "ANÁLISIS COMPARATIVO DE ADAPTIVE BOOSTING Y DECISION TREE PARA PREDECIR LA MOROSIDAD DE CUOTAS SOCIALES DEL COLEGIO DE INGENIEROS DEL PERU CONSEJO DEPARTAMENTAL LAMBAYEQUE"

Por el presente, el que suscribe, señor ING. DENNY JOHN FUENTES ADRIANZEN, presidente del CAPÍTULO DE ING. INDUSTRIAL Y DE SISTEMAS DEL COLEGIO DE INGENIEROS DEL PERU, CONSEJO DEPARTAMENTAL DE LAMBAYEQUE, AUTORIZO al alumno: MAXIMO GABRIEL SILVA PARRAGUEZ, identificado con DNI N° 76399476, estudiante de la Escuela Profesional de ING. DE SISTEMAS DE LA UNIVERSIDAD SEÑOR DE SIPÁN, y autor del trabajo de investigación denominado: "ANÁLISIS COMPARATIVO DE ADAPTIVE BOOSTING Y DECISION TREE PARA PREDECIR LA MOROSIDAD DE CUOTAS SOCIALES DEL COLEGIO DE INGENIEROS DEL PERU CONSEJO DEPARTAMENTAL LAMBAYEQUE", al uso de dicha información que conforma el expediente técnico así como hojas de memorias, cálculos entre otros como planos para efectos exclusivamente académicos de la elaboración de tesis, enunciada líneas arriba de quien solicita se garantice la absoluta confidencialidad de la información solicitada.

Atentamente,



Ing. Denny John Fuentes Adrianzen
Presidente de Capitulo

ANEXO 03: Instrumentos de recolección de datos.

ANEXO 3.1.: Constancia de validación por Juicio de Expertos.



CONSTANCIA DE VALIDACIÓN POR JUICIO DE EXPERTOS

Quien suscribe, Mg. Ing. **Victor Alexci Tuesta Monteza**, con documento Nacional de Identidad N° 42722929, de profesión Ingeniero de Sistemas con Grado de **Ingeniero de Sistemas**, docente de la escuela profesional de Ingeniería de Sistemas de la Universidad Señor de Sipán, mediante la presente hago constatar que el instrumento utilizado para la recolección de datos del proyecto de tesis para obtener el grado de **Ingeniero de Sistemas**, denominado **"ANÁLISIS COMPARATIVO DE ADAPTIVE BOOSTING Y DECISION TREE PARA PREDECIR LA MOROSIDAD DE CUOTAS SOCIALES DEL COLEGIO DE INGENIEROS DEL PERU CONSEJO DEPARTAMENTAL LAMBAYEQUE."**, elaborado por el Bach. **MÁXIMO GABRIEL SILVA PARRAGUEZ**, identificado con DNI N° 76399476 de la Universidad Señor de Sipán, reúne los requisitos suficientes y necesarios para ser considerados válidos y confiables, por tanto, están aptos para ser aplicados en el logro de los objetivos que se plantearon en la investigación.

Atentamente

Fecha 07 de Diciembre del 2018



FIRMA DEL JUEZ DE EXPERTO

Mg. Ing: Victor Alexci Tuesta Monteza

Cargo : Docente de la E.P. Ing. Sistemas

ANEXO 3.2.: Fichas de Observación tanto para Adaboost como Decision Tree usados en cada mes.

Ficha de observación para Adaboost

Numero de Análisis:	1
Mes Analizado:	Enero
Tiempo de Construcción del Modelo (seg):	43.68
Tiempo de realización de la predicción (seg):	2.48

Matriz de confusión

	NO	SI
NO	1046	185
SI	251	1818

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.79
----------------------------	--------------

Otras Evaluaciones (%)

Exactitud (Precision)	87.87
Sensibilidad	80.65
Especificidad	90.76
AUC	0.857

Ficha de observación para Árbol de Decisión

Numero de Análisis:	2
Mes Analizado:	Enero
Tiempo de Construcción del Modelo (seg):	2.64
Tiempo de realización de la predicción (seg):	0.84

Matriz de confusión

	NO	SI
NO	1080	218
SI	217	1785

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.82
----------------------------	--------------

Otras Evaluaciones (%)

Exactitud (Precision)	89.16
Sensibilidad	83.27
Especificidad	89.12
AUC	0.860

Ficha de observación para Adaboost

Numero de Análisis:	3
Mes Analizado:	Febrero
Tiempo de Construcción del Modelo (seg):	49.60
Tiempo de realización de la predicción (seg):	2.46

Matriz de confusión

	NO	SI
NO	1035	181
SI	271	1813

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.30
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	87.00
Sensibilidad	79.25
Especificidad	90.92
AUC	0.851

Ficha de observación para Árbol de Decisión

Numero de Análisis:	4
Mes Analizado:	Febrero
Tiempo de Construcción del Modelo (seg):	5.94
Tiempo de realización de la predicción (seg):	0.68

Matriz de confusión

	NO	SI
NO	1088	229
SI	218	1765

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.45
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	89.01
Sensibilidad	83.31
Especificidad	88.52
AUC	0.85

Ficha de observación para Adaboost

Numero de Análisis:	5
Mes Analizado:	Marzo
Tiempo de Construcción del Modelo (seg):	52.61
Tiempo de realización de la predicción (seg):	2.50

Matriz de confusión

	NO	SI
NO	1089	190
SI	238	1783

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	87.03
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	88.22
Sensibilidad	82.06
Especificidad	90.37
AUC	0.862

Ficha de observación para Árbol de Decisión

Numero de Análisis:	6
Mes Analizado:	Marzo
Tiempo de Construcción del Modelo (seg):	1.84
Tiempo de realización de la predicción (seg):	0.66

Matriz de confusión

	NO	SI
NO	1021	163
SI	306	1810

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	85.79
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	85.54
Sensibilidad	76.94
Especificidad	91.74
AUC	0.84

Ficha de observación para Adaboost

Numero de Análisis:	7
Mes Analizado:	Abril
Tiempo de Construcción del Modelo (seg):	53.53
Tiempo de realización de la predicción (seg):	2.59

Matriz de confusión

	NO	SI
NO	1050	154
SI	284	1812

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.73
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	86.45
Sensibilidad	78.71
Especificidad	92.17
AUC	0.854

Ficha de observación para Árbol de Decisión

Numero de Análisis:	8
Mes Analizado:	Abril
Tiempo de Construcción del Modelo (seg):	2.79
Tiempo de realización de la predicción (seg):	0.86

Matriz de confusión

	NO	SI
NO	1110	200
SI	224	1766

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	87.15
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	88.74
Sensibilidad	83.21
Especificidad	89.83
AUC	0.86

Ficha de observación para Adaboost

Numero de Análisis:	9
Mes Analizado:	Mayo
Tiempo de Construcción del Modelo (seg):	71.01
Tiempo de realización de la predicción (seg):	2.54

Matriz de confusión

	NO	SI
NO	1021	135
SI	296	1848

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.94
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	86.19
Sensibilidad	77.52
Especificidad	93.19
AUC	0.854

Ficha de observación para Árbol de Decisión

Numero de Análisis:	10
Mes Analizado:	Mayo
Tiempo de Construcción del Modelo (seg):	2.82
Tiempo de realización de la predicción (seg):	0.62

Matriz de confusión

	NO	SI
NO	1098	217
SI	219	1766

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.79
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	88.97
Sensibilidad	83.37
Especificidad	89.06
AUC	0.86

Ficha de observación para Adaboost

Numero de Análisis:	11
Mes Analizado:	Junio
Tiempo de Construcción del Modelo (seg):	92.36
Tiempo de realización de la predicción (seg):	2.47

Matriz de confusión

	NO	SI
NO	1026	150
SI	302	1822

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.3
---------------------	------

Otras Evaluaciones (%)

Exactitud (Precision)	85.78
Sensibilidad	77.26
Especificidad	92.39
AUC	0.848

Ficha de observación para Árbol de Decisión

Numero de Análisis:	12
Mes Analizado:	Junio
Tiempo de Construcción del Modelo (seg):	3.05
Tiempo de realización de la predicción (seg):	0.94

Matriz de confusión

	NO	SI
NO	1083	212
SI	245	1760

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.15
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	87.78
Sensibilidad	81.55
Especificidad	89.25
AUC	0.85

Ficha de observación para Adaboost

Numero de Análisis:	13
Mes Analizado:	Julio
Tiempo de Construcción del Modelo (seg):	65.57
Tiempo de realización de la predicción (seg):	3.14

Matriz de confusión

	NO	SI
NO	1004	114
SI	326	1856

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.67
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	85.06
Sensibilidad	75.49
Especificidad	94.21
AUC	0.848

Ficha de observación para Árbol de Decisión

Numero de Análisis:	14
Mes Analizado:	Julio
Tiempo de Construcción del Modelo (seg):	2.31
Tiempo de realización de la predicción (seg):	0.66

Matriz de confusión

	NO	SI
NO	1077	1901
SI	253	1779

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.55
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	87.55
Sensibilidad	80.98
Especificidad	90.3
AUC	0.86

Ficha de observación para Adaboost

Numero de Análisis:	15
Mes Analizado:	Agosto
Tiempo de Construcción del Modelo (seg):	89.56
Tiempo de realización de la predicción (seg):	2.50

Matriz de confusión

	NO	SI
NO	988	118
SI	200	1994

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	90.36
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	90.88
Sensibilidad	83.16
Especificidad	94.41
AUC	0.887

Ficha de observación para Árbol de Decisión

Numero de Análisis:	16
Mes Analizado:	Agosto
Tiempo de Construcción del Modelo (seg):	5.72
Tiempo de realización de la predicción (seg):	0.64

Matriz de confusión

	NO	SI
NO	1056	163
SI	132	1949

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	91.06
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	93.66
Sensibilidad	88.89
Especificidad	92.28
AUC	0.91

Ficha de observación para Adaboost

Numero de Análisis:	17
Mes Analizado:	Septiembre
Tiempo de Construcción del Modelo (seg):	70.8
Tiempo de realización de la predicción (seg):	2.27

Matriz de confusión

	NO	SI
NO	963	98
SI	355	1884

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.27
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	84.14
Sensibilidad	73.07
Especificidad	95.06
AUC	0.84

Ficha de observación para Árbol de Decisión

Numero de Análisis:	18
Mes Analizado:	Septiembre
Tiempo de Construcción del Modelo (seg):	2.89
Tiempo de realización de la predicción (seg):	0.64

Matriz de confusión

	NO	SI
NO	1061	212
SI	257	1770

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	85.79
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	87.32
Sensibilidad	80.5
Especificidad	89.3
AUC	0.85

Ficha de observación para Adaboost

Numero de Análisis:	19
Mes Analizado:	Octubre
Tiempo de Construcción del Modelo (seg):	62.17
Tiempo de realización de la predicción (seg):	2.31

Matriz de confusión

	NO	SI
NO	986	112
SI	343	1859

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.21
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	84.42
Sensibilidad	74.19
Especificidad	94.32
AUC	0.84

Ficha de observación para Árbol de Decisión

Numero de Análisis:	20
Mes Analizado:	Octubre
Tiempo de Construcción del Modelo (seg):	2.23
Tiempo de realización de la predicción (seg):	0.64

Matriz de confusión

	NO	SI
NO	1001	106
SI	328	1865

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.85
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	85.04
Sensibilidad	75.32
Especificidad	94.62
AUC	0.85

Ficha de observación para Adaboost

Numero de Análisis:	21
Mes Analizado:	Noviembre
Tiempo de Construcción del Modelo (seg):	57.73
Tiempo de realización de la predicción (seg):	2.19

Matriz de confusión

	NO	SI
NO	988	105
SI	355	1852

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.06
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	83.91
Sensibilidad	73.57
Especificidad	94.63
AUC	0.84

Ficha de observación para Árbol de Decisión

Numero de Análisis:	22
Mes Analizado:	Noviembre
Tiempo de Construcción del Modelo (seg):	2.62
Tiempo de realización de la predicción (seg):	0.61

Matriz de confusión

	NO	SI
NO	999	106
SI	344	1851

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	86.36
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	84.33
Sensibilidad	74.39
Especificidad	94.58
AUC	0.84

Ficha de observación para Adaboost

Numero de Análisis:	23
Mes Analizado:	Diciembre
Tiempo de Construcción del Modelo (seg):	63.39
Tiempo de realización de la predicción (seg):	2.4

Matriz de confusión

	NO	SI
NO	991	109
SI	390	1810

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	84.88
---------------------	-------

Otras Evaluaciones (%)

Exactitud (Precision)	82.27
Sensibilidad	71.76
Especificidad	94.32
AUC	0.830

Ficha de observación para Árbol de Decisión

Numero de Análisis:	24
Mes Analizado:	Diciembre
Tiempo de Construcción del Modelo (seg):	3.14
Tiempo de realización de la predicción (seg):	0.56

Matriz de confusión

	NO	SI
NO	983	98
SI	398	1821

Evaluación Precisión (Accuracy) (%)

Porcentaje Dividido	84.97
---------------------	-------

Otras Evaluaciones (%)

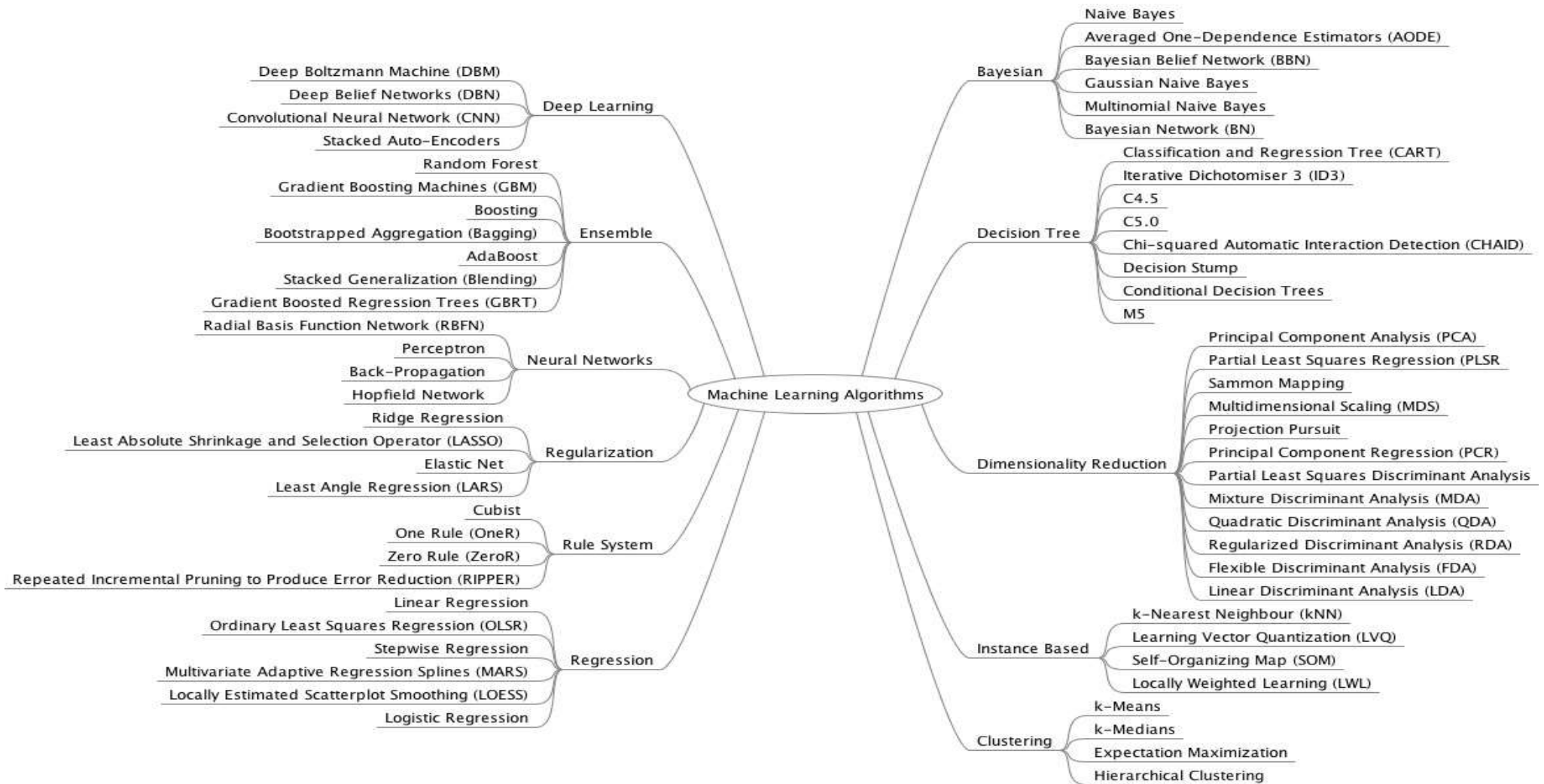
Exactitud (Precision)	82.06
Sensibilidad	71.18
Especificidad	94.89
AUC	0.83

ANEXO 04: Técnicas de aprendizaje automático mencionado en “The Use of Machine Learning Algorithms in Recommender Systems: A Systematic Review”

Type of Machine Learning Algorithm	Number of ML algorithms	Number of studies	Studies
Ensemble	22	14	(Aouay et al., 2014; Bar et al., 2013; Biancalana et al., 2011; Braida et al., 2015; Buabin, 2012; Elmongui et al., 2015; T.-K. Fan & Chang, 2010; Islam et al., 2015; Kao & Fahn, 2013; Middleton et al., 2004; Szymański & Rzeniewicz, 2016; Tsuji et al., 2014; Vialardi et al., 2011; Yan et al., 2013)
K Means	22	4	(T.-J. Lee & Tseng, 2012; Degemmis et al., 2007; Z. Fan et al., 2016; Zahra et al., 2015)
Support Vector Machines (SVM)	20	17	(Agarwal, 2011; Banerjee et al., 2012; Brouard & Pomot, 2016; Diaby et al., 2013, 2014; Ghazarian & Nematbakhsh, 2015; Jun, 2005; Kong et al., 2013; Pecli et al., 2015; Pronoza et al., 2016; I. Song et al., 2011; Sun et al., 2015; Szymański & Rzeniewicz, 2016; Tsuji et al., 2014; Verma et al., 2016; Yap et al., 2005; K. Zhao & Pan, 2015)
Bayesian	14	12	(Aouay et al., 2014; Banerjee et al., 2012; Costa et al., 2012, 2013; De Gemmis et al., 2008; Hernández del Olmo et al., 2009; Lops et al., 2009; Musto et al., 2010; Paparrizos et al., 2011; Pecli et al., 2015; Pronoza et al., 2016; S.-Z. Zhang et al., 2007)
Decision Tree	14	13	(Alemeye & Getahun, 2015; Aouay et al., 2014; Banerjee et al., 2012; Bellogín et al., 2011; Caraballo et al., 2014; Costa et al., 2013, 2013; Haiduc et al., 2013; Hernández del Olmo et al., 2009; Hussain et al., 2015; W.-P. Lee & Lu, 2003; T. Liu et al., 2011; Musto et al., 2010; Wei et al., 2011)
Matrix Factorization	13	7	(Bauer & Nanopoulos, 2014; Huang & Nikulin, 2014; Krohn-Grimberghe et al., 2011; J. Lu et al., 2013; Takács et al., 2008; Takács et al., 2009; Zhai & Li, 2015)
k Nearest Neighbors	11	10	(Aouay et al., 2014; Castro-Herrera et al., 2009; Das Dóres et al., 2016; Hernández del Olmo et al., 2009; Kong et al., 2013; Liang et al., 2014; R. Liu et al., 2014; Marović et al., 2011; McLaughlin & Herlocker, 2004; Pecli et al., 2015)
Latent Semantic Analysis	7	4	(Hotmann, 2003, 2004; Marović et al., 2011; Y. Zhang et al., 2009)
Logistic Regression	7	6	(Cai et al., 2012; Das et al., 2013; Krzywicki et al., 2015; Montañés et al., 2009; Pronoza et al., 2016; Sun et al., 2015)
Various	6	6	(Anaissi & Goyal, 2015; Forsati & Meybodi, 2010; Jung & Lee, 2004; Marović et al., 2011; Murfi & Obermayer, 2009; Roh et al., 2003)
Clustering	5	4	(Bjelica, 2010; Hassan et al., 2010; J. Li & Zaccari, 2004; Xuan et al., 2014)
Slope One	5	1	(Wang et al., 2012)
Association Rule	4	2	(Leopairete et al., 2013; X.-Z. Zhang, 2007)
Kernel Methods	4	4	(Devi & Venkatesh, 2013; Dinuzzo et al., 2011; X. Li & Chen, 2013; X. Li et al., 2014)
Bandit	3	3	(Bouneffouf et al., 2012; Nicol et al., 2014; Wang et al., 2014)

Frequency Counting	3	1	(Luong, Huynh, Gauch, Do, & Hoang, 2012)
Least Squares	3	2	(Gedikli et al., 2011; Takács et al., 2008)
Neural Network	3	3	(Geng et al., 2016; Marović et al., 2011; Aouay et al., 2014)
Regression Tree	3	2	(Marović et al., 2011; Peclí et al., 2015)
Sim. metric - Cosine Similarity	3	3	(Banerjee et al., 2012; Halder et al., 2014; Q.-C. Li et al., 2008)
Dictionary Learning	2	1	(Szabó et al., 2012)
Gradient Descent	2	2	(Cai et al., 2010; Pessiot et al., 2007)
Latent Dirichlet Allocation	2	2	(Xin et al., 2014; Yeh & Wu, 2010)
Linear Model	2	2	(R. Zhang & Tran, 2010; Moreno et al., 2012)
Linear Regression	2	2	(W. Zhang et al., 2008; Q. Zhao et al., 2015)
Pearson Correlation	2	2	(Halder et al., 2014; Wan et al., 2009)
Staked Regression	2	1	(Q. Liu et al., 2014)
Cross-temporal Link Prediction	1	1	(Oyama et al., 2012)
Euclidean Distance	1	1	(Buettner, 2016)
Gaussian Processes	1	1	(Y. Song et al., 2011)
Graphical Model	1	1	(Yuan et al., 2014)
Learning Automata	1	1	(Forsati et al., 2009)
Mahalanobis Classifier	1	1	(Tsapatsoulis et al., 2015)
Markov Model	1	1	(Baldominos et al., 2015)
Lagrange Multiplier	1	1	(Nie et al., 2013)
Mixture Model	1	1	(Y. Song et al., 2011)
Optimal Path Forest	1	1	(Marques et al., 2011)
Personality Diagnosis	1	1	(Marović et al., 2011)
Probabilistic Latent Semantic Analysis	1	1	(Jin et al., 2005)
Q-Learning	1	1	(Taghipour et al., 2007)
Regularization Methods	1	1	(Takács et al., 2008)
Shortest Path	1	1	(Luong, Huynh, Gauch, & Hoang, 2012)
Simil. metric - Geosemantic Proximity	1	1	(C. Lu et al., 2015)
Simil. metric - Aggregate Function	1	1	(Karahodza & Donko, 2015)
Simil. metric - WordNet Class Distance	1	1	(Nguyen et al., 2016)
Single Value Decomposition (SVD)	1	1	(Pantraki & Kotropoulos, 2015)

ANEXO 05: Técnicas de aprendizaje automático mencionado en “A Tour of Machine Learning Algorithms”



Fuente: <https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/>

ANEXO 06: Listado de Técnicas de aprendizaje automático Total (Población).

MACHINE LEARNING ALGORITHMS - SUPERVISED - NOT SUPERVISED	
1	RANDOM FOREST
2	GRADIENT BOOSTING MACHINES
3	BOOSTING
4	BOOTSTRAPPED AGGREGATION (BAGGING)
5	ADABOOST
6	STACKED GENERALIZATION (BLENDING)
7	GRADIENT BOOSTED REGRESION TREES (GBRT)
8	XGBOOST
9	CLASSIFICATION AND REGRESSION TREE (CART)
10	ITERATIVE DICHOTOMISER 3 (ID3)
11	DECISION TREE (C4.5)
12	DECISION TREE (C5.0)
13	CHI-SQUARED AUTOMATIC INTERACTION DETECTION (CHAID)
14	DECISION STUMP
15	CONDITIONAL DECISION TREES
16	M5
17	RADIAL BASIS FUNCTION NETWORK (RBFN)
18	PERCEPTRON
19	BACK-PROPAGATION
20	HOPFIELD NETWORK
21	NAIVE BAYES
22	AVERAGED ONE-DEPENDENCE ESTIMATORS
23	BAYESIAN BELIEF NETWORK
24	GAUSIAN NAIVE BAYES
25	MULTINOMIAL NAIVE BAYES
26	BAYESIAN NETWORK
27	LINEAR REGRESION
28	ORDINARY LEAST SQUARES REGRESSION (OLSR)
29	STEPWISE REGRESSION
30	MULTIVARIATE ADAPTIVE REGRESSION SPLINES (MARS)
31	LOCALLY ESTIMATED SCATTERPLOT SMOOTHING (LOESS)
32	LOGISTIC REGRESION
33	K-NEAREST NEIGHBOUR (KNN)
34	LEARNING VECTOR QUANTIZATION (LVQ)
35	SELF ORGANIZING MAP (SOM)
36	LOCALLY WEIGHTED LEARNING (LWL)
37	K-MEANS
38	K-MEDIAS
39	EXPECTATION MAXIMIZATION
40	HIERARCHICAL CLUSTERING
41	DEEP BOLTZMANN MACHINE (DBM)

42	DEEP BELIEF NETWORKS (DBN)
43	CONVOLUTIONAL NEURAL NETWORK (CNN)
44	STACKED AUTO-ENCODERS
45	CUBIST
46	ZERO RULE (OneR)
47	ZERO RULE (ZeroR)
	REPEATED INCREMENTAL PRUNING TO PRODUCE ERROR REDUCTION
48	(RIPPER)
49	PRINCIPAL COMPONENT ANALYSIS (PCA)
50	PARTIAL LEAST SQUARES REGRESSION (PLSR)
51	SAMMON MAPPING
52	MULTIDIMENSIONAL SCALING (MDS)
53	PROJECTION PURSUIT
54	PRINCIPAL COMPONENT REGRESSION (PCR)
55	PARTIAL LEAST SQUARES DISCRIMINANT ANALYSIS
56	MIXTURE DISCRIMINANT ANALYSIS (QDA)
57	QUADRATIC DISCRIMINANT ANALYSIS (QDA)
58	REGULARIZED DISCRIMINANT ANALYSIS (RDA)
59	FLEXIBLE DISCRIMINANT ANALYSIS (FDA)
60	LINEAR DISCRIMINANT ANALYSIS (LDA)
61	SUPPORT VECTOR MACHINES
62	ROUGH SETS
63	GENETIC ALGORITHMS
64	CASE BASED REASONING
65	LINEAR MODEL
66	LINEAR CLASSIFICATION
67	GRADIENT DESCENT
68	MATRIX FACTORIZATION
69	LATENT SEMANTIC ANALYSIS
70	CLUSTERING
71	SLOPE ONE
72	ASSOCIATION RULE
73	LEAST SQUARES
74	BANDIT
75	FREQUENCY COUNTING
76	SIM METRIC - COSINE SIMILARITY
77	DICTIONARY LEARNING
78	LATENT DIRICHLET ALLOCATION
79	PEARSON CORRELATION
80	STAKED REGRESSION
81	CROSS-TEMPORAL LINK PREDICTION
82	EUCLIDEAN DISTANCE
83	GAUSSIAN PROCESSES
84	GRAPHICAL MODEL
85	LEARNING AUTOMATA

86	MAHALANOBIS CLASSIFIER
87	MARKOV MODEL
88	LAGRANGE MULTIPLIER
89	MIXTURE MODEL
90	OPTIMAL PATH FOREST
91	PERSONALITY DIAGNOSIS
92	PROBABILISTIC LATENT SEMANTIC ANALYSIS
93	Q-LEARNING
94	REGULARIZATION METHODS
95	SHORTEST PATH
96	SIMIL METRIC GEOSEMANTIC PROXIMITY
97	SIMIL METRIC AGGREGATE FUNCTION
98	SIMIL METRIC WORDNET CLASS DISTANCE
99	SINGLE VALUE DECOMPOSITION (SVD)

ANEXO 07: Artículos Revisado para el top de técnicas mas usadas y con mejores resultados.

Nº	PAPER	¿QUÉ SE HIZO?	ALGORITMOS	RESULTADOS	CONCLUSION
1	Machine Learning Methods of Bankruptcy Prediction Using Accounting Ratios.(Li and Wang 2018)	Realizaron una comparación del metodo estadistico y el metodo de aprendizaje automático, para predecir la bancarrota con la utilizacion de empresas que cotizan en china. Se utilizo la base de datos Wind Financial Terminal y la base de datos economica financiera CCER.		Accuracy:	Los metodos empiricos muestran que los metodos de aprendizaje automático son superiores a los metodos estadisticos.
			STATISTICAL	70.8%	
			KNN	80.5%	
			SUPPORT VECTOR MACHINES (SVM)	84.0%	
			LOGISTIC REGRESION	79.0%	
			DECISION TREE	92.9%	
RANDOM FOREST	95.9%				
2	A PREDICTIVE SYSTEM FOR DETECTION OF BANKRUPTCY USING	Propusieron un sistema de predicción y para ello se evaluaron		Muestran Accuracy de:	Los resultados sugieren que se pueden implementar

	MACHINE LEARNING TECHNIQUES.(Nagaraj and Sridhar 2015)	los algoritmos de aprendizaje automatico y de estadística. Aplicados a un conjunto de datos cualitativo de UCI. El conjunto de datos consta de 250 instancias basadas en 6 atributos. La salida tenía dos clases de tipo nominal que describían la instancia como "Quiebra" y "No Quiebra". Se eliminó los atributos redundantes del conjunto de datos y para ello se utilizó la técnica de selección de características basadas en la correlación. El conjunto de datos se divide en entrenamiento (2/3 del conjunto de datos) y	LOGISTIC REGRESION	97.2%	0.97	técnicas de aprendizaje automático para la predicción de quiebra. Se implementó el sistema de predicción de quiebra. El sistema predictivo ayuda a predecir la quiebra de un conjunto de datos del cliente basado en el modelo SVM.
			RANDOM FOREST	97.4%	0.97	
			NAIVE BAYES	98.3%	0.98	
			NEURAL NETWORK	98.6%	0.98	
			SUPPORT VECTOR MACHINES (SVM)	99.6%	0.99	

		<p>conjunto de pruebas (1/3 del conjunto de datos). Se aplicó validación cruzada de 10 veces para probar la precisión. La precisión de cada algoritmo se estimó a partir de los resultados de validación cruzada.</p>				
3	<p>Financial distress prediction: The case of French small and medium-sized firms. (Mselmi, Lahiani, and Hamza 2017)</p>	<p>Tuvo como objetivo predecir las dificultades financieras de las pequeñas y medianas empresas francesas. Los datos se recogieron de la base de datos DIANE que contiene datos financieros anuales sobre empresas francesas pequeñas y</p>		Muestran Accuracy:		<p>La precisión de la predicción de ANN, SVM Y PLS-DA son estables en el tiempo. Un año antes de la quiebra, SVM supera a los otros modelos de acuerdo con la precisión general. Según la especificidad, los modelos ANN y Logit son similares y superan a SVM Y PLS-SVM en un 2.85%.</p>
				UN AÑO ANTES	DOS AÑOS ANTES	
			LOGIT MODEL	85.71%	92.86%	
			ARTIFICIAL NEURAL NETWORK	87.14%	88.57%	
			SUPPORT VECTOR MACHINES (SVM)	88.57%	92.86%	
PARTIAL LEAST SQUARES	84.29%	90%				
HYBRID MODEL INTEGRATING (SVM Y PARTIAL LEAST SQUARES)	85.72%	94.28%				

		medianas desde 2010 hasta 2013.			Dos años antes de la quiebra, PLS-SVM supera a SVM y Logit Model, que asu vez supera a PLS-DA Y ANNs de acuerdo con la precision general. Segun la especificidad y los errores Tipo II, SVM y PLS-DA son mejores que los modelos Logit, ANN y PLS-SVM.
4	Predicting financial distress of the South Korean manufacturing industries. (Bae 2012)	Desarrollaron un modelo de predicción de problemas financieros basado en SVM con la funcion radial para la prediccion de las dificultades financieras de empresas manufactureras con datos financieros de 1888 empresas		Accuracy:	Los experimentos demuestran que RSVM siempre supera a otros modelos en el desempeño de la prediccion de dificultades financieras corporativas. Nos menciona que debemos de utilizar RSVM para una mejor toma de decisiones
			SUPPORT VECTOR MACHINES (RSVM - Radial)	82.35%	
			DECISION TREE (C5.0)	74.47%	
			NEURAL NETWORK (MLP)	78.31%	
			MDA	77.65%	
			LOGIT MODEL	78.45%	

		recopiladas por el fondo de Garantía de Crédito de Korea (KODIT). Compararon el rendimiento de precisión de la clasificación entre RSVM y las técnicas de inteligencia artificial.	BAYESIAN NETWORKS	70.50%	de financiamiento e inversión, lo que puede conducir a mayores ganancias y valores de la empresa eventualmente.
5	Systematic review of bankruptcy prediction models: Towards a framework for tool selection. (Alaka et al. 2018)	Realizaron una revision sistemática de 49 articulos publicados entre 2010 y 2015. Esta revision muestra 6 herramientas de inteligencia artificial mas populares y prometedoras dentro del area de investigacion de modelos de prediccion de bancarrota.		Accuracy:	Se concluye que ninguna herramienta individual es predominantemente mejor que otras herramientas. Ademas menciona, que solo se puede encontrar un mejor modelo de rendimiento general mediante la integración informada de herramientas para formar un modelo hibrido.
			SUPPORT VECTOR MACHINES	94.54%	
			ARTIFICIAL NEURAL NETWORK	93.79%	
			DECISION TREE	91.83%	
			ROUGH SETS	91.10%	
			GENETIC ALGORITHMMS	90.36%	
			CASE BASED REASONING	59.24%	
6	Predicting corporate bankruptcy:	Se realizo un analisis critico de		Accuracy:	La revision muestra que las tecnicas

	where we stand?.(Aziz and Dar 2006)	un gran numero de estudios empiricos de prediccion de quiebra corporativa, basados en diversos modelos estadisticos, AIES y teoricos.	DECISION TREE	87.0%	estadisticas (MDA y Logit) se han usado con mayor frecuencia, que el enfoque AIES es relativamente nuevo y que los modelos teoricos son relativamente infrecuentes. Se observo que la precision predictiva era generalmente buena en todos los modelos, la revision sugiere que el AIES y los modelos teoricos tienen una precision predictiva promedio ligeramente mejor que los modelos estadisticos.
			CBR	83.0%	
			NEURAL NETWORKS	87.0%	
			GENETIC ALGORITHMS	88.0%	
			ROUGH SETS	91.0%	

7	C5.0 Classification Algorithm and Application on Individual Credit Evaluation of Banks.(Pang and Gong 2009)	Los registros de crédito individual incluyen datos tanto numéricos como no numéricos. El árbol de decisiones es una buena solución para este tipo de problema. Este artículo se centra en la evaluación crediticia individual de un banco comercial. En este artículo, realizamos una investigación profunda sobre el algoritmo C5.0 al incorporar la tecnología de "impulso" en la matriz de costos y el árbol sensible al costo para establecer un nuevo modelo para la evaluación de crédito individual de	DECISION TREE C5.0	<p>PRECISION:</p> <p>ENTRENAMIENTO (TIPO A) MODELO PRIMARIO:63.90% MODELO AJUSTADO:85.89%</p> <p>PRUEBA (TIPO A) MODELO PRIMARIO:28.81% MODELO AJUSTADO:67.80%</p>	MODELO AJUSTADO ES MEJOR QUE EL MODELO PRIMARIO
---	---	---	-----------------------	---	---

		CommercialBank. Aplicamos nuestro nuevo modelo para evaluar los registros de crédito individuales de un banco alemán y comparamos los resultados del modelo de árbol de decisión ajustado y el original.			
8	A practical approach to bankruptcy prediction for small businesses: Substituting the unavailable financial data for credit card sales information.	El propósito de este documento es desarrollar un modelo de predicción de quiebra para pequeñas empresas	SUPPORT VECTOR MACHINES BACKPROPAGATION NEURAL NETWORKS (BPN)	PRECISION: ENTRENAMIENTO-VALIDACION SVM: 79.0% - 74.2% CART: 74.7% - 70.3% BPN: 78.5% - 73.1% MDA: 69.0% - 70.1% C5.0: 78.2% - 70.8% LRA: 68.9% - 70.1%	Se obtuvo mayor precision la tecnica de SVM.

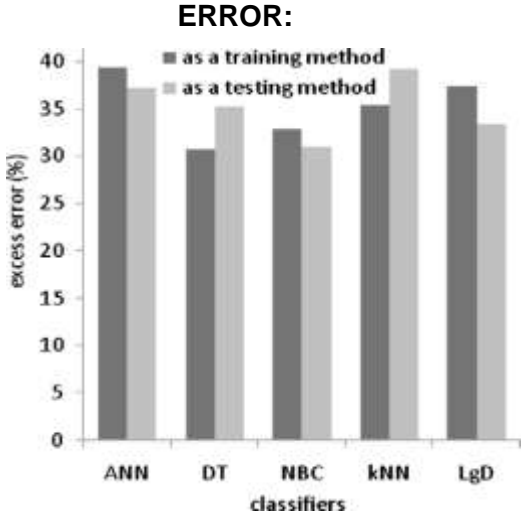
	(Yoon and Kwon 2010)	<p>utilizando información de ventas con tarjeta de crédito en lugar de utilizar datos financieros insuficientes para pequeñas empresas. De este modo, derivamos una variedad de variables que están altamente relacionadas con la información de ventas de tarjetas de crédito y adoptamos una técnica de aprendizaje automático de máquinas de vectores de soporte (SVM), y se compara con otras técnicas de aprendizaje automático.</p>	CART		
			DECISION TREE C5.0		
			ANALISIS DISCRIMINANTE MULTIPLE (MDA)		
			ANALISIS DE REGRESION LOGISTICA (LRA)		

9	<p>Applying a nonparametric random forest algorithm to assess the credit risk of the energy industry in China.(Tang, Cai, and Ouyang 2018)</p>	<p>El objetivo de este estudio es medir científicamente el riesgo crediticio de las tarjetas de crédito utilizadas en la industria energética de China y sentar las bases para una gestión integral del riesgo crediticio. Basado en un análisis de los factores que influyen en los factores que influyen en el riesgo de crédito, este estudio aplica el algoritmo de bosques aleatorios y los datos mensuales de las tarjetas de crédito utilizadas por los clientes de la industria de la energía en una sucursal del Postal Savings Bank of China</p>	RANDOM FOREST	<p>Accuracy: 91.5%</p>	<p>Este estudio reveló que el método de RF tiene una mayor precisión de predicción, es más tolerante a los valores atípicos y el ruido, y es menos probable que tenga problemas de sobrealimentación. Además, se puede usar para estimar la importancia de cada característica y, por lo tanto, es más eficaz para tratar datos de alta dimensión.</p>
---	--	--	---------------	-----------------------------------	--

		desde abril de 2014 hasta junio de 2017 para crear un modelo efectivo de evaluación de riesgo crediticio y mide científicamente el riesgo crediticio en la industria energética de China.			
10	Exploring the behaviour of base classifiers in credit scoring ensembles.(Marqués, García, and Sánchez 2012)	Durante los últimos años, se han aplicado con éxito diferentes enfoques para los conjuntos de	1 VECINO MAS CERCANO CLASIFICADOR INGENUO DE BAYES (NBC) REGRESION LOGISTICA	<p>Accuracy:</p> <ul style="list-style-type: none"> ▶ Se hicieron muchas pruebas y se obtuvieron diferentes resultados, pero en general: <ul style="list-style-type: none"> ▶ Los mejores modelos son el árbol de decisiones, el perceptrón multicapa y la regresión logística. ▶ El vecino más cercano y el ingenuo Bayes son 	Al considerar la precisión, se concluyo que los clasificadores 1-NN y NBC tienen el peor desempeño,

		<p>clasificadores a los problemas de calificación crediticia, demostrando ser más precisos que los modelos de predicción individuales. El presente documento evalúa el desempeño de siete técnicas de predicción individuales cuando se utilizan como miembros de cinco métodos conjuntos diferentes. El objetivo final de este estudio es sugerir clasificadores apropiados para cada enfoque conjunto en el contexto de la calificación crediticia.</p>	<p>PERCEPTRON MULTICAPA</p> <hr/> <p>REDES NEURONALES DE FUNCION BASICA RADIAL (RBF)</p> <hr/> <p>MAQUINA DE VECTORES DE SOPORTE (SVM)</p> <hr/> <p>DECISION TREE (C4.5)</p>	<p>los peores modelos, independientemente del conjunto.</p>	<p>independientemente del método de conjunto utilizado. También es importante tener en cuenta que para la mayoría de los enfoques conjuntos, C4.5 parece ser el mejor clasificador de base, seguido de cerca por MLP, regresión logística y SVM.</p>
11	An experimental comparison of	El objetivo de este estudio es	<p>ADABOOST</p> <hr/> <p>REGRESION</p>	AUC:	En un estudio comparativo de 9

	<p>classification techniques in debt recoveries scoring: Evidence from South Africa's unsecured lending market. (Colebatch and Engel 1974)</p>	<p>investigar el poder predictivo de algunas de las técnicas de clasificación más populares actualmente en uso, con atención específica para predecir la propensión de un prestatario que tiene 90 días o más de atraso en un préstamo sin garantía para pagar en un plazo fijo.</p>	<p>LOGISTICA</p> <p>ANALISIS DISCRIMINANTE CLASICO</p> <p>MODELOS ADITIVOS GENERALIZADOS (GAM_LR Y GAM_GEV)</p> <p>ANALISIS DISCRIMINANTE FLEXIBLE (FDA)</p> <p>VALORES ESTREMOS GENERALIZADOS BINARIOS FLEXIBLES (BGEVA)</p> <p>MAQUINA DE VECTORES DE SOPORTE (SVM)</p> <p>BOSQUES ALEATORIOS (RANDOM FOREST)</p>	<p>Los resultados anteriores han establecido que un modelo GAM con un enlace GEV (GAM_GEV) ocupa el número 1 de los 9 clasificadores considerados en este estudio.</p>	<p>metodologías de clasificación en 10 conjuntos de datos, hubo pruebas estadísticamente significativas de que los modelos de aditivos generalizados (GAM) con una función de enlace de valor extremo generalizado (GEV) realizan técnicas de calificación estándar de la industria como la regresión logística (donde - Las características han sido discretizadas y algunas técnicas populares de aprendizaje automático, tales como bosques aleatorios equilibrados, AdaBoost (con árboles de clasificación como base) y máquinas de vectores de soporte.</p>
--	--	--	---	--	--

12	Multiple classifier application to credit risk assessment. (Twala 2010)	Se investiga la precisión de cinco clasificadores aplicados para la predicción de riesgo de crédito y los probamos en cuatro conjuntos de datos del mundo real.	REDES NEURONALES ARTIFICIALES DECISION TREE NBC KNN LgD	<p>ERROR:</p> 	kNN es evidentemente el peor método global debido a la difícil elección que se tiene que hacer al determinar la métrica de distancia k que se desconoce para n finito. NBC es el más eficaz con una seria competencia de LgD y DT.
13	An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes.(Jones et al. 2015)	En este estudio, se comparó el rendimiento predictivo de 20 clasificadores binarios en una gran muestra de datos de cambios en las calificaciones crediticias corporativas internacionales que se producen entre 1983 y 2013. Los clasificadores se	ADABOOST GENERALIZED BOOSTING RANDOM FOREST RANDOM FOREST OBLIQUE NEURAL NETWORKS SUPPORT VECTOR MACHINES SUPPORT VECTOR MACHINES PENALISED LOGISTIC GAM LOGISTIC MARS PROBIT GAM	<p>AUC:</p> AdaBoost -> 0.9436 GeneralisedBoosting -> 0.9571 RandomForests -> 0.9442 RandomForests_Oblique -> 0.9229 NeuralNetworks -> 0.8359 SVM -> 0.8385 SVM_Penalised -> 0.7897 Logistic_GAM -> 0.9077 Logistic_MARS -> 0.8850 Probit_GAM -> 0.9115 Probit_MARS -> 0.8864 MixedLogit -> 0.8141 Probit_Boosted -> 0.7840 Logistic_Subset -> 0.8023 Logistic_Stepwise -> 0.7972 Logistic_Penalised -> 0.7876	El análisis de la puntuación AUC y H indica que los clasificadores más nuevos, como el aumento generalizado, AdaBoost y los bosques aleatorios, superaron ampliamente a todos los demás clasificadores en las muestras de prueba de sección transversal y longitudinal, y en

		prueban empíricamente en ocho variaciones del conjunto de datos que incluyen muestras de prueba transversales y longitudinales; y la comparación del rendimiento predictivo a través de conjuntos de datos transformados y no transformados, con imputación de valor faltante y sin él.	PROBIT MARS MIXED LOGIT PROBIT BOOSTED LOGISTIC SUBSET LOGISTIC STEPWISE LOGISTIC PENALISED PROBIT SUNSET PROBIT STEPWISE QDA LDA	Probit_Subset -> 0.7944 Probit_Stepwise -> 0.7910 QDA -> 0.7931 LDA -> 0.7971	casi todas las variaciones de los datos muestreados. Dado que estos clasificadores parecen ser notablemente precisos y estables, pueden ser una promesa significativa para futuras investigaciones y prácticas en este campo.
14	Machine learning models and bankruptcy prediction.(Barboza, Kimura, and Altman 2017)	En este estudio, probamos modelos de aprendizaje automático (máquinas de vectores de soporte, ensacado, refuerzo y bosques aleatorios) para	BAGGING BOOSTING (VARIANTE ADABOOST) RANDOM FOREST SUPPORT VECTOR MACHINES (SVM-LIN Y SVM-RBF) ARTIFICIAL NEURAL NETWORK	Accuracy: SVM-Linear -> 71.52% SVM-RBF -> 79.77% Boosting -> 86.65% Bagging -> 85.67% Random forest -> 87.06% Neural networks -> 72.98% Logit -> 76.29% MDA -> 52.18%	Con respecto a los índices de precisión, los resultados muestran que los modelos tradicionales (MDA, LR y ANN) tienen una capacidad predictiva más baja (entre 52% y 77%) que los modelos de aprendizaje

		predecir la bancarrota un año antes del evento y comparamos su desempeño con los resultados del análisis discriminante, la regresión logística y las redes neuronales.	LOGISTIC REGRESION		automático (71% a 87%).
			MULTIVARIATE DISCRIMINANT ANALYSIS (MDA)		
15	AdaBoost based bankruptcy forecasting of Korean construction companies. (Heo and Yang 2014)	En este documento, se muestra que el AdaBoost (impulso adaptativo) es un modelo apropiado para juzgar el riesgo financiero de las empresas de construcción coreanas. Clasificamos las empresas de construcción en tres grupos: grande, mediano y pequeño según el capital de una empresa.	ADABOOST	<p>Accuracy: Adaboost -> 78.5% ANN -> 77.1% SVM -> 73.3% Arbol de Decisiones -> 73.1% Z-Score -> 51.3%</p>	Los resultados experimentales mostraron que AdaBoost tiene más poder predictivo que otros, especialmente para el gran grupo de empresas que tiene el capital más de 50 mil millones de won.
			ARTIFICIAL NEURAL NETWORK		
			SUPPORT VECTOR MACHINES		
			DECISION TREE		
			Z-SCORE		

		Analizamos la capacidad predictiva de AdaBoost y otros algoritmos para cada grupo de empresas.			
16	Classifier selection and clustering with fuzzy assignment in ensemble model for credit scoring.(Zhang, He, and Zhang 2018)	Este estudio se basa en el conjunto de cinco de los clasificadores básicos más reconocidos en calificación crediticia: regresión logística, máquina de vectores de soporte, red neuronal, árbol de decisión de aumento de gradiente y bosque aleatorio. Se prueba en tres conjuntos de datos de puntuación de crédito (australiano, alemán, japonés) y tres de medidas		Evaluaron 3 países: Australian, German, Japanese Accuracy, AUC, F-Score	En resumen, LR, SVM, RF y GBDT son los mejores clasificadores base entre los nueve que hemos probado, pero todos ellos se superan con el clasificador conjunto de LR, SVM, RF, GBDT y NN, lo que demuestra la efectividad de este conjunto.
			KNN	(0.7913-0.7095-0.8000)	
			LOGISTIC REGRESION	(0.8688-0.7560-0.8232)	
			RANDOM FOREST	(0.8609-0.8649-0.7230)	
			GBDT	(0.8551-0.7530-0.8616)	
			LDA	(0.8091-0.5350-0.7717)	
			SUPPORT VECTOR MACHINES	(0.8583-0.6960-0.8630)	
			NEURAL NETWORKS	(0.8380-0.7395-0.8181)	
			ADABOOST	(0.8323-0.7270-0.8580)	
	DECISION TREE	(0.8207-0.6610-0.8152)			

		de rendimiento (precisión, AUC, puntuación F).			
17	A genetic algorithm approach for SMEs bankruptcy prediction: Empirical evidence from Italy. (Gordini 2014)	Comparan el potencial de los algoritmos genéticos (AG) con los de logística Regresión (LR) y máquina de vectores de soporte (SVM), en la predicción de quiebra. Se aplican a una gran muestra de 3.100 pymes manufactureras italianas, tres, dos y un año antes de la quiebra. Los resultados indican que las AG son un instrumento muy eficaz y prometedor para evaluar la probabilidad de Quiebra de PYME comparada con		Accuracy:	Los resultados indican que las AG son un instrumento muy eficaz y prometedor para evaluar la probabilidad de quiebra de las PYMES comparada con LR y SVM
			GENETIC ALGORITHMS	71.50%	
			SUPPORT VECTOR MACHINES	69.50%	
			LOGISTIC REGRESION	66.80%	

		LR y SVM.			
18	Bankruptcy forecasting: An empirical comparison of AdaBoost and neural networks. (Alfaro and Elizondo 2008)	Se compara la precisión de predicción de ambas técnicas (Redes Neuronales y AdaBoost) en un conjunto de empresas europeas, considerando las variables de predicción usuales, como las razones financieras, así como las variables cualitativas, como el tamaño de la empresa, la actividad y la estructura legal.		ERROR:	En este estudio, se compararon dos métodos de clasificación, que muestran la mejora en la precisión que AdaBoost logra contra la red neuronal.
			AdaBoost	8.898%	
			Neural Networks	12.712%	

ANEXO 08: Detallando Todas las Tablas de La base de datos Colegio_Ingenieros (Base de Datos Original).

Nro	Tabla	Descripción	Nro Registros	VARIABLES DE PREDICCIÓN							
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de	Especialidad	Categoría
1	aag01ns	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR
2	acn0000	MOVIMIENTOS Y PRESUPUESTO DEL COLEGIO DE INGENIEROS	1125	SR	SR	SR	SR	SR	SR	SR	SR
3	acn0100	MOVIMIENTOS Y PRESUPUESTO DEL COLEGIO DE INGENIEROS	32069	SR	SR	SR	SR	SR	SR	SR	SR
4	ce0000	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR
5	cgm01011992	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1906	SR	SR	SR	SR	SR	SR	SR	SR
6	cgm01011993	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1906	SR	SR	SR	SR	SR	SR	SR	SR
7	cgm01011994	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR
8	cgm01011995	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1889	SR	SR	SR	SR	SR	SR	SR	SR

9	cgm01011996	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
10	cgm01011997	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
11	cgm01011998	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
12	cgm01011999	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
13	cgm01012001	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
14	cgm01012002	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
15	cgm01012003	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
16	cgm01012004	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
17	cgm01012005	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1857	SR	SR	SR	SR	SR	SR	SR	SR	SR
18	cgm01012006	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1857	SR	SR	SR	SR	SR	SR	SR	SR	SR
19	cgm01012007	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1857	SR	SR	SR	SR	SR	SR	SR	SR	SR
20	cgm01012008	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
21	cgm01012009	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR

22	cgm01012010	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1871	SR	SR	SR	SR	SR	SR	SR	SR	SR
23	cgm01012011	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1896	SR	SR	SR	SR	SR	SR	SR	SR	SR
24	cgm01012012	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1898	SR	SR	SR	SR	SR	SR	SR	SR	SR
25	cgm01012013	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1902	SR	SR	SR	SR	SR	SR	SR	SR	SR
26	cgm01012014	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1896	SR	SR	SR	SR	SR	SR	SR	SR	SR
27	cgm01012015	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1901	SR	SR	SR	SR	SR	SR	SR	SR	SR
28	cgm01012016	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1908	SR	SR	SR	SR	SR	SR	SR	SR	SR
29	cgm01012017	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1901	SR	SR	SR	SR	SR	SR	SR	SR	SR
30	cgm01012018	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1900	SR	SR	SR	SR	SR	SR	SR	SR	SR
31	cgm01012019	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
32	cgm01012020	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
33	cgm01012021	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
34	cgm01012022	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR

35	cgm01012023	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
36	cgm01012024	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
37	cgm01012025	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
38	cgm01012026	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
39	cgm01012027	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
40	cgm01012028	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
41	cgm01012029	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
42	cgm01021992	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
43	cgm01021993	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
44	cgm01021994	PAGO DETALLADO POR MESES DE UN COLEGIADO	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
45	cgm01021995	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
46	cgm01021996	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
47	cgm01021997	SALDO INICIAL	1173	SR	SR	SR	SR	SR	SR	SR	SR	SR
48	cgm01021998	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
49	cgm01021999	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
50	cgm01022000	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
51	cgm01022001	CUOTAS MENSUALES DE COLEGIADOS/ DATOS	1116	SR	SR	SR	SR	SR	SR	SR	SR	SR

		INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)											
52	cgm01022002	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	1260	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
53	cgm01022003	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	1788	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
54	cgm01022004	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	2475	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
55	cgm01022005	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	2858	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
56	cgm01022006	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	3278	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
57	cgm01022007	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	3818	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
58	cgm01022008	CUOTAS MENSUALES DE COLEGIADOS/ DATOS	2862	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR

		INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)											
59	cgm01022009	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	3298	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
60	cgm01022010	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	178474	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
61	cgm01022011	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES (REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	384322	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
62	cgm01022012	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	422948	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
63	cgm01022013	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	385505	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
64	cgm01022014	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	424996	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
65	cgm01022015	CUOTAS MENSUALES DE COLEGIADOS/ DATOS	532793	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR

		INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)										
66	cgm01022016	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	579615	SR	SR	SR	SR	SR	SR	SR	SR	SR
67	cgm01022017	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	619743	SR	SR	SR	SR	SR	SR	SR	SR	SR
68	cgm01022018	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	322043	SR	SR	SR	SR	SR	SR	SR	SR	SR
69	cgm01022019	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	4927	SR	SR	SR	SR	SR	SR	SR	SR	SR
70	cgm01022020	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	106	SR	SR	SR	SR	SR	SR	SR	SR	SR
71	cgm01022021	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	50	SR	SR	SR	SR	SR	SR	SR	SR	SR
72	cgm01022022	CUOTAS MENSUALES DE COLEGIADOS/ DATOS	24	SR	SR	SR	SR	SR	SR	SR	SR	SR

		INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)											
73	cgm01022023	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	24	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
74	cgm01022024	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	24	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
75	cgm01022025	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	24	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
76	cgm01022026	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	24	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
77	cgm01022027	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	24	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
78	cgm01022028	CUOTAS MENSUALES DE COLEGIADOS/ DATOS INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)	24	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
79	cgm01022029	CUOTAS MENSUALES DE COLEGIADOS/ DATOS	24	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR

		INCONSISTENTES(REDUNDANTES Y PRESENTES EN LA TABLA mst01ccc)										
80	cgm01031992	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1906	SR	SR	SR	SR	SR	SR	SR	SR	SR
81	cgm01031993	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1906	SR	SR	SR	SR	SR	SR	SR	SR	SR
82	cgm01031994	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
83	cgm01031995	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1889	SR	SR	SR	SR	SR	SR	SR	SR	SR
84	cgm01031996	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
85	cgm01031997	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
86	cgm01031998	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
87	cgm01031999	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
88	cgm01032000	FLUJO DE CAJA	842	SR	SR	SR	SR	SR	SR	SR	SR	SR
89	cgm01032001	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
90	cgm01032002	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
91	cgm01032003	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
92	cgm01032004	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR

93	cgm01032005	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1857	SR	SR	SR	SR	SR	SR	SR	SR	SR
94	cgm01032006	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1857	SR	SR	SR	SR	SR	SR	SR	SR	SR
95	cgm01032007	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1857	SR	SR	SR	SR	SR	SR	SR	SR	SR
96	cgm01032008	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
97	cgm01032009	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
98	cgm01032010	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1871	SR	SR	SR	SR	SR	SR	SR	SR	SR
99	cgm01032011	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1896	SR	SR	SR	SR	SR	SR	SR	SR	SR
100	cgm01032012	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1898	SR	SR	SR	SR	SR	SR	SR	SR	SR
101	cgm01032013	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1902	SR	SR	SR	SR	SR	SR	SR	SR	SR
102	cgm01032014	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1896	SR	SR	SR	SR	SR	SR	SR	SR	SR
103	cgm01032015	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1901	SR	SR	SR	SR	SR	SR	SR	SR	SR
104	cgm01032016	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1908	SR	SR	SR	SR	SR	SR	SR	SR	SR
105	cgm01032017	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1901	SR	SR	SR	SR	SR	SR	SR	SR	SR

106	cgm01032018	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1900	SR	SR	SR	SR	SR	SR	SR	SR	SR
107	cgm01032019	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
108	cgm01032020	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
109	cgm01032021	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
110	cgm01032022	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
111	cgm01032023	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
112	cgm01032024	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
113	cgm01032025	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
114	cgm01032026	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
115	cgm01032027	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
116	cgm01032028	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
117	cgm01032029	LISTADO DE ACTIVOS Y PASIVOS CORRIENTES DEL CIP	1849	SR	SR	SR	SR	SR	SR	SR	SR	SR
118	cgm01041992	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
119	cgm01041993	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
120	cgm01041994	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

121	cgm01041995	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
122	cgm01041996	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
123	cgm01041997	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
124	cgm01041998	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
125	cgm01041999	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
126	cgm01042001	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
127	cgm01042002	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
128	cgm01042003	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
129	cgm01042004	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
130	cgm01042005	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
131	cgm01042006	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
132	cgm01042007	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
133	cgm01042008	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
134	cgm01042009	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
135	cgm01042010	SIN RELEVANCIA	210	SR	SR	SR	SR	SR	SR	SR	SR	SR
136	cgm01042011	SIN RELEVANCIA	317	SR	SR	SR	SR	SR	SR	SR	SR	SR
137	cgm01042012	SIN RELEVANCIA	513	SR	SR	SR	SR	SR	SR	SR	SR	SR
138	cgm01042013	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
139	cgm01042014	SIN RELEVANCIA	408	SR	SR	SR	SR	SR	SR	SR	SR	SR
140	cgm01042015	SIN RELEVANCIA	169	SR	SR	SR	SR	SR	SR	SR	SR	SR
141	cgm01042016	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
142	cgm01042017	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

143	cgm01042018	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
144	cgm01042019	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
145	cgm01042020	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
146	cgm01042021	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
147	cgm01042022	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
148	cgm01042023	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
149	cgm01042024	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
150	cgm01042025	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
151	cgm01042026	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
152	cgm01042027	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
153	cgm01042028	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
154	cgm01042029	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
155	cgm01051992	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
156	cgm01051993	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
157	cgm01051994	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
158	cgm01051995	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
159	cgm01051996	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
160	cgm01051997	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
161	cgm01051998	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
162	cgm01051999	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
163	cgm01052001	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
164	cgm01052002	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

165	cgm01052003	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
166	cgm01052004	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
167	cgm01052005	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
168	cgm01052006	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
169	cgm01052007	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
170	cgm01052008	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
171	cgm01052009	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
172	cgm01052010	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
173	cgm01052011	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
174	cgm01052012	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
175	cgm01052013	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
176	cgm01052014	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
177	cgm01052015	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
178	cgm01052016	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
179	cgm01052017	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
180	cgm01052018	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
181	cgm01052019	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
182	cgm01052020	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
183	cgm01052021	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
184	cgm01052022	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
185	cgm01052023	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
186	cgm01052024	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

187	cgm01052025	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
188	cgm01052026	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
189	cgm01052027	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
190	cgm01052028	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
191	cgm01052029	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
192	cgp01021992	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
193	cgp01021993	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
194	cgp01021994	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
195	cgp01021995	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
196	cgp01021996	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
197	cgp01021997	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
198	cgp01021998	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
199	cgp01021999	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
200	cgp01022001	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
201	cgp01022002	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
202	cgp01022003	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
203	cgp01022004	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
204	cgp01022005	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
205	cgp01022006	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
206	cgp01022007	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
207	cgp01022008	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
208	cgp01022009	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

209	cgp01022010	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
210	cgp01022011	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
211	cgp01022012	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
212	cgp01022013	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
213	cgp01022014	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
214	cgp01022015	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
215	cgp01022016	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
216	cgp01022017	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
217	cgp01022018	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
218	cgp01022019	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
219	cgp01022020	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
220	cgp01022021	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
221	cgp01022022	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
222	cgp01022023	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
223	cgp01022024	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
224	cgp01022025	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
225	cgp01022026	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
226	cgp01022027	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
227	cgp01022028	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
228	cgp01022029	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
229	cgp01031992	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
230	cgp01031993	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

231	cgp01031994	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
232	cgp01031995	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
233	cgp01031996	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
234	cgp01031997	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
235	cgp01031998	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
236	cgp01031999	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
237	cgp01032001	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
238	cgp01032002	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
239	cgp01032003	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
240	cgp01032004	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
241	cgp01032005	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
242	cgp01032006	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
243	cgp01032007	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
244	cgp01032008	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
245	cgp01032009	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
246	cgp01032010	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
247	cgp01032011	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
248	cgp01032012	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
249	cgp01032013	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
250	cgp01032014	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
251	cgp01032015	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
252	cgp01032016	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

253	cgp01032017	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
254	cgp01032018	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
255	cgp01032019	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
256	cgp01032020	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
257	cgp01032021	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
258	cgp01032022	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
259	cgp01032023	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
260	cgp01032024	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
261	cgp01032025	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
262	cgp01032026	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
263	cgp01032027	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
264	cgp01032028	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
265	cgp01032029	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
266	cgr01011992	SIN RELEVANCIA	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
267	cgr01011993	SIN RELEVANCIA	3	SR	S	SR	SR	SR	SR	SR	SR	SR
268	cgr01011994	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
269	cgr01011995	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
270	cgr01011996	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
271	cgr01011997	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
272	cgr01011998	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
273	cgr01011999	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
274	cgr01012001	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR

275	cgr01012002	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
276	cgr01012003	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
277	cgr01012004	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
278	cgr01012005	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
279	cgr01012006	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
280	cgr01012007	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
281	cgr01012008	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
282	cgr01012009	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
283	cgr01012010	SIN RELEVANCIA	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
284	cgr01012011	SIN RELEVANCIA	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
285	cgr01012012	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
286	cgr01012013	SIN RELEVANCIA	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
287	cgr01012014	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
288	cgr01012015	SIN RELEVANCIA	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
289	cgr01012016	SIN RELEVANCIA	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
290	cgr01012017	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
291	cgr01012018	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
292	cgr01012019	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
293	cgr01012020	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
294	cgr01012021	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
295	cgr01012022	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
296	cgr01012023	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR

297	cgr01012024	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
298	cgr01012025	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
299	cgr01012026	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
300	cgr01012027	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
301	cgr01012028	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
302	cgr01012029	SIN RELEVANCIA	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
303	cgr01021992	SIN RELEVANCIA	38	SR	SR	SR	SR	SR	SR	SR	SR	SR
304	cgr01021993	SIN RELEVANCIA	38	SR	SR	SR	SR	SR	SR	SR	SR	SR
305	cgr01021994	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
306	cgr01021995	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
307	cgr01021996	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
308	cgr01021997	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
309	cgr01021998	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
310	cgr01021999	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
311	cgr01022001	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
312	cgr01022002	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
313	cgr01022003	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
314	cgr01022004	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
315	cgr01022005	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
316	cgr01022006	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
317	cgr01022007	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
318	cgr01022008	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR

319	cgr01022009	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
320	cgr01022010	SIN RELEVANCIA	66	SR	SR	SR	SR	SR	SR	SR	SR	SR
321	cgr01022011	SIN RELEVANCIA	38	SR	SR	SR	SR	SR	SR	SR	SR	SR
322	cgr01022012	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
323	cgr01022013	SIN RELEVANCIA	38	SR	SR	SR	SR	SR	SR	SR	SR	SR
324	cgr01022014	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
325	cgr01022015	SIN RELEVANCIA	38	SR	SR	SR	SR	SR	SR	SR	SR	SR
326	cgr01022016	SIN RELEVANCIA	38	SR	SR	SR	SR	SR	SR	SR	SR	SR
327	cgr01022017	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
328	cgr01022018	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
329	cgr01022019	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
330	cgr01022020	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
331	cgr01022021	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
332	cgr01022022	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
333	cgr01022023	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
334	cgr01022024	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
335	cgr01022025	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
336	cgr01022026	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
337	cgr01022027	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
338	cgr01022028	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
339	cgr01022029	SIN RELEVANCIA	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
340	cgr01031992	SIN RELEVANCIA	1071	SR	SR	SR	SR	SR	SR	SR	SR	SR

341	cgr01031993	SIN RELEVANCIA	1071	SR	SR	SR	SR	SR	SR	SR	SR	SR
342	cgr01031994	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
343	cgr01031995	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
344	cgr01031996	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
345	cgr01031997	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
346	cgr01031998	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
347	cgr01031999	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
348	cgr01032001	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
349	cgr01032002	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
350	cgr01032003	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
351	cgr01032004	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
352	cgr01032005	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
353	cgr01032006	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
354	cgr01032007	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
355	cgr01032008	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
356	cgr01032009	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
357	cgr01032010	SIN RELEVANCIA	2053	SR	SR	SR	SR	SR	SR	SR	SR	SR
358	cgr01032011	SIN RELEVANCIA	1071	SR	SR	SR	SR	SR	SR	SR	SR	SR
359	cgr01032012	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
360	cgr01032013	SIN RELEVANCIA	1071	SR	SR	SR	SR	SR	SR	SR	SR	SR
361	cgr01032014	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
362	cgr01032015	SIN RELEVANCIA	1071	SR	SR	SR	SR	SR	SR	SR	SR	SR

363	cgr01032016	SIN RELEVANCIA	1071	SR	SR	SR	SR	SR	SR	SR	SR	SR
364	cgr01032017	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
365	cgr01032018	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
366	cgr01032019	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
367	cgr01032020	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
368	cgr01032021	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
369	cgr01032022	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
370	cgr01032023	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
371	cgr01032024	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
372	cgr01032025	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
373	cgr01032026	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
374	cgr01032027	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
375	cgr01032028	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
376	cgr01032029	SIN RELEVANCIA	580	SR	SR	SR	SR	SR	SR	SR	SR	SR
377	cgt01011992	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
378	cgt01011993	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
379	cgt01011994	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
380	cgt01011995	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
381	cgt01011996	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
382	cgt01011997	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
383	cgt01011998	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
384	cgt01011999	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR

385	cgt01012001	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
386	cgt01012002	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
387	cgt01012003	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
388	cgt01012004	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
389	cgt01012005	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
390	cgt01012006	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
391	cgt01012007	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
392	cgt01012008	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
393	cgt01012009	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
394	cgt01012010	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
395	cgt01012011	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
396	cgt01012012	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
397	cgt01012013	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
398	cgt01012014	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
399	cgt01012015	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
400	cgt01012016	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
401	cgt01012017	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
402	cgt01012018	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
403	cgt01012019	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
404	cgt01012020	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
405	cgt01012021	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
406	cgt01012022	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR

407	cgt01012023	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
408	cgt01012024	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
409	cgt01012025	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
410	cgt01012026	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
411	cgt01012027	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
412	cgt01012028	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
413	cgt01012029	SIN RELEVANCIA	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
414	cgt01031992	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
415	cgt01031993	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
416	cgt01031994	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
417	cgt01031995	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
418	cgt01031996	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
419	cgt01031997	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
420	cgt01031998	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
421	cgt01031999	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
422	cgt01032001	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
423	cgt01032002	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
424	cgt01032003	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
425	cgt01032004	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
426	cgt01032005	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
427	cgt01032006	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
428	cgt01032007	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR

429	cgt01032008	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
430	cgt01032009	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
431	cgt01032010	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
432	cgt01032011	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
433	cgt01032012	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
434	cgt01032013	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
435	cgt01032014	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
436	cgt01032015	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
437	cgt01032016	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
438	cgt01032017	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
439	cgt01032018	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
440	cgt01032019	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
441	cgt01032020	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
442	cgt01032021	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
443	cgt01032022	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
444	cgt01032023	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
445	cgt01032024	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
446	cgt01032025	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
447	cgt01032026	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
448	cgt01032027	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
449	cgt01032028	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
450	cgt01032029	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR

451	cgt01061992	MESES DEL AÑO Y CANTIDAD DE DIAS	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
452	cgt01061993	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
453	cgt01061994	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
454	cgt01061995	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
455	cgt01061996	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
456	cgt01061997	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
457	cgt01061998	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
458	cgt01061999	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
459	cgt01062001	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
460	cgt01062002	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
461	cgt01062003	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
462	cgt01062004	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
463	cgt01062005	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
464	cgt01062006	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR

465	cgt01062007	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
466	cgt01062008	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
467	cgt01062009	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
468	cgt01062010	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
469	cgt01062011	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
470	cgt01062012	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
471	cgt01062013	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
472	cgt01062014	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
473	cgt01062015	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
474	cgt01062016	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
475	cgt01062017	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
476	cgt01062018	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
477	cgt01062019	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR

478	cgt01062020	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
479	cgt01062021	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
480	cgt01062022	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
481	cgt01062023	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
482	cgt01062024	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
483	cgt01062025	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
484	cgt01062026	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
485	cgt01062027	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
486	cgt01062028	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
487	cgt01062029	LISTADO DE NOMBRES / SIN RELEVANCIA	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
488	cgt01071992	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
489	cgt01071993	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
490	cgt01071994	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
491	cgt01071995	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
492	cgt01071996	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
493	cgt01071997	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

494	cgt01071998	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
495	cgt01071999	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
496	cgt01072001	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
497	cgt01072002	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
498	cgt01072003	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
499	cgt01072004	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
500	cgt01072005	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
501	cgt01072006	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
502	cgt01072007	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
503	cgt01072008	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
504	cgt01072009	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
505	cgt01072010	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
506	cgt01072011	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
507	cgt01072012	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
508	cgt01072013	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
509	cgt01072014	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
510	cgt01072015	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
511	cgt01072016	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
512	cgt01072017	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
513	cgt01072018	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
514	cgt01072019	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
515	cgt01072020	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

516	cgt01072021	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
517	cgt01072022	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
518	cgt01072023	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
519	cgt01072024	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
520	cgt01072025	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
521	cgt01072026	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
522	cgt01072027	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
523	cgt01072028	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
524	cgt01072029	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
525	cgt01081992	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
526	cgt01081993	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
527	cgt01081994	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
528	cgt01081995	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
529	cgt01081996	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
530	cgt01081997	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
531	cgt01081998	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
532	cgt01081999	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR

533	cgt01082001	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
534	cgt01082002	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
535	cgt01082003	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
536	cgt01082004	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
537	cgt01082005	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
538	cgt01082006	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
539	cgt01082007	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
540	cgt01082008	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
541	cgt01082009	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
542	cgt01082010	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
543	cgt01082011	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
544	cgt01082012	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
545	cgt01082013	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR

546	cgt01082014	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
547	cgt01082015	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
548	cgt01082016	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
549	cgt01082017	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
550	cgt01082018	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
551	cgt01082019	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
552	cgt01082020	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
553	cgt01082021	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
554	cgt01082022	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
555	cgt01082023	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
556	cgt01082024	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
557	cgt01082025	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
558	cgt01082026	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR

559	cgt01082027	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
560	cgt01082028	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
561	cgt01082029	TIPOS DE COSTOS RELACIONADOS CON EL CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
562	clientes_vino	Datos de colegiados / Datos existentes en las tablas mst01cli y tbl01cac	8117	SR	SR	SR	SR	SR	SR	SR	SR	SR
563	CodigoBarras	Descripcion del codigo de barras con su fecha respectiva	856	SR	SR	SR	SR	SR	SR	SR	SR	SR
564	COLEG	Codigo y nombre de cada capitulo que le pertenece a cada colegiado	14	SR	SR	SR	SR	SR	SR	SR	SR	SR
565	comparacionFinalMayo_Octubre	Datos de colegiados / Datos existentes en las tablas mst01cli y tbl01cac	197	SR	SR	SR	SR	SR	SR	SR	SR	SR
566	daot	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
567	Daot1	SIN RELEVANCIA	52	SR	SR	SR	SR	SR	SR	SR	SR	SR
568	data_web_09012014	Descripcion del colegiado con datos generales de cada uno	7940	I	D	I	SR	SR	SR	D	SR	
569	db_flujocaja_2008	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
570	db_flujocaja_2009	SIN RELEVANCIA	45	SR	SR	SR	SR	SR	SR	SR	SR	SR
571	db_flujocaja_2010	SIN RELEVANCIA	1017	SR	SR	SR	SR	SR	SR	SR	SR	SR
572	db_flujocaja_2011	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
573	db_flujocaja_2012	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
574	db_flujocaja_2013	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
575	db_flujocaja_2014	SIN RELEVANCIA	22164	SR	SR	SR	SR	SR	SR	SR	SR	SR

576	db_flujocaja_2015	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
577	db_flujocaja_2016	SIN RELEVANCIA	18082	SR	SR	SR	SR	SR	SR	SR	SR	SR
578	db_flujocaja_2017	SIN RELEVANCIA	11451	SR	SR	SR	SR	SR	SR	SR	SR	SR
579	db_flujocaja_2018	SIN RELEVANCIA	17024	SR	SR	SR	SR	SR	SR	SR	SR	SR
580	detmst01cobanul	LISTADO DE TRAMITES REALIZADOS POR COLEGIADOS	2575	SR	SR	SR	SR	SR	SR	SR	SR	SR
581	dga0000	DESCRIPCION DE SOFTWARES Y OTROS COMPONENTES	11595	SR	SR	SR	SR	SR	SR	SR	SR	SR
582	dnava	Listado de pagos mensuales de colegiados	61619	SR	SR	SR	D	SR	I	SR	SR	SR
583	dnava-ant	Listado de pagos mensuales de colegiados / datos antiguos presentes en la tabla dnava	58718	SR	SR	SR	SR	SR	SR	SR	SR	SR
584	drolly	Listado de pagos mensuales de colegiados	48012	I	SR	SR	D	SR	I	SR	SR	SR
585	dtl_anexocon	CODIGO DE CUENTAS DEL COLEGIO CON SUS ANEXOS	901	SR	SR	SR	SR	SR	SR	SR	SR	SR
586	Dtl_artpro	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
587	Dtl_Asistencia	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
588	dtl_audi_lispre	Actualizaciones en el nuevo sistema de NavaSoft (algunos registros)	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
589	DTL_Cargo_awb	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
590	dtl_certificado_hab	Descripcion de habilidad de los solicitantes y fechas vigentes	94276	SR	SR	SR	SR	SR	SR	SR	SR	SR

591	dtl_cliente_espe	Codigo de colegiado con su codigo de especialidad y capitulo al que pertenece, presentes en la tabla mst01cli	13690	SR	SR	SR	SR	SR	SR	SR	SR	SR
592	Dtl_ComMat	Inventario de algunos equipos del colegio con sus respectivas características	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
593	Dtl_contacto	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
594	Dtl_curricula	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
595	DTL_DepAct	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
596	Dtl_DetCom	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
597	dtl_flujocaja_licencia	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
598	dtl_GesCob	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
599	dtl_gridglosa	Codigo de productos con su repctivo numero de documento y cuenta a la que a sido abonada	97121	SR	SR	SR	SR	SR	SR	SR	SR	SR
600	Dtl_Lectura	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
601	DTL_ManAct	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
602	dtl_mensajeria	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
603	DTL_MovAct	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
604	DTL_ObrasAct	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
605	Dtl_Ocurrencia	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
606	dtl_ordpro	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
607	dtl_ordprogas	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
608	dtl_ordprores	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
609	dtl_pack_gene	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

610	dtl_pcge_defcta	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
611	dtl_precios_cliente	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
612	dtl_precios_lavagna	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
613	dtl_presupuesto_caja	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
614	Dtl_Produ	SERVICIOS DE EQUIPOS DEL COLEGIO DE INGENIEROS	96	SR	SR	SR	SR	SR	SR	SR	SR	SR
615	dtl_provee_consig	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
616	dtl_provee_item	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
617	dtl_provee_precios	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
618	dtl_receta	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
619	dtl_rolcob	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
620	dtl_servconfirm	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
621	DTL_Stock	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
622	Dtl_Tmp_Inven	SIN RELEVANCIA	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
623	Dtl01Apl	Ingresos al colegio de ingenieros no detallados especificamente	210604	SR	SR	SR	SR	SR	SR	SR	SR	SR
624	dtl01bco	TRANSACCIONES DEL COLEGIO DEL INGENIEROS	3081	SR	SR	SR	SR	SR	SR	SR	SR	SR
625	dtl01ccc	Ingresos por solicitud de documentos e conferencias que se brinan	2575	SR	SR	SR	SR	SR	SR	SR	SR	SR
626	dtl01cch	Empresas que estan vinculadas con el colegio de ingenieros (gastos del colegio)	9193	SR	SR	SR	SR	SR	SR	SR	SR	SR
627	dtl01cco	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

628	dtl01ccp	Ingresos y gastos del colegio de ingenieros	75971	SR	SR	SR	SR	SR	SR	SR	SR	SR
629	dtl01cfa	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
630	dtl01che	Proveedores que estan registrados en el colegio de ingenieros	4624	SR	SR	SR	SR	SR	SR	SR	SR	SR
631	dtl01cob	Pagos echos por el Colegio de Ingenieros	818967	SR	SR	SR	SR	SR	SR	SR	SR	SR
632	dtl01com	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
633	Dtl01Con	Datos de Algunos Familiares de Colegiados	8906	SR	SR	SR	SR	I	SR	SR	SR	SR
634	dtl01cot	SIN RELEVANCIA - SIN DATOS	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
635	dtl01die	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
636	dtl01erc	Gastos pagados por el Colegio de Ingeniero	1586	SR	SR	SR	SR	SR	SR	SR	SR	SR
637	dtl01fac	Pagos inconsistentes del CIP	98520	SR	SR	SR	SR	SR	SR	SR	SR	SR
638	dtl01gfa	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
639	dtl01gim	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
640	dtl01gsm	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
641	dtl01gui	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
642	Dtl01guiOlb	SIN RELEVANCIA - SIN DATOS	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
643	dtl01hch	Compas echas por el Colegio de Ingeniero	4614	SR	SR	SR	SR	SR	SR	SR	SR	SR
644	dtl01itm	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
645	dtl01lch	Liquidaciones echas	201	SR	SR	SR	SR	SR	SR	SR	SR	SR
646	dtl01nfa	SIN RELEVANCIA - SIN DATOS	17	SR	SR	SR	SR	SR	SR	SR	SR	SR
647	dtl01ocm	SIN RELEVANCIA - SIN DATOS	2	SR	SR	SR	SR	SR	SR	SR	SR	SR

648	dtl01opm	Deudas por el Colegio de Ingenieros	2300	SR	SR	SR	SR	SR	SR	SR	SR	SR
649	dtl01pac	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
650	dtl01ped	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
651	Dtl01PGA	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
652	dtl01ple	Deudas por el Colegio de Ingenieros	3790	SR	SR	SR	SR	SR	SR	SR	SR	SR
653	dtl01pli	Deudas por el Colegio de Ingenieros	323	SR	SR	SR	SR	SR	SR	SR	SR	SR
654	dtl01pmo	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
655	dtl01pre	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
656	dtl01rco	Gastos pagados por el Colegio de Ingeniero	40695	SR	SR	SR	SR	SR	SR	SR	SR	SR
657	dtl01rec	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
658	dtl01res	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
659	dtl01rle	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
660	dtl01seg	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
661	dtl01sri	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
662	dtlfac	Pagos de Colegiados	143	SR	SR	SR	SR	SR	SR	SR	SR	SR
663	dtproperties	Nombre e Bases de Datos - Sin Importancia	21	SR	SR	SR	SR	SR	SR	SR	SR	SR
664	Efectivo_Dia_Caja	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
665	Egresos_No_Provisionados	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
666	espec	Codigos de Especialidad	41	SR	SR	SR	SR	SR	SR	SR	SR	SR
667	EstDtIFac	Lista De Colegios de Ingenieros	1629721	SR	SR	SR	SR	SR	SR	SR	SR	SR

668	EstDtlRco	Pagos de Comisiones del Colegio de Ingenieros	40695	SR	SR	SR	SR	SR	SR	SR	SR	SR
669	EstMstRco	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
670	exportEmailNava	Datos de Algunos Colegiados / datos presentes en la tabla mst01cli	9924	SR	SR	SR	SR	SR	SR	SR	SR	SR
671	FacturasDelDia	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
672	fami_choco2013	Datos de Familias de Colegiados / datos presentes en la tabla dtl01con	527	SR	SR	SR	SR	SR	SR	SR	SR	SR
673	familias2013	Datos de Familias de Colegiados / datos presentes en la tabla dtl01con	2531	SR	SR	SR	SR	SR	SR	SR	SR	SR
674	fcu0000	USUARIOS DE EMPRESAS EN EL CIP	80	SR	SR	SR	SR	SR	SR	SR	SR	SR
675	frx0101	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
676	fto0101e	Reportes	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
677	fto0101o	Reportes	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
678	fto0102e	Reportes	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
679	fto0102o	Reportes	48	SR	SR	SR	SR	SR	SR	SR	SR	SR
680	GrossDtlFac	Diversos Tramites realizados por empresas y colegiados	191817	SR	SR	SR	SR	SR	SR	SR	SR	SR
681	habil2204201	ESTADOS DE HABILES Y NO HABILES DE COLEGIADOS (NO ACTUALIZADOS)	9102	SR	SR	SR	SR	SR	SR	SR	SR	SR
682	informe	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
683	iss_cnn012013	Listado de Colegiados (colegiados presentes en mst01cli)	2120	SR	SR	SR	SR	SR	SR	SR	SR	SR
684	kdd0101	Documentos echos por Colegiados	785	SR	SR	SR	SR	SR	SR	SR	SR	SR
685	leame	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

686	lpm0101e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
687	lpm0102e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
688	MST_Activo	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
689	Mst_awb	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
690	Mst_curricula	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
691	Mst_Matriz	Gastos pagados por el Colegio de Ingeniero	32	SR	SR	SR	SR	SR	SR	SR	SR	SR
692	mst_ordpro	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
693	Mst_Produ	Gastos pagados por el Colegio de Ingeniero	26	SR	SR	SR	SR	SR	SR	SR	SR	SR
694	Mst_rolcob	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
695	mst0101e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
696	mst0101o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
697	mst0102e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
698	mst0102o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
699	mst0103e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
700	mst0103o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
701	mst0104e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
702	mst0104o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
703	mst0105e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
704	mst0105o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
705	mst0106e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
706	mst0106o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

707	mst0107e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
708	mst0107o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
709	mst0108e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
710	mst0108o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
711	mst0109e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
712	mst0109o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
713	mst0110a	SIN RELEVANCIA - SIN DATOS	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
714	mst0110e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
715	mst0110o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
716	Mst0111a	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
717	mst0111e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
718	mst0111o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
719	mst0112a	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
720	mst0112e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
721	mst0112o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
722	mst0113e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
723	mst0113o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
724	mst0114e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
725	mst0114o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
726	mst0115e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
727	mst0115o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
728	mst0116e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR

729	mst0116o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
730	mst0117e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
731	mst0117o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
732	mst0118e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
733	mst0118o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
734	mst0120e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
735	mst0120o	SIN RELEVANCIA - SIN DATOS - Fechas	1352	SR	SR	SR	SR	SR	SR	SR	SR	SR
736	mst0121e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
737	mst0121o	Tramites Pagados	45	SR	SR	SR	SR	SR	SR	SR	SR	SR
738	mst0122e	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
739	mst0122o	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
740	mst01bco	Listado de Diferentes Tramites de Colegiados	33662	SR	SR	SR	SR	SR	SR	SR	SR	SR
741	mst01buz	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
742	mst01ccc	Pagos de Colegiados desde el 2001 hasta el 2029	1763275	SR	SR	SR	D	SR	I	SR	SR	SR
743	mst01cch	Ingresos Y Egresos del Colegio	8063	SR	SR	SR	SR	SR	SR	SR	SR	SR
744	mst01ccp	Cuotas o Donaciones	37524	SR	SR	SR	SR	SR	SR	SR	SR	SR
745	mst01chd	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
746	mst01che	Pagos echos por el Colegio de Ingenieros	1712	SR	SR	SR	SR	SR	SR	SR	SR	SR
747	mst01cli	Listado de Colegiados	19356	I	D	I	SR	SR	SR	D	D	SR
748	mst01cli_web14012014	Listado de Colegiados(Datos presentes en la tabla mst01cli)	1812	SR	SR	SR	SR	SR	SR	SR	SR	SR

749	mst01cli09122013	Listado de Colegiados(Datos presentes en la tabla mst01cli)	12035	SR	SR	SR	SR	SR	SR	SR	SR	SR
750	mst01cli29112013	Listado de Colegiados(Datos presentes en la tabla mst01cli)	11998	SR	SR	SR	SR	SR	SR	SR	SR	SR
751	mst01cob	Certificados echas Por el Cip	367734	SR	SR	SR	SR	SR	SR	SR	SR	SR
752	mst01cobanul	Certificados echas Por el Cip	1108	SR	SR	SR	SR	SR	SR	SR	SR	SR
753	mst01com	SIN RELEVANCIA - SIN DATOS	3527	SR	SR	SR	SR	SR	SR	SR	SR	SR
754	Mst01con	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
755	mst01cot	SIN RELEVANCIA - SIN DATOS	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
756	mst01daot	Remuneraciones según el RUC	96	SR	SR	SR	SR	SR	SR	SR	SR	SR
757	mst01erc	Recibos dados por el Colegio	389	SR	SR	SR	SR	SR	SR	SR	SR	SR
758	mst01fac	Pagos hacia el Colegio	96707	SR	SR	SR	SR	SR	SR	SR	SR	SR
759	mst01fax	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
760	mst01gim	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
761	mst01gsm	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
762	mst01gui	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
763	mst01guiOlb	SIN RELEVANCIA - SIN DATOS	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
764	mst01ocm	SIN RELEVANCIA - SIN DATOS	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
765	mst01opm	GASTOS PRESTAMOS Y RECIBOS DEL CIP	2109	SR	SR	SR	SR	SR	SR	SR	SR	SR
766	mst01ped	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
767	mst01pro	Se describen datos de algunos proveedores de los cuales hace uso el CIP	4458	SR	SR	SR	SR	SR	SR	SR	SR	SR

768	mst01rco	Se describen datos de proveedores de los cuales hace uso el CIP	37131	SR	SR	SR	SR	SR	SR	SR	SR	SR
769	mst01rec	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
770	Mst01Ret	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
771	Mst01snt	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
772	mst01tcr	Detalla deudas de proveedores hacia el CIP	226	SR	SR	SR	SR	SR	SR	SR	SR	SR
773	mst01var	Contiene ruc, nombre y direccion de proveedores.	467	SR	SR	SR	SR	SR	SR	SR	SR	SR
774	persona\$	Contiene un listado de algunos colegiados, esto se deduce porque todos contienen numero de CIP (datos presentes en mst01cli)	627	SR	SR	SR	SR	SR	SR	SR	SR	SR
775	pla0000	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
776	prd0101	Es un listado de pagos que se realizan en el CIP	47	SR	SR	SR	SR	SR	SR	SR	SR	SR
777	prove	Listado de Proveedores, solo contiene rc y nombre de proveedor, son pocos registros.	164	SR	SR	SR	SR	SR	SR	SR	SR	SR
778	psn0100	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
779	Ranking_Vend_Fac	Es una vista que se encuentra en la base de datos, la cual brinda informacion acerca de ingresos de dinero que tiene el CIP, por diversos motivos, entre algunos: Alquiler de auditorios, cuotas de sonidos, etc.	198520	SR	SR	SR	SR	SR	SR	SR	SR	SR

780	reportes	Listado de consultas sql, las cuales, al ejecutarse, muestran reportes, pero casi todas sin registro alguno al ser ejecutadas.	37	SR	SR	SR	SR	SR	SR	SR	SR	SR
781	ric_35_2804	Almacena registros de pagos que se realizan para el CIP, por ejemplo: CANCELACION DE DOCUMENTOS, CERTIFICADOS,etc.	9934	SR	SR	SR	SR	SR	SR	SR	SR	SR
782	roly	Contiene pagos de 1911 colegiados, detallando su numero de CIP, su nombre, la fecha de cancelacion y el monto pagado.	244666	I	SR	SR	D	SR	I	SR	SR	
783	sysdiagrams	Diagramas Entidad - Relacion	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
784	SysNavaAudi	Almacena todas las transacciones echas por Navasoft.	1049939	SR	SR	SR	SR	SR	SR	SR	SR	SR
785	SysNavaError	Registros que hacen referencias a errores en transacciones en la base de datos	7585	SR	SR	SR	SR	SR	SR	SR	SR	SR
786	SysNavaFK	Registros de llaves foraneas de tablas de la base de datos.	136	SR	SR	SR	SR	SR	SR	SR	SR	SR
787	SysNavaLogin	Registros que contienen usuario y maquina como columnas mas intuitivas y entendibles.	83755	SR	SR	SR	SR	SR	SR	SR	SR	SR
788	SysNavaPk	Sin Relevancia	31	SR	SR	SR	SR	SR	SR	SR	SR	SR
789	SysNavaUser	Registros que contienen usuario y maquina como columnas mas intuitivas y entendibles.	33	SR	SR	SR	SR	SR	SR	SR	SR	SR
790	tbl_act_navidad	Listado de Personas asistentes a actividades de navidad	51442	SR	SR	SR	SR	SR	SR	SR	SR	SR
791	TBL_Agente_awb	Sin Relevancia	2	SR	SR	SR	SR	SR	SR	SR	SR	SR

792	tbl_ajuste	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
793	tbl_alternativa	Sin Relevancia	11	SR	SR	SR	SR	SR	SR	SR	SR	SR
794	Tbl_Amnistia_Anual	Sin Relevancia	19	SR	SR	SR	SR	SR	SR	SR	SR	SR
795	Tbl_AnexoCon	Sin Relevancia	127	SR	SR	SR	SR	SR	SR	SR	SR	SR
796	tbl_anexocon_pn	Sin Relevancia	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
797	Tbl_Aporte_ConsejoNac	Sin Relevancia	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
798	TBL_BajasAct	Sin Relevancia	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
799	tbl_cajamayor	Listado de Cajas de Pago	8	SR	SR	SR	SR	SR	SR	SR	SR	SR
800	tbl_capitulo	Listado de Capítulos del CIP	16	SR	SR	SR	SR	SR	SR	I	SR	SR
801	TBL_Cargo_awb	Sin Relevancia	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
802	TBL_CategAct	Categoría de Activos	17	SR	SR	SR	SR	SR	SR	SR	SR	SR
803	TBL_CCostoAct	Listado de oficinas presentes en el Cip	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
804	tbl_cg_addcol	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
805	tbl_cg_aduana	Listado de Lugares de Aduana	30	SR	SR	SR	SR	SR	SR	SR	SR	SR
806	tbl_cg_catitem	Listado de Categorías de Productos	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
807	tbl_cg_dni	Tipos de Documento de Identidad	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
808	tbl_cg_glosainv	Listado de Actividades que generalmente se realizan en el CIP	17	SR	SR	SR	SR	SR	SR	SR	SR	SR
809	tbl_cg_intangible	Contiene 3 registros, que les llaman intangibles	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
810	tbl_cg_libro	Categorías de Libros que el CIP utiliza para guardar registros de Ingresos y Egresos de Dinero.	31	SR	SR	SR	SR	SR	SR	SR	SR	SR

811	tbl_cg_mediopago	Tipos de formas de realizacion de pagos.	20	SR	SR	SR	SR	SR	SR	SR	SR	SR
812	tbl_cg_plancta	Sin Relevancia	35	SR	SR	SR	SR	SR	SR	SR	SR	SR
813	tbl_cg_tipbco	Listado de bancos	33	SR	SR	SR	SR	SR	SR	SR	SR	SR
814	tbl_cg_tipdoc	listado de tipos de documentos emitidos que emplean en el CIP	48	SR	SR	SR	SR	SR	SR	SR	SR	SR
815	tbl_cg_umdiv	Listado de tipos de unidades de inventarios	16	SR	SR	SR	SR	SR	SR	SR	SR	SR
816	TBL_ciudad	Sin Relevancia	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
817	tbl_clave_especial	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
818	tbl_clave_motivo	Sin Relevancia	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
819	TBL_Comision	Sin Relevancia	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
820	TBL_Commodity	Sin Relevancia	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
821	Tbl_Conprepla	Sin Relevancia	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
822	TBL_Continente	Listado de Continentes	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
823	Tbl_CorMatriz	Sin Relevancia	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
824	tbl_desktop_agenda	Sin Relevancia	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
825	tbl_desktop_agente	Listado de mensajes enviados al usar el software y quedaron grabados en la base de datos.	1857	SR	SR	SR	SR	SR	SR	SR	SR	SR
826	tbl_desktop_apuntes	Sin Relevancia	28	SR	SR	SR	SR	SR	SR	SR	SR	SR
827	tbl_desktop_buscar	Listado de Mensajes predeterminados que debe de brindar el software que utiliza el CiP a ciertas operaciones que se quieren realizar en dicho programa.	14	SR	SR	SR	SR	SR	SR	SR	SR	SR

828	tbl_desktop_estad	Listado de Mensajes predeterminados que debe de brindar el software que utiliza el CiP a ciertas operaciones que se quieren realizar en dicho programa.	80	SR	SR	SR	SR	SR	SR	SR	SR	SR
829	tbl_desktop_favorito	Listado de Mensajes predeterminados que debe de brindar el software que utiliza el CiP a ciertas operaciones que se quieren realizar en dicho programa.	41	SR	SR	SR	SR	SR	SR	SR	SR	SR
830	tbl_desktop_visita	Listado de Mensajes predeterminados que debe de brindar el software que utiliza el CiP a ciertas operaciones que se quieren realizar en dicho programa.	403	SR	SR	SR	SR	SR	SR	SR	SR	SR
831	tbl_detraccion	Listado de productos y Servicios.	40	SR	SR	SR	SR	SR	SR	SR	SR	SR
832	TBL_DptoAct	Listado de Departamentos del CiP	8	SR	SR	SR	SR	SR	SR	SR	SR	SR
833	tbl_escala_ch	Sin Relevancia	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
834	tbl_especialidad	Listado de Especialidades, menciona tambien el capitulo al que pertenece cada especialidad	98	SR	SR	SR	SR	SR	SR	D	SR	SR
835	Tbl_Fotos	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
836	Tbl_grupo_Empre	Sin Relevancia	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
837	tbl_insumo	Listado de Insumos	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
838	TBL_Larea_awb	Sin Relevancia	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
839	TBL_Moneda	Listado de Tipos de Monedas	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
840	Tbl_motivo_anulacion	Datos relacionados con software del CIP	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
841	TBL_MOTIVO_CIP	Motivos de pagos adicionales del CIP	4	SR	SR	SR	SR	SR	SR	SR	SR	SR

842	tbl_motivo_opm	Motivos de Ingreso y Egreso de dinero del CIP	20	SR	SR	SR	SR	SR	SR	SR	SR	SR
843	Tbl_notacontable	Sin Relevancia	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
844	tbl_ocurrencia	Sin Relevancia	8	SR	SR	SR	SR	SR	SR	SR	SR	SR
845	TBL_PARAMETRO	Sin Relevancia	49	SR	SR	SR	SR	SR	SR	SR	SR	SR
846	tbl_parametros	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
847	Tbl_Parentesco	Listado de Parentesco que pueda existir entre un familiar y un colegiado.	8	SR	SR	SR	SR	I	SR	SR	SR	SR
848	tbl_pcge_catccp	Listado de descripcion de cuentas por cobrar	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
849	tbl_pcge_clasif	Sin Relevancia	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
850	tbl_plantilla_dtlgasto	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
851	tbl_plantilla_gasto	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
852	tbl_proceso	Sin Relevancia	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
853	TBL_Producto	Sin Relevancia	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
854	tbl_residuo	Sin Relevancia	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
855	tbl_subdiario	Sin Relevancia	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
856	Tbl_Sucursal	Sin Relevancia	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
857	tbl_sunatlgv	Sin Relevancia	13	SR	SR	SR	SR	SR	SR	SR	SR	SR
858	TBL_tcambio_awb	Sin Relevancia	11	SR	SR	SR	SR	SR	SR	SR	SR	SR
859	Tbl_TcDefault	Listados de datos para un aplicativos que utilizan en el CIP	22	SR	SR	SR	SR	SR	SR	SR	SR	SR
860	tbl_tienda	Sin Relevancia	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
861	Tbl_TipMatriz	Sin Relevancia	3	SR	SR	SR	SR	SR	SR	SR	SR	SR

862	Tbl_TipSer	Sin Relevancia	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
863	Tbl_Turno	Listado de Turnos (Hora Entrada y Hora Salida)	15	SR	SR	SR	SR	SR	SR	SR	SR	SR
864	tbl_universidad	Listado de Universidades	33	SR	SR	SR	SR	SR	SR	SR	SR	SR
865	TBL_ZonaAct	Sin Relevancia	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
866	tbl00con	Listado de Años	37	SR	SR	SR	SR	SR	SR	SR	SR	SR
867	tbl0101a	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
868	tbl0101e	Sin Relevancia	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
869	tbl0101o	Sin Relevancia	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
870	tbl0102a	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
871	tbl0102e	Sin Relevancia	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
872	tbl0102o	Sin Relevancia	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
873	tbl0103a	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
874	tbl0103e	Sin Relevancia	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
875	tbl0103o	Sin Relevancia	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
876	tbl0104a	Sin Relevancia - Sin Datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
877	tbl0104e	Listado de Carreras	42	SR	SR	SR	SR	SR	SR	SR	I	SR
878	tbl0104o	Listado de Cargos de Areas del Cip	8	SR	SR	SR	SR	SR	SR	SR	SR	SR
879	tbl0105e	Listado de pagos a trabajadores (SUELDO BASICO, BONIFICACION, etc)	31	SR	SR	SR	SR	SR	SR	SR	SR	SR
880	tbl0105o	Listado de pagos a trabajadores (SUELDO BASICO, BONIFICACION, etc)	30	SR	SR	SR	SR	SR	SR	SR	SR	SR
881	tbl0106e	Descuento de pagos a trabajadores (ESSALUD, AFP, etc)	21	SR	SR	SR	SR	SR	SR	SR	SR	SR

882	tbl0106o	Descuento de pagos a trabajadores (ESSALUD, AFP, etc)	17	SR	SR	SR	SR	SR	SR	SR	SR	SR
883	tbl0107e	Listado de rubros	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
884	tbl0107o	Listado de rubros	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
885	tbl0108e	Lista de Codigo de rubro y su formula para cada usuario	42	SR	SR	SR	SR	SR	SR	SR	SR	SR
886	tbl0108o	Lista de Codigo de rubro y su formula para cada usuario	32	SR	SR	SR	SR	SR	SR	SR	SR	SR
887	tbl0109e	Lista de departamento	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
888	tbl0109o	Lista de departamento	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
889	tbl0110e	Lista de seguro AFP	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
890	tbl0110o	Lista de seguro AFP	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
891	tbl0111e	Lista de seguro AFP	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
892	tbl0111o	Lista de seguro AFP	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
893	tbl0112o	Lista de distritos	22	SR	SR	SR	SR	SR	SR	SR	SR	SR
894	tbl0113o	Lista de provincias	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
895	tbl0114o	Lista de departamento	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
896	tbl0115o	Lista de pais	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
897	tbl0116e	Lista de contacto	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
898	tbl0116o	Lista de contacto	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
899	tbl0117e	Nivel de educacion	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
900	tbl0117o	Nivel de educacion	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
901	tbl0118e	Catagoria de estado civil	5	SR	SR	SR	SR	SR	SR	SR	SR	SR

902	tbl0118o	Catagoria de estado civil	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
903	tbl0119e	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
904	tbl0119o	Gastos del CIP	776	SR	SR	SR	SR	SR	SR	SR	SR	SR
905	tbl0120e	Tipo de contrato	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
906	tbl0120o	Tipo de contrato	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
907	tbl0121e	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
908	tbl0121o	Sin relevancia	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
909	tbl0122e	Sin relevancia	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
910	tbl0122o	Sin relevancia	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
911	tbl0123e	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
912	tbl0123o	Detalle de año por mes	36	SR	SR	SR	SR	SR	SR	SR	SR	SR
913	tbl0124e	Detalle de cada mes (proximo mes, cierre de mes)	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
914	tbl0124o	Lista de periodo por mes	897	SR	SR	SR	SR	SR	SR	SR	SR	SR
915	tbl0125e	Lista de turno	13	SR	SR	SR	SR	SR	SR	SR	SR	SR
916	tbl0126e	Sin relevancia (codocu, desocu [faltas, descanso medico], verlec, afedsc, ediocu)	28	SR	SR	SR	SR	SR	SR	SR	SR	SR
917	tbl0126o	Sin relevancia (codocu, desocu [faltas, descanso medico], verlec, afedsc, ediocu)	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
918	Tbl0127e	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
919	tbl01act	Sin relevancia (nomact: NINGUNO)	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
920	tbl01alm	Sin relevancia (Lista de almacenamiento)	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
921	TBL01AYU	Lista de ayudantes	7	SR	SR	SR	SR	SR	SR	SR	SR	SR

922	tbl01bco	Lista de bancos	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
923	tbl01cac	Lista de categoria (ORDINARIO, VITALICIO)	8	SR	SR	SR	SR	SR	SR	SR	SR	D
924	tbl01cai	Lista de categoria (MERCADERIA, SERVICIOS)	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
925	tbl01caj	Sin relevancia (SOLO UNA COLUMNA PERO VACIA)	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
926	tbl01cap	Lista de categoria (LOCALES, EXTERIOR)	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
927	tbl01ccc	Lista de pagos por banco realizado por los clientes	18	SR	SR	SR	SR	SR	SR	SR	SR	SR
928	tbl01cch	Lista de caja	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
929	Tbl01cci	Lista por año, mes y codigo de almacen	93	SR	SR	SR	SR	SR	SR	SR	SR	SR
930	TBL01CCP	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
931	tbl01cda	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
932	tbl01cdc	Gastos del colegio de ingenieros	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
933	tbl01cdv	Tipos de pago (contado, credito)	3	SR	SR	SR	I	SR	SR	SR	SR	SR
934	tbl01Coc	Tipos de Familiares del colegiado / tabla no utilizada por ninguna otra tabla, solo usan la tabla tbl_parentesco	5	SR	SR	SR	SR	SI	SR	SR	SR	SR
935	tbl01com	Sin relevancia	71	SR	SR	SR	SR	SR	SR	SR	SR	SR
936	tbl01cor	Sin relevancia	105	SR	SR	SR	SR	SR	SR	SR	SR	SR
937	tbl01cpo	Sin relevancia	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
938	tbl01cta	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
939	tbl01dep	Nombres de Departamentos del Perú	25	SR	SR	SR	SR	SR	SR	SR	SR	SR

940	tbl01det	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
941	tbl01dis	Distritos del Perú	1828	SR	SR	SR	SR	SR	SR	SR	SR	SR
942	tbl01doc	Nombres de Tipos de Documentos	71	SR	SR	SR	SR	SR	SR	SR	SR	SR
943	tbl01erc	Nomrbes de Usuario	99	SR	SR	SR	SR	SR	SR	SR	SR	SR
944	tbl01est	Sin relevancia	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
945	tbl01est_subest	Sin relevancia	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
946	tbl01fam	sin relevancia	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
947	tbl01fba	Glosario del Cip	42	SR	SR	SR	SR	SR	SR	SR	SR	SR
948	tbl01fdp	Sin relevancia	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
949	Tbl01Fic	Sin relevancia	237	SR	SR	SR	SR	SR	SR	SR	SR	SR
950	tbl01for	Sin relevancia	8	SR	SR	SR	SR	SR	SR	SR	SR	SR
951	tbl01gli	Sin relevancia	15	SR	SR	SR	SR	SR	SR	SR	SR	SR
952	tbl01gls	Sin relevancia	20	SR	SR	SR	SR	SR	SR	SR	SR	SR
953	tbl01glv	Sin relevancia	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
954	tbl01gna	Sin relevancia	8	SR	SR	SR	SR	SR	SR	SR	SR	SR
955	tbl01grp	Sin relevancia	11	SR	SR	SR	SR	SR	SR	SR	SR	SR
956	tbl01itm	Tipos de servicios del CIP	48	SR	SR	SR	SR	SR	SR	SR	SR	SR
957	tbl01lca	sin relevancia - son datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
958	tbl01lch	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
959	tbl01mar	Sin relevancia (CODMAR, NOMMAR, ABMAR)	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
960	tbl01pai	Lista de pais	4	SR	SR	SR	SR	SR	SR	SR	SR	SR

961	tbl01pcs	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
962	tbl01pos	Lista de cursos dictado por el CIP	269	SR	SR	SR	SR	SR	SR	SR	SR	SR
963	Tbl01Pre	Sin relevancia (CODPRE, CODCLI: NULL, NOMPRES)	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
964	tbl01pro	Lista de provincias	194	SR	SR	SR	SR	SR	SR	SR	SR	SR
965		Lista de cajas del CIP	12	SR	SR	SR	SR	SR	SR	SR	SR	SR
966	Tbl01pyt	Sin relevancia (CODPYT, DESPUT, PROPYT)	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
967	tbl01sbf	Servicios del CIP	8	SR	SR	SR	SR	SR	SR	SR	SR	SR
968	tbl01scc	Gastos del CIP	1217	SR	SR	SR	SR	SR	SR	SR	SR	SR
969	tbl01tar	Lista de tarjetas	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
970	tbl01tca	Tasa de compra y venta del dolar por año y mes	12432	SR	SR	SR	SR	SR	SR	SR	SR	SR
971	tbl01tcl	Lista de CD por departamento (CD LAMBAYEQUE, CD AMAZONAS, etc)	29	SR	SR	SR	SR	SR	SR	SR	SR	SR
972	tbl01tra	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
973	tbl01umd	Unidad de medicion	14	SR	SR	SR	SR	SR	SR	SR	SR	SR
974	tbl01ven	Lista de nombres de ventanilla	44	SR	SR	SR	I	SR	SR	SR	SR	SR
975	tbl01vta	Gastos del CIP	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
976	tbl01zon	Lista de Zona	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
977	tblColCap	Lista de colegiados, datos presentes en mst01cli	49	SR	SR	SR	SR	SR	SR	SR	SR	SR
978	tblColDir	Lista de colegiados, datos presentes en mst01cli	26	SR	SR	SR	SR	SR	SR	SR	SR	SR

979	TblRes	SIN RELEVANCIA - SIN DATOS	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
980	tbltxt	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
981	temp_col	Lista de colegiados	8422	SR	SR	SR	SR	SR	SR	SR	SR	SR
982	tipo_tbl_act	Servicio del CIP	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
983	tmp_adel	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	13261	SR	SR	SR	SR	SR	SR	SR	SR	SR
984	tmp_asientos2016	Sin Relevancia	54	SR	SR	SR	SR	SR	SR	SR	SR	SR
985	tmp_CA2017	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	22354	SR	SR	SR	SR	SR	SR	SR	SR	SR
986	tmp_CA2018	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	22354	SR	SR	SR	SR	SR	SR	SR	SR	SR
987	tmp_CA2019	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	136	SR	SR	SR	SR	SR	SR	SR	SR	SR
988	tmp_CA2020	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	8	SR	SR	SR	SR	SR	SR	SR	SR	SR
989	TMP_CCAN_2015AT	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	20964	SR	SR	SR	SR	SR	SR	SR	SR	SR
990	tmp_cob_ago	Lista Pagos de colegiados en cuanto a tramites realizados	1811	SR	SR	SR	SR	SR	SR	SR	SR	SR
991	tmp_cob1407_01	Lista Pagos de colegiados en cuanto a tramites realizados	1415	SR	SR	SR	SR	SR	SR	SR	SR	SR
992	tmp_CtaAde	Lista Pagos de colegiados en cuanto a tramites realizados	2412	SR	SR	SR	SR	SR	SR	SR	SR	SR
993	tmp_ctapn	Lista de cuenta	24	SR	SR	SR	SR	SR	SR	SR	SR	SR
994	tmp_cxc	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	552397	SR	SR	SR	SR	SR	SR	SR	SR	SR

995	tmp_subccosto	Lista de sucursales	18	SR	SR	SR	SR	SR	SR	SR	SR	SR
996	TMPACN0000OLD	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
997	TmpAnaPro	Sin relevancia (AÑO:2016, MES: DICIEMBRE)	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
998	TmpBuscarSensitive	Sin relevancia (medina, medin)	2	SR	SR	SR	SR	SR	SR	SR	SR	SR
999	update_18	Lista de cliente mas su celular y telefono	7934	SR	SR	SR	SR	SR	SR	SR	SR	SR
1000	V_COBCUOTA	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	674130	SR	SR	SR	SR	SR	SR	SR	SR	SR
1001	V_COBINGDIV	Lista de colegiados con certificado de habilidad (No esta actualizado)	101960	SR	SR	SR	SR	SR	SR	SR	SR	SR
1002	V_CUOMESRET	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	20038	SR	SR	SR	SR	SR	SR	SR	SR	SR
1003	v_dtlrolcob	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
1004	V_PGCUOMESADL	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	223084	SR	SR	SR	SR	SR	SR	SR	SR	SR
1005	V_PGCUOUNMESRET	Lista de nroCIP mas monto, saldo Y fecha, datos presentes en la tabla mst01ccc	20038	SR	SR	SR	SR	SR	SR	SR	SR	SR
1006	V_PLANILLAEMP	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
1007	V_PLANILAOBR	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
1008	v_ric12_35	Tramites de Colegiados	364	SR	SR	SR	SR	SR	SR	SR	SR	SR
1009	v_ric35_1104	Tramites de Colegiados	381	SR	SR	SR	SR	SR	SR	SR	SR	SR
1010	v_transeuntes	Lista de colegiado mas su categoria: VITALICIO, FALLECIDO, ETC. , prsentes en la tabla mst01cli	19356	SR	SR	SR	SR	SR	SR	SR	SR	SR
1011	VentaPtoFam	Servicios que brinda el CIP	198520	SR	SR	SR	SR	SR	SR	SR	SR	SR

1012	vista_cuotas_migradas	Lista de colegiado mas migracion de cuota ordinaria	420206	SR	SR	SR	SR	SR	SR	SR	SR	SR
1013	vs_clientes	Lista de clientes, datos en la tabla mst01cli	19358	SR	SR	SR	SR	SR	SR	SR	SR	SR
1014	Vs_Compras_Año_Tri	Sin relevancia (AÑO, T1, T2, T3, T4)	10	SR	SR	SR	SR	SR	SR	SR	SR	SR
1015	Vs_RC_Conta	Lista de proveedores	37131	SR	SR	SR	SR	SR	SR	SR	SR	SR
1016	VT_CONTACTOS	Lista de familiares de colegiados	8886	SR	SR	SR	SR	SR	SR	SR	SR	SR
1017	VT-SOCIOS	Lista de socios mas su categoria (VITALICIO, FALLECIDO, ETC)., datos presentes en la tabla mst01cli	13392	SR	SR	SR	SR	SR	SR	SR	SR	SR
1018	zbk10_psn0100	Sin relevancia	1	SR	SR	SR	SR	SR	SR	SR	SR	SR
1019	zbk10_tbl_cajamayor	Lista de caja perteneciente al CIP	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
1020	zbk10_tbl_parametro	Lista de parametros	39	SR	SR	SR	SR	SR	SR	SR	SR	SR
1021	zbk10_tbl01cac	Lista de categoria: ORDINARIO, VITALICIO, FALLECIDO. (Presente tambien en la tabla tbl01cac)	7	SR	SR	SR	SR	SR	SR	SR	SR	SR
1022	zbk10_tbl01cai	Lista de categoria: MERCADERIA, SERVICIOS, PRODUCTO TERMINADO	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
1023	zbk10_tbl01ccc	Lista de bancos: BCP, BANCO CONTINENTAL, CMAC PUIRA SAC.	9	SR	SR	SR	SR	SR	SR	SR	SR	SR
1024	zbk10_tbl01cch	Sin relevancia - Sin datos	0	SR	SR	SR	SR	SR	SR	SR	SR	SR
1025	zbk10_tbl01cdc	Sin relevancia	5	SR	SR	SR	SR	SR	SR	SR	SR	SR
1026	zbk10_tbl01cdv	Sin relevancia	3	SR	SR	SR	SR	SR	SR	SR	SR	SR
1027	zbk10_tbl01erc	Lista de usuario	37	SR	SR	SR	SR	SR	SR	SR	SR	SR
1028	zbk10_tbl01est	Sin relevancia	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
1029	zbk10_tbl01est_subest	Sin relevancia	4	SR	SR	SR	SR	SR	SR	SR	SR	SR

1030	zbk10_tbl01gli	Gastos del CIP	15	SR	SR	SR	SR	SR	SR	SR	SR	SR
1031	zbk10_tbl01gls	Servicio del CIP	20	SR	SR	SR	SR	SR	SR	SR	SR	SR
1032	zbk10_tbl01glv	Servicio del CIP	22	SR	SR	SR	SR	SR	SR	SR	SR	SR
1033	zbk10_tbl01grp	Servicio del CIP	8	SR	SR	SR	SR	SR	SR	SR	SR	SR
1034	zbk10_tbl01itm	Gastos del CIP	36	SR	SR	SR	SR	SR	SR	SR	SR	SR
1035	zbk10_tbl01sbf	Servicio del CIP	6	SR	SR	SR	SR	SR	SR	SR	SR	SR
1036	zbk10_tbl01tar	Lista de taarjetas: WIESE CASH, CREDIMAS, MASTERCARD.	4	SR	SR	SR	SR	SR	SR	SR	SR	SR
1037	zbk10_tbl01vta	Gastos del CIP	7	SR	SR	SR	SR	SR	SR	SR	SR	SR

ANEXO 09: Detallando Columnas de tablas con información relevante de la Base de Datos Colegio_Ingenieros.

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN							
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría
data_web_09012014	7940	idcolegiado	Clave primaria de la tabla.	SR	SR	SR	SR	SR	SR	SR	SR
		tipo	Sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		cip	Codigo que se le asigna a cada colegiado	SR	SR	SR	SR	SR	SR	SR	SR
		nombres	Nombres de cada colegiado	SR	SR	SR	SR	SR	SR	SR	SR
		apellidos	Apellidos de cada colegiado	SR	SR	SR	SR	SR	SR	SR	SR
		idsexo	Este campo pertenece al genero (1: Masculino, 2: Femenino)	I	SR	SR	SR	SR	SR	SR	SR
		idespecialidad	Hace referencia a la especialidad del colegiado y tiene registros 0 (cero).	SR	SR	SR	SR	SR	SR	SR	SR
		fechanac	Hace referencia a la fecha de nacimiento de los colegiados	SR	SR	I	SR	SR	SR	SR	SR
		idestadocivil	Hace referencia al estado civil de cada colegiado y esta relacionada con la tabla tbl0118e	SR	D	SR	SR	SR	SR	SR	SR
		idnacionalidad	Hace referencia a la nacionalidad de cada colegiado y tiene registros null	SR	SR	SR	SR	SR	SR	SR	SR
		dni	Hace referencia al documento de identidad de cada colegiado	SR	SR	SR	SR	SR	SR	SR	SR
		ruc	Hace referencia al número que identifica a una empresa a nivel nacional y tiene registros null.	SR	SR	SR	SR	SR	SR	SR	SR
		email1	Hace referencia al primer correo electronico	SR	SR	SR	SR	SR	SR	SR	SR
		email2	Hace referencia al segundo	SR	SR	SR	SR	SR	SR	SR	SR

			correo electronico y tiene registros null.								
	emailcip		Hace referencia al correo electronico del colegiado y tiene registros null.	SR	SR	SR	SR	SR	SR	SR	SR
	iddistrito		Hace referencia al distrito que pertenece el colegiado y esta relacionado con la tabla tbl01dis.	SR	SR	SR	SR	SR	SR	SR	SR
	direccion		Hace referencia a la direccion de cada colegiado.	SR	SR	SR	SR	SR	SR	SR	SR
	telefono		Hace referencia al telefono de cada colegiado.	SR	SR	SR	SR	SR	SR	SR	SR
	celular		Hace referencia al celular de cada colegiado.	SR	SR	SR	SR	SR	SR	SR	SR
	idsituacion		Sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR
	contrasena		Hace referencia a la contraseña asignado a cada colegiado y tiene como registros null.	SR	SR	SR	SR	SR	SR	SR	SR
	appaterno		Hace referencia al apellido paterno del colegiado.	SR	SR	SR	SR	SR	SR	SR	SR
	apmaterno		Hace referencia al apellido materno del colegiado.	SR	SR	SR	SR	SR	SR	SR	SR
	fechaactualizacion		Hace referencia a la actualizacion del nuevo sistema del cip (navasoft).	SR	SR	SR	SR	SR	SR	SR	SR
	cuo		Sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR
	fecing		Hace referencia a la fecha de ingreso del colegiado al cip.	SR	SR	SR	SR	SR	SR	SR	SR
	codcap_		Hace referencia al codigo de capitulo respecto al colegiado que se relaciona con la tabla COLEG.	SR	SR	SR	SR	SR	SR	SR	SR
	codesp_		Hace referencia al codigo de especialidad respecto al colegiado que se relaciona con la tabla espec.	SR	SR	SR	SR	SR	SR	D	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN								
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría	
dnava	61619	fecha	Fecha de apertura para que se empiece a realizar el pago de la cuota mensual	SR	SR	SR	SR	SR	SR	SR	SR	SR
		nrocip	Número de colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
		nomcli	Nombre completo del colegiado asociado al CIP	I	SR	SR	SR	SR	SR	SR	SR	SR
		monto	Monto total del pago realizado de la cuota mensual	SR	SR	SR	D	SR	SR	SR	SR	SR
		saldo	Saldo pendiente de la cuota mensual	SR	SR	SR	SR	SR	SR	SR	SR	SR
		uabo	Fecha de cancelación de pago de la cuota mensual	SR	SR	SR	SR	SR	I	SR	SR	SR
		verif	Verificación del estado de los datos en ambos sistemas	SR	SR	SR	SR	SR	SR	SR	SR	SR
		flag	Bandera de los datos (Existente, Nuevo)	SR	SR	SR	SR	SR	SR	SR	SR	SR
		codcap	Codigo que pertenece al capitulo de cada asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN								
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría	
dro lly	48 01 2	fecha	Fecha vencimiento de pago de cada asociado	SR	SR	SR	SR	SR	SR	SR	SR	SR

		Descripcion	Mes en el que pago cada asociado de acuerdo a la fecha de pago	SR	SR	SR	SR	SR	SR	SR	SR
		nrocip	Numero de colegiados asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR
		nomcli	Nombre de cada asociado al CIP	I	SR	SR	SR	SR	SR	SR	SR
		Monto	Monto que pago el asociado de acuerdo a su cuota mensual	SR	SR	SR	D	SR	SR	SR	SR
		saldo	Saldo pendiente que debe el asociado	SR	SR	SR	SR	SR	I	SR	SR
		uabo	Fecha en la que cancelo el asociado	SR	SR	SR	SR	SR	I	SR	SR
		verif	verificacion de cumplimiento de pagos(Sin Registros)	SR	SR	SR	SR	SR	SR	SR	SR
		flag	Flagancia que tiene cada asociado de acuerdo a sus cuotas pagadas	SR	SR	SR	SR	SR	SR	SR	SR
		codcap	Codigo de capitulo al cual pertenece cada asociado	SR	SR	SR	SR	SR	SR	SR	SR
		estado	Estado que se encuentra cada asociado de acuerdo a sus pagos realizados	SR	SR	SR	I	SR	SR	SR	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN									
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría		
Dt101Con	8906	Codn	Codigo del Asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR	
		NomCon	Nombres y Apellidos de Asociado al CIP	I	SR	SR	SR	SR	SR	SR	SR	SR	SR
		CarCon	Cargo del Asociado al CIP (sin registro)	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
		TelCon	Telefono del Asociado al CIP (sin registro)	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
		CelCon	Celular del Asociado al CIP (sin registro)	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
		Email	Correo Electronico del Asociado al CIP (sin registro)	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
		idcon	Sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR

		FecNac	Fecha de Nacimiento del Asociado al CIP	SR	SR	D	SR	SR	SR	SR	SR
		NroDni	Numero del DNI del Asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR
		CodPar	Codigo del Parentesco de la Familia del Asociado al CIP	SR	SR	SR	SR	D	SR	SR	SR
		Obser	No contiene Datos - Sin revelancia	SR	SR	SR	SR	SR	SR	SR	SR
		RutFot	No contiene Datos - Sin revelancia	SR	SR	SR	SR	SR	SR	SR	SR
		rutfor	No contiene Datos - Sin revelancia	SR	SR	SR	SR	SR	SR	SR	SR
		item	No contiene Datos - Sin revelancia	SR	SR	SR	SR	SR	SR	SR	SR
		editpor	No contiene Datos - Sin revelancia	SR	SR	SR	SR	SR	SR	SR	SR
		dnicon	No contiene Datos - Sin revelancia	SR	SR	SR	SR	SR	SR	SR	SR
		notif	No contiene Datos - Sin revelancia	SR	SR	SR	SR	SR	SR	SR	SR
		Fla_lla	No contiene Datos - Sin revelancia	SR	SR	SR	SR	SR	SR	SR	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN								
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría	
mst01ccc	1763275	fecha	Fecha de Pago del Asociado al CIP	SR	SR	SR	D	SR	D	SR	SR	
		cdocu	Codigo de Documento	SR	SR	SR	SR	SR	SR	SR	SR	SR
		ndocu	Numero de Documento	SR	SR	SR	SR	SR	SR	SR	SR	SR
		crefe	Codigo de Referencia	SR	SR	SR	SR	SR	SR	SR	SR	SR
		nrefe	Numero de Referencia	SR	SR	SR	SR	SR	SR	SR	SR	SR
		codcli	Codigo de Asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
		nomcli	Nombres Completos de Asociados al CIP	I	SR	I	SR	SR	SR	SR	SR	SR
		ruccli	RUC de Proveedores al Cip	SR	SR	SR	SR	SR	SR	SR	SR	SR
		codcdv	Codigo de Tipo de Pago del Asociado	SR	SR	SR	SR	SR	SR	SR	SR	SR
		monto	Monto de Pago de Asociado al CIP	SR	SR	SR	D	SR	SR	SR	SR	SR
		saldo	Saldo pendiente por el Asociado al CIP	SR	SR	SR	SR	SR	I	SR	SR	SR

glosa	Descripcion de Pago echo por el Asociado	SR	SR	SR	D	SR	SR	SR	SR	SR
dias	En este campo se encuentran la cantidad de dias la cual algunos tienen dias diferentes (30,9,27, otros) sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
fven	En este campo el Asociado del CIP hizo el pago del ultimo dia de vencimiento	SR	SR	SR	SR	SR	SR	SR	SR	SR
uabo	Mes de pago que realizo el Asociado del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
tcam	Tipo de cambio en Dolares	SR	SR	SR	SR	SR	SR	SR	SR	SR
mone	tipo de moneda	SR	SR	SR	SR	SR	SR	SR	SR	SR
flag	Segun el campo de SALDO si este campo tiene datos el campo flag cogera el caracter 0 y si no tiene datos tomara el caracter 1	SR	SR	SR	SR	SR	I	SR	SR	SR
codest	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
codbco	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
nombco	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
unibco	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
agebco	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
fcam	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
flacje	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
fren	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
fpro	El unico dato extraido es el caracter 0 - Rin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
codven	Codigo de venta (solicitudes) de los asociados al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
cuenta	Numero de la cuenta para el deposito de las ventas (solicitudes)	SR	SR	SR	SR	SR	SR	SR	SR	SR
flagi	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
codcob	Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
selchk	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
marchk	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
flaenv	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
codenv	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
fecenv	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
flarec	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
codrec	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
fecrec	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR

		flasit	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		flachd	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		compro	Codigo de compra de los asociados al CIP	SR	SR	SR	SR	SR	SR	SR	SR
		obser	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		selec	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		flarol	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		codscc	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		FecAju	Fecha de Ajusta - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		TcaOld	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		codsub	Contiene campos vacios y llenmos desde 00 hasta 04 - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		subest	Contiene un solo dato (01) - sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		codpto	Contiene campos vacios y llenmos desde 01 hasta 05 - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		fsubest	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		fdieta	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		cuentaold	Numero de Cuenta de pago antigua	SR	SR	SR	SR	SR	SR	SR	SR
		nrocta	Numero de cuenta - sin datos	SR	SR	SR	SR	SR	SR	SR	SR
		idalias	No contiene datos - Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN									
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría		
mst01cli	19356	codcli	Hace referencia al código del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR	
		nomcli	Hace referencia al nombre completo del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR
		dircli	Hace referencia a la dirección del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR	SR

ruccli	Hace referencia al RUC de proveedores del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
nrodni	Hace referencia al número de documento de Identidad del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
flaper	Sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
dirent	Sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
telcli	Teléfono de colegiados asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
faxcli	Fax de colegiados asociado al CIP (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
codpos	Código postal del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
coddis	Código de distrito del que proviene el colegiado	SR	SR	SR	SR	SR	SR	SR	SR	SR
codpro	Código de provincia del que proviene el colegiado	SR	SR	SR	SR	SR	SR	SR	SR	SR
coddep	Código de departamento del que proviene el colegiado	SR	SR	SR	SR	SR	SR	SR	SR	SR
codpai	Código de país del que proviene el colegiado	SR	SR	SR	SR	SR	SR	SR	SR	SR
codzon	Código de zona del que proviene el colegiado	SR	SR	SR	SR	SR	SR	SR	SR	SR
codven	Código de vendedor del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
vensup	Vendedor suplente del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
codcob	Código de cobrador del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
codact	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
codcdv	Hace referencia al tipo de pago que puede efectuar el colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
fecing	Fecha de ingreso del asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
flalin	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
mcredi	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
fecapr	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
dsctxv	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
dsctcr	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
tipocl	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
diapag	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
hrspag	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR

ficha	Descripción en participación en elecciones de los colegiados asociados al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
nomcaj	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
telcaj	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
codgrp	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
email	Correo del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
fecreg	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
estado	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
nomava	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
dirava	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
rucava	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
telava	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
capdec	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
flaloc	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
nombco1	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
nrocta1	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
nomsec1	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
telsec1	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
nombco2	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
nrocta2	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
nomsec2	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
telsec2	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
nomref1	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
monref1	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
conref1	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
nomref2	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
monref2	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
conref2	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
nomref3	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
monref3	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
conref3	sin relevancia (sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
codcat	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
FecPag	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
HorPag	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR

TipFax	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
VisCli	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
ctlper	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
catsnt	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
perobl	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
NomEmp	Nombre de empresas que proveen y requieren servicios del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
DirEmp	Dirección de empresas que proveen y requieren servicios del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
Cargo	Cargo del representante de la empresa que provee o requiere de servicios del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
tipsex	Hace referencia al género de los colegiados (Masculino =1 , Femenino =2)	D	SR	SR	SR	SR	SR	SR	SR	SR
TelTra	Teléfono de trabajo del colegiado asociado al CIP (Sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
TelCel	Teléfono celular del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
CarExt	Carnet de extranjería del colegiado asociado al CIP (Sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
LibMil	Libreta militar del colegiado asociado al CIP (Sin registros)	SR	SR	SR	SR	SR	SR	SR	SR	SR
fecnac	Fecha de nacimiento del colegiado asociado al CIP	SR	SR	I	SR	SR	SR	SR	SR	SR
codcar	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
DirCob	Lugar de nacimiento del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
CodOlb	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
capitulo	Hace referencia al capítulo al cual pertenece el colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
especialidad	Hace referencia a la especialidad del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	D	SR
tiposoc	Hace referencia al tipo de colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	D

codcap	Hace referencia al código de los capítulos que conforman el CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
codesp	Hace referencia al código de las especialidades que conforman a los capítulos del CIP	SR	SR	SR	SR	SR	SR	SR	I	SR
coduni	Hace referencia al código de la universidad del que egresó el colegiado	SR	SR	SR	SR	SR	SR	SR	SR	SR
feccol	Hace referencia a la fecha de colegiatura del colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
estciv	Hace referencia al estado civil de los colegiados asociados del CIP	SR	D	SR	SR	SR	SR	SR	SR	SR
consoc	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
nroCIP	Hace referencia al número de CIP del colegiado	SR	SR	SR	SR	SR	SR	SR	SR	SR
CodGru_empre	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
Cuota_venta	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
codtipo	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
codmonlinea	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
mampcre	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
FECTRA	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
FECDES_TRA	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
FECHAS_TRA	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
MOTIVO_TRA	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
fecupdate	Fecha de actualización de los registros de fecha de colegiatura de los asociados al CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
nomcli_update	Hace referencia a los nombres de los asociados actualizados en el nuevo sistema del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
email_update	Hace referencia a los correos de los asociados actualizados en el nuevo sistema del CIP	SR	SR	SR	SR	SR	SR	SR	SR	SR
is_registro_sticker	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
is_registro_web	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
fecing_dos	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR
feccol_dos	sin relevancia	SR	SR	SR	SR	SR	SR	SR	SR	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN							
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría
roly	244666	fecha	Fecha de apertura para que se empiece a realizar el pago de la cuota mensual	SR	SR	SR	SR	SR	SR	SR	SR
		nrocip	Número de colegiado asociado al CIP	SR	SR	SR	SR	SR	SR	SR	SR
		nomcli	Nombre completo del colegiado	I	SR	SR	SR	SR	SR	SR	SR
		Monto	Monto total del pago realizado de la cuota mensual	SR	SR	SR	D	SR	SR	SR	SR
		fcan	Fecha de cancelación de la cuota mensual	SR	SR	SR	SR	SR	I	SR	SR
		Saldo	Saldo pendiente de la cuota mensual	SR	SR	SR	SR	SR	SR	SR	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN							
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría
tbl_capitulo	244666	CodCap	Campo Identificador de la Tabla	SR	SR	SR	SR	SR	SR	I	SR
		NomCap	Nombre del Capitulo de Cada Especialidad del Colegiado	SR	SR	SR	SR	SR	SR	I	SR
		CodScc	Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		ctavtaloc	Tipo de Cuentas	SR	SR	SR	SR	SR	SR	SR	SR
		ctavtaext	Tipo de Cuentas	SR	SR	SR	SR	SR	SR	SR	SR
		comparador	Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		orden_cap	Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR

				VARIABLES DE PREDICCIÓN							
Tabla	Nro Registros	Columnas	Descripcion	Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría
tbl_especialidad	98	CodEsp	Campo que contiene el Identificador de la Tabla	SR	SR	SR	SR	SR	SR	SR	SR
		NomEsp	Nombre de la Especialidad a la cual pertenece el Colegiado	SR	SR	SR	SR	SR	SR	D	SR
		CodCap	Campo que Contiene la Clave Foranea de la Tabla tbl_capitulo	SR	SR	SR	SR	SR	SR	SR	SR

				VARIABLES DE PREDICCIÓN							
Tabla	Nro Registros	Columnas	Descripcion	Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría
Tbl_Parentesco	8	CodPar	Campo Identificador de la Tabla	SR	SR	SR	SR	SR	SR	SR	SR
		NomPar	Campo que contiene el Parentesco de los Familiares de los Colegiados	SR	SR	SR	SR	I	SR	SR	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN							
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría
tbl01cdv	3	codcdv	Campo Identificador de la Tabla	SR	SR	SR	I	SR	SR	SR	SR
		nomcdv	Campo que contiene la Forma de Pago del Colegiado	SR	SR	SR	I	SR	SR	SR	SR
		nrolet	Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		diaven	Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		cuenta	Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR
		CtaCoL	Tipo de Cuentas	SR	SR	SR	SR	SR	SR	SR	SR
		CtaCoE	Tipo de Cuentas	SR	SR	SR	SR	SR	SR	SR	SR
		CtaCrL	Tipo de Cuentas	SR	SR	SR	SR	SR	SR	SR	SR
		CtaCrE	Tipo de Cuentas	SR	SR	SR	SR	SR	SR	SR	SR
		Codcon	Sin Relevancia	SR	SR	SR	SR	SR	SR	SR	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN								
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría	
tbl01cac	8	codcat	Codigo que pertenece a la categoria de cada asociado al CIP	SR	SR	SR	SR	SR	SR	SR	D	
		nomcat	Nombre de la categoria que pertenece a cada colegiado	SR	SR	SR	SR	SR	SR	SR	SR	D
		ctafacloc	Sin Relevancia - Datos inconsistentes	SR	SR	SR	SR	SR	SR	SR	SR	SR
		ctafacext	Sin Relevancia - Datos inconsistentes	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC01	Importe de la cuota mensual del mes Enero	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC02	Importe de la cuota mensual del mes Febrero	SR	SR	SR	SR	SR	SR	SR	SR	SR

		IMPC03	Importe de la cuota mensual del mes Marzo	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC04	Importe de la cuota mensual del mes Abril	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC05	Importe de la cuota mensual del mes Mayo	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC06	Importe de la cuota mensual del mes Junio	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC07	Importe de la cuota mensual del mes Julio	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC08	Importe de la cuota mensual del mes Agosto	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC09	Importe de la cuota mensual del mes Septiembre	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC10	Importe de la cuota mensual del mes Octubre	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC11	Importe de la cuota mensual del mes Noviembre	SR	SR	SR	SR	SR	SR	SR	SR	SR
		IMPC12	Importe de la cuota mensual del mes Diciembre	SR	SR	SR	SR	SR	SR	SR	SR	SR

Tabla	Nro Registros	Columnas	Descripcion	VARIABLES DE PREDICCIÓN								
				Género	Estado Civil	Edad	Monto Pagado	Numero de Hijos	Dias de Incumplimiento	Especialidad	Categoría	
tbl01ven	44	codven	Codigo de la ventanilla donde se realiza la venta	SR	SR	SR	I	SR	SR	SR	SR	SR
		nomven	Nombre de la persona quien realiza la venta	SR	SR	SR	I	SR	SR	SR	SR	SR
		estado	Estado activo y inactivo	SR	SR	SR	SR	SR	SR	SR	SR	SR
		comven		SR	SR	SR	SR	SR	SR	SR	SR	SR
		comcob		SR	SR	SR	SR	SR	SR	SR	SR	SR
		codadm		SR	SR	SR	SR	SR	SR	SR	SR	SR
		ubica	Ubicación donde se realiza la venta	SR	SR	SR	SR	SR	SR	SR	SR	SR

ANEXO 10: Diccionario de Datos del Modelo Entidad – Relacional.

Nombre Columna	Tipo Dato	Null Option	Descripción	PK	FK
CodEsp	Char(6)	Not null	Código Especialidad	PK	-
NomEsp	Varchar(100)	Not null	Nombre especialidad	-	-
CodCap	Char(3)	Null	Código del capitulo	-	FK

Nombre Columna	Tipo Dato	Null Option	Descripción	PK	FK
CodCap	Char(3)	Not null	Código Capitulo	PK	-
NomCap	Varchar(100)	Not null	Nombre Capitulo	-	-
Orden_Cap	Int	Null	Orden de los Capitulo	-	FK

Nombre Columna	Tipo Dato	Null Option	Descripción	PK	FK
IdFamiliar	Int	Not null	Código del Familiar del Asociado al CIP	PK	-
Familiar	Varchar(255)	Not null	Nombres Completos del Familiar del Asociado al CIP	-	-
FechaNac	Datetime	Null	Fecha de Nacimiento del Familiar del Asociado al CIP	-	-
Edad	Int	Null	Edad del Familiar Asociado al	-	-

NumDni	Char(10)	Null	CIP Numero de DNI del familiar Asociado al CIP	-	-
NroCIP	Char(7)	Not null	Numero del CIP del Asociado	-	FK
NomPar	Varchar(20)	Not null	Nombre del Parentesco del Asociado al CIP	-	-

Nombre Columna	Tipo Dato	Null Option	Descripción	PK	FK
CodCdv	Char(2)	Not null	Codigo del Cobro del Asociado al CIP	PF	-
NomCdv	Char (20)	Not null	Nombre del Cobro del Asociado al CIP	-	-
Diaven	Int	Null	Número de días a pagar	-	-
CtaCol	Char(12)	Null	Numero de Cuota a pagar	-	-
CtaCoe	Char(12)	Null	Numero de Cuota a pagar	-	-
CtaCrL	Char(12)	Null	Numero de Cuota a pagar	-	-
CtaCrE	Char(12)	Null	Numero de Cuota a pagar	-	-

Nombre Columna	Tipo Dato	Option Null	Descripción	PK	FK
Cod Pago	Int	Not null	Código del Pago	PK	-
NroCIP	Char(7)	Not null	Numero CIP de los Asociados	-	-
Mes_Pago	Varchar(10)	Not null	Monto totañ del mes	-	-
FechaApertura	Date	Not null	Fecha de inicio de la apertura	-	-
FechaVencimiento	Date	Not null	Fecha de vencimiento	-	-
FechaPago	Date	Null	Fecha del Pago asociados al CIP	-	-
Monto_Programacion	Float	Not null	Monto programado en la U	-	-
Monto_Pagado	Float	Not null	Monto total para pagar de los asociados al CIP.	-	-
Saldo	Float	Floatnot null	Saldo del Asociado al CIP	-	-
Días	Int	Null		-	-
CodCdv	Char(2)	Null	Codigo de Cobro	-	FK
CodVen	Char(2)	Null	Codigo de Venta	-	FK

Nombre Columna	Tipo Dato	Null Option	Descripción	PK	FK
CodVen	Char(5)	Not null	Codigo del Cajero	PK	-
NomVen	Char(30)	Not null	Nombre del Cajero	-	-

ANEXO 11: Algoritmo Completo de Adaboost.M1 utilizado.

#-----UTILIZANDO EL Percentage Split (Porcentaje Dividido)

Cargar todos los paquetes necesarios

library(ggplot2)

library(lattice)

library(caret)

library(nnet)

library(mlbench)

library(rpart)

library(adabag)#contiene al algoritmo adaboost

Carga de datos inicial

colegiados <- read.csv("Enero_Algoritmo.csv",dec = ".")

View(colegiados)

Selección de una submuestra (el 70% de los datos)

set.seed(101) #semilla aleatoria

tamano.total <- nrow(colegiados)

**tamano.entreno <- round(tamano.total*0.7)#total para entrenamiento: el
70%(0.7)**

colegiados.indices <- sample(1:tamano.total , size=tamano.entreno)

#Dividiendo el conjunto de datos

**colegiados.entreno <- colegiados[colegiados.indices,]#entrenamiento (el
70%)**

**colegiados.test <- colegiados[-colegiados.indices,] #test(el restante, osea el
30%)**

**#Asegurarnos la proporcion de Colegiados que pagaron y no pagaron es
aproximadamente**

```

#la misma en ambos grupos de datos ejecutamos lo siguiente:
t1<-table(colegiados.entreno$PAGO)
as.numeric(t1[1]/t1[2])
t2<-table(colegiados.test$PAGO)
as.numeric(t2[1]/t2[2])
#Ambos Tienen una proporcion parecida por lo que la proporcion se mantiene

```

```

# Ejecución de la técnica AdaBoost.M1(demora en crear el modelo)
t <- proc.time() # Inicia el cronómetro
modelito <- boosting(PAGO ~ ., data = colegiados.entreno, coeflearn =
  "Breiman", boos = FALSE, mfinal = 100)
proc.time()-t # Detiene el cronómetro

```

```

summary(modelito)

```

```

#Importancia de cada variable
modelito$importance #importancia de cada variable
modelito$trees #arboles creados
modelito$weights #pesos para todos los arboles
errorevol(modelito,colegiados) #Evolucion del error
modelito$formula#variables utilizadas. Separando variable dependiente de
independiente.

```

```

# test (Calculando Tiempo en segundos)
t <- proc.time() # Inicia el cronómetro
predicciones <- predict(object = modelito, newdata=colegiados.test, type =
  "class")
proc.time()-t # Detiene el cronómetro

```

```

predicciones$class
# Matriz de confusión
confusionMatrix(predicciones$class)

```



```

# Correctamente clasificados y predecidos
100 * (sum(diag(predicciones$confusion)) / sum(predicciones$confusion))

#Finalmente, podemos llamar al paquete Tree y trazar árboles generados por
el
#paquete Adabag, utilizando el siguiente código:
t1<-modelito$trees[[2]] #arbol 2
library(tree)
plot(t1)
text(t1,pretty=0)

#Graficar Importancia de variables
barplot(sort(modelito$importance,dec=T),col="blue",horiz=T,las=1)

#-----CURVA ROC
library(caTools)
colAUC(as.numeric(as.factor(predicciones$class)),as.factor(colegiados.test$
PAGO),plotROC = TRUE ,alg = c("Wilcoxon", "ROC"))
abline(a=0, b=1)

#-----PREDICCION

# Ingresamos colegiados sin indicarle si pagó o no, para que el algoritmo
prediga
nuevoT <- read.csv("Enero_Prededir.csv",dec = ".")
View(nuevoT)

#PRediccion de pago de colegiados
b <-predict(modelito,nuevoT, type = "prob")
#c <- predict(modelito,nuevoT[,-42]) #hace lo mismo
head(b) #visualizacion de predicciones

```

```
#Adicionar columna de Prediccion a la tabla nuevo  
nuevoT$Prediccion_PAGO <- b[["class"]]  
#Guarda la tabla "nuevo" en un csv con la prediccion  
write.csv(nuevoT,"C:/Users/MAX/Desktop/ALGORITMO/Enero_Predecir3.csv  
    ")
```

ANEXO 12: Algoritmo Completo de C5.0 utilizado.

#-----UTILIZANDO EL Percentage Split (Porcentaje Dividido)

Carga el paquete específico del Árbol de clasificación C5.0

**#Si no se tiene se puede instalar: install.packages("C50", dependencies =
TRUE)**

library(C50)

Carga de datos inicial

datos <- read.csv("Diciembre_Algoritmo.csv",dec = ".")

View(datos)

Selección de una submuestra del 70% de los datos

set.seed(101)#Semilla aleatoria

tamano.total <- nrow(datos)

tamano.entreno <- round(tamano.total*0.7)

datos.indices <- sample(1:tamano.total , size=tamano.entreno)

#Dividiendo el conjunto de datos

datos.entreno <- datos[datos.indices,] #entrenamiento (el 70%)

datos.test <- datos[-datos.indices,] #test(el restante, osea el 30%)

#Asegurarnos la proporcion de Colegiados que pagaron y no pagaron

**#es aproximadamente la misma en ambos grupos de datos ejecutamos lo
siguiente:**

t1<-table(datos.entreno\$PAGO)

as.numeric(t1[1]/t1[2])

t2<-table(datos.test\$PAGO)

as.numeric(t2[1]/t2[2])

**#Tienen una proporcion de 0.65 aprox. ambos por lo que la proporcion se
mantiene**

```

# Ejecución de la técnica de Arbol de Decision C5.0
t <- proc.time() # Inicia el cronómetro
modelo1 <- C5.0(PAGO ~ ., data = datos.entreno,
               trial = 100,
               control = C5.0Control(noGlobalPruning = TRUE,CF=0.20))
proc.time()-t # Detiene el cronómetro

summary(modelo1) # Información sobre el modelo

#Graficar Arbol
plot(modelo1) # Gráfico

# Para detallar un nodo en particular se usaria la siguiente función
plot(modelo1, subtree=5) #Muestra un nodo en particular

# test (Calculando Tiempo en segundos)
t <- proc.time() # Inicia el cronómetro
precision1 <- predict(modelo1,newdata=datos.test)
proc.time()-t # Detiene el cronómetro

precision1
# Matriz de confusión
library(caret)
tabla1 <- table(precision1, datos.test$PAGO)
confusionMatrix(tabla1)

# % correctamente clasificados
100 * sum(diag(tabla1)) / sum(tabla1)

#-----CURVA ROC
library(caTools)

```

```

colAUC(as.numeric(as.factor(precision1)),as.factor(datos.test$PAGO),plotR
      OC = TRUE)
abline(a=0, b=1)

#-----PREDICCION
# Cargamos colegiados sin indicarle si pagó o no para que el algoritmo
  prediga
nuevo <- read.csv("Enero_Prededir.csv",dec = ".")#nuevo sed de datos
View(nuevo)

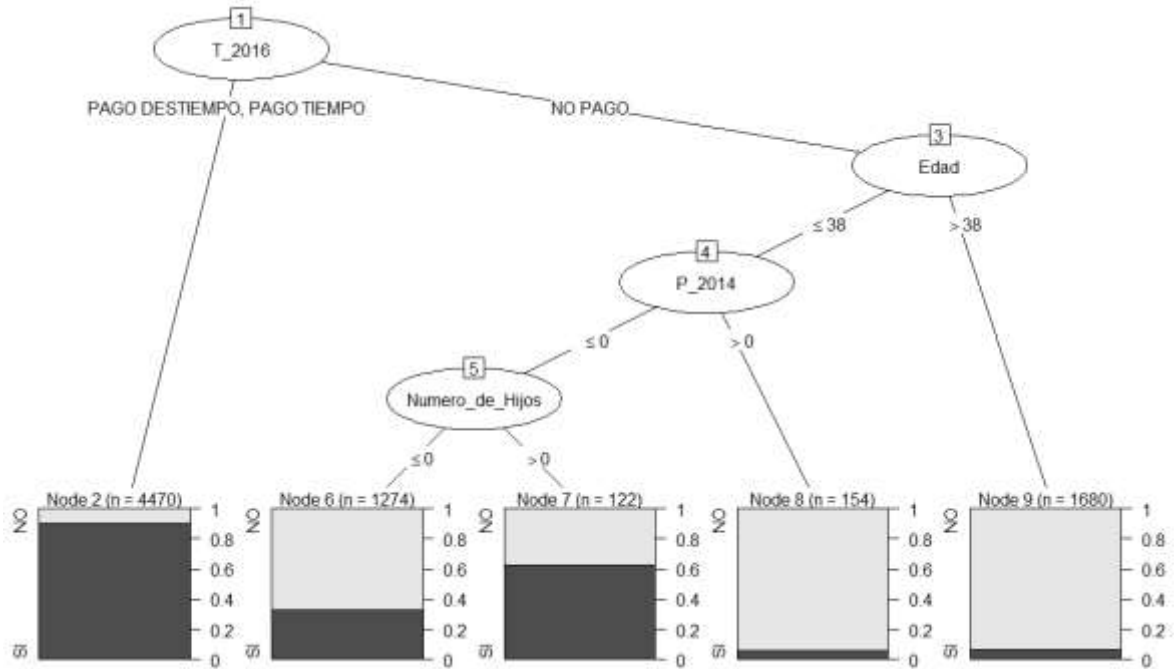
#PRedicción con algunos colegiados
b <- predict(modelo1,nuevo[,-42])#se ignora la columna a predecir
#a <- predict(modelo1,nuevo, type = "prob")
head(b)#se pueden visualizar algunas las predicciones

#Adicionar columna de Predicción a la tabla nuevo
nuevo$Predicción_PAGO <- b
#Guarda la tabla "nuevo" en un csv con la predicción
write.csv(nuevo,"C:/Users/MAX/Desktop/ALGORITMO/Enero_Prededir2.csv")

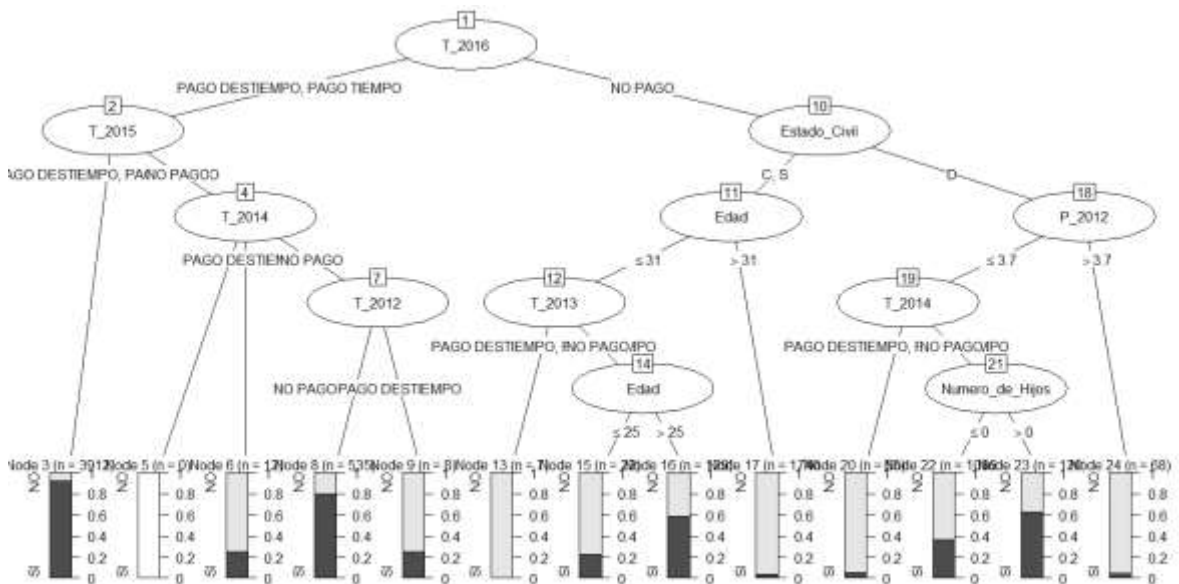
```

ANEXO 13: Árboles generados por Decision Tree C5.0.

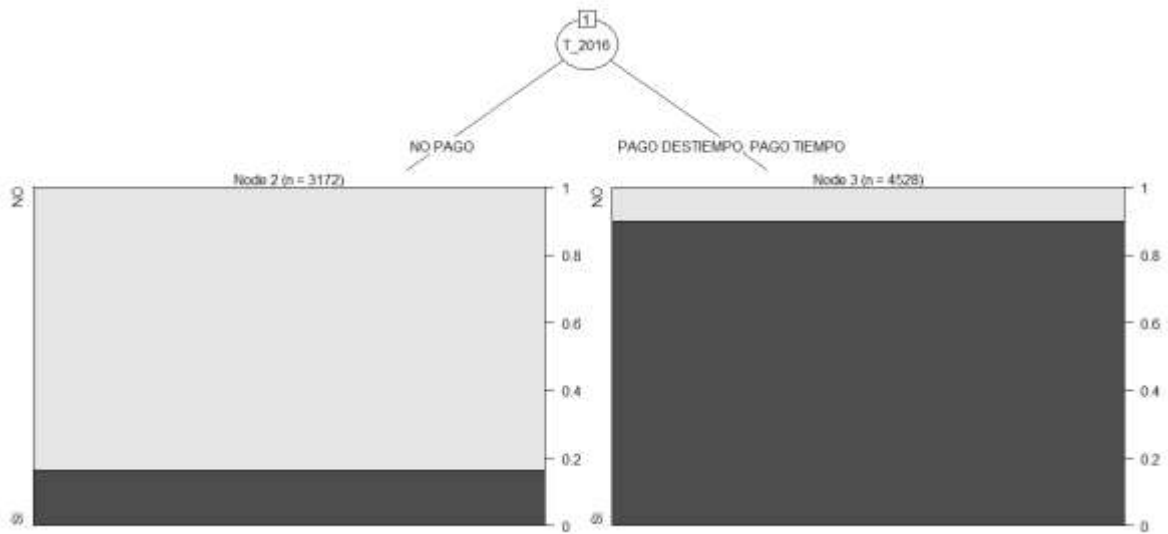
Enero



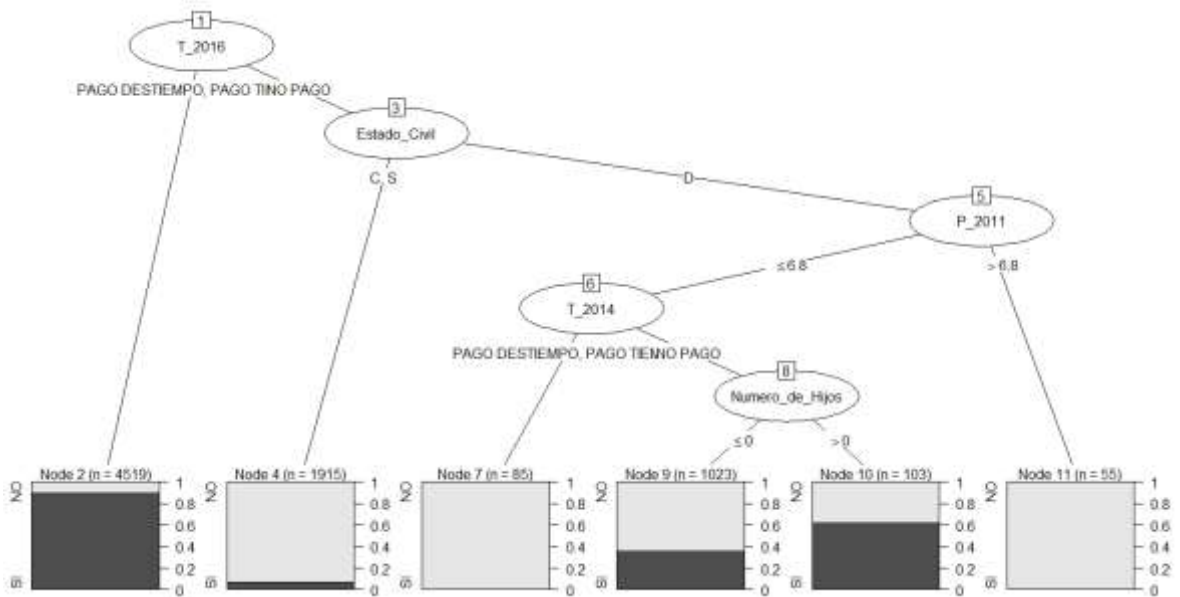
Febrero



Marzo



Abril



Mayo

```
Numero_de_Hijos > 2: SI (207.2/76.2)
Numero_de_Hijos <= 2:
:...T_2016 in {PAGO DESTIEMPO,PAGO TIEMPO}: SI (4252.8/1992.6)
  T_2016 = NO PAGO:
    :...T_2002 in {PAGO DESTIEMPO,PAGO TIEMPO}: NO (22.8)
      T_2002 = NO PAGO:
        :...T_2007 in {PAGO DESTIEMPO,PAGO TIEMPO}: NO (13.5)
          T_2007 = NO PAGO:
            :...T_2014 = PAGO TIEMPO: NO (17)
              T_2014 in {NO PAGO,PAGO DESTIEMPO}:
                :...T_2013 = PAGO TIEMPO: NO (12.7)
                  T_2013 in {NO PAGO,PAGO DESTIEMPO}:
                    :...T_2012 = PAGO TIEMPO: NO (9.1)
                      T_2012 in {NO PAGO,PAGO DESTIEMPO}:
                        :...T_2015 = PAGO TIEMPO: NO (32.5/5.4)
                          T_2015 in {NO PAGO,PAGO DESTIEMPO}:
                            :...Edad <= 36: [SI]
                              Edad > 36:
                                :...T_2003 in {NO PAGO,
                                  : PAGO TIEMPO}: NO (1077.8/430.3)
                                T_2003 = PAGO DESTIEMPO: SI (8.4/1.6)
```

Junio

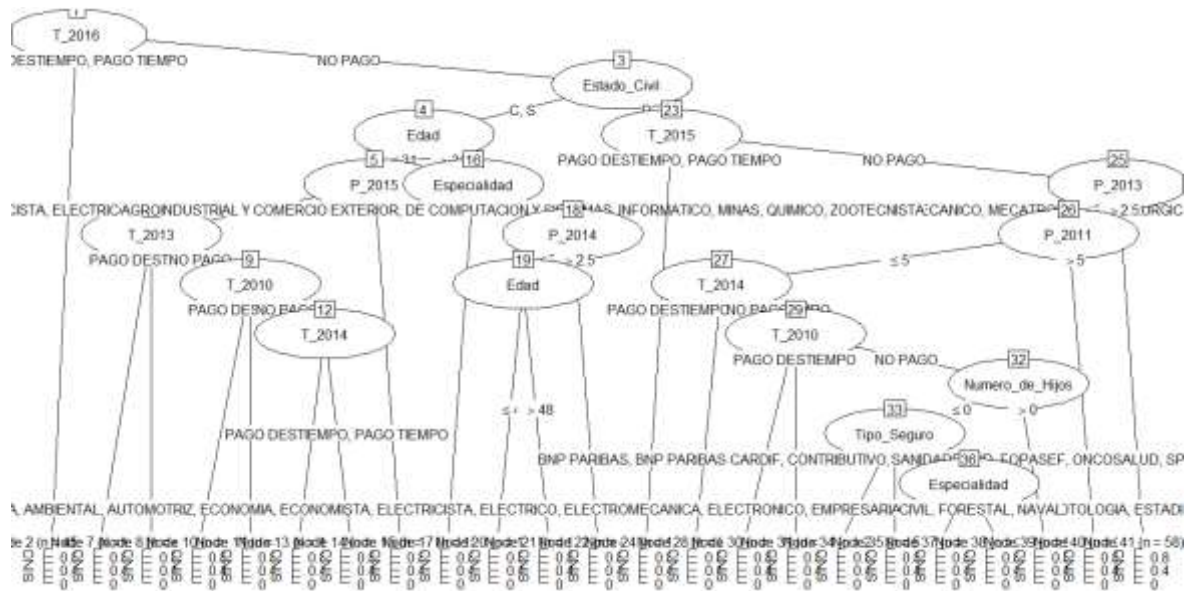
```
Numero_de_Hijos > 3: SI (57.6/18.1)
Numero_de_Hijos <= 3:
:...T_2007 = PAGO TIEMPO: SI (61/25.8)
  T_2007 in {NO PAGO,PAGO DESTIEMPO}:
    :...T_2015 = PAGO TIEMPO:
      :...P_2016 <= 0: NO (53.3/13.9)
        : P_2016 > 0: SI (1313.7/583.6)
      T_2015 in {NO PAGO,PAGO DESTIEMPO}:
        :...T_2002 in {PAGO DESTIEMPO,PAGO TIEMPO}: NO (209.4/99.5)
          T_2002 = NO PAGO:
            :...T_2003 = NO PAGO: NO (5980.6/2910.6)
              T_2003 in {PAGO DESTIEMPO,PAGO TIEMPO}: SI (24.2/5.5)
```


Julio

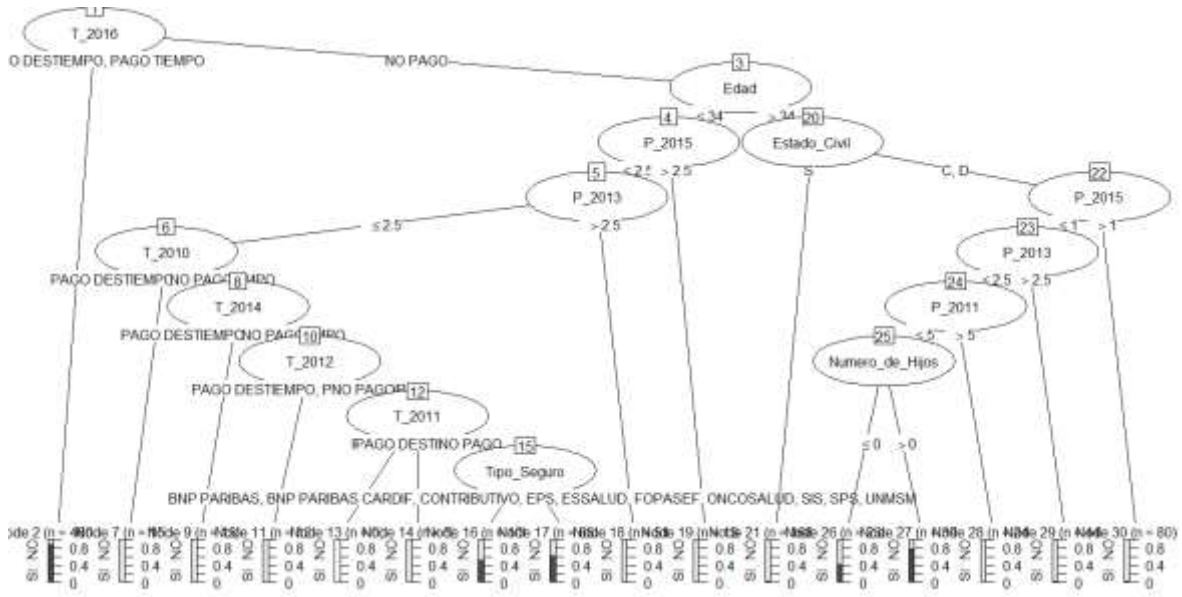
```

Numero_de_Hijos > 3: SI (43.5/10.1)
Numero_de_Hijos <= 3:
...T_2016 in {NO PAGO,PAGO DESTIEMPO}: NO (5914.7/2744.8)
  T_2016 = PAGO TIEMPO:
  ...P_2005 > 2.5: SI (82.2/21.7)
    P_2005 <= 2.5:
    ...T_2004 = PAGO DESTIEMPO: NO (11.2/2.2)
      T_2004 = PAGO TIEMPO: SI (1)
      T_2004 = NO PAGO:
      ...T_2006 = PAGO DESTIEMPO: SI (5.7)
        T_2006 = PAGO TIEMPO: NO (5.2/0.6)
        T_2006 = NO PAGO:
        ...T_2009 in {NO PAGO,PAGO TIEMPO}: SI (1599.1/736.5)
          T_2009 = PAGO DESTIEMPO: NO (37.4/15.1)
  
```

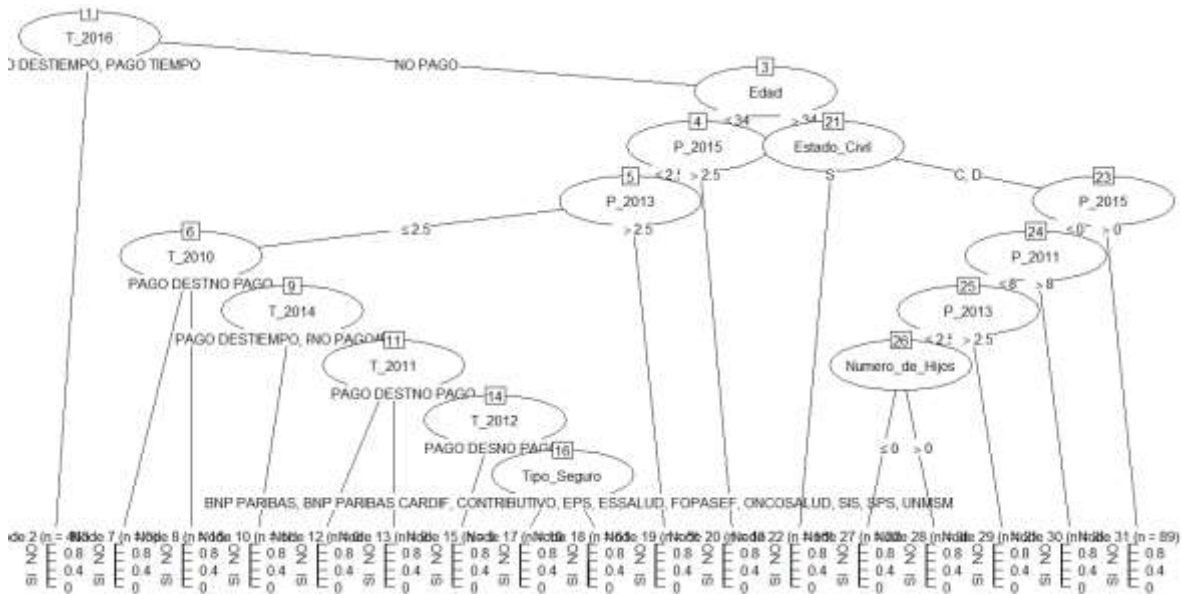
Agosto



Septiembre



Octubre



Noviembre

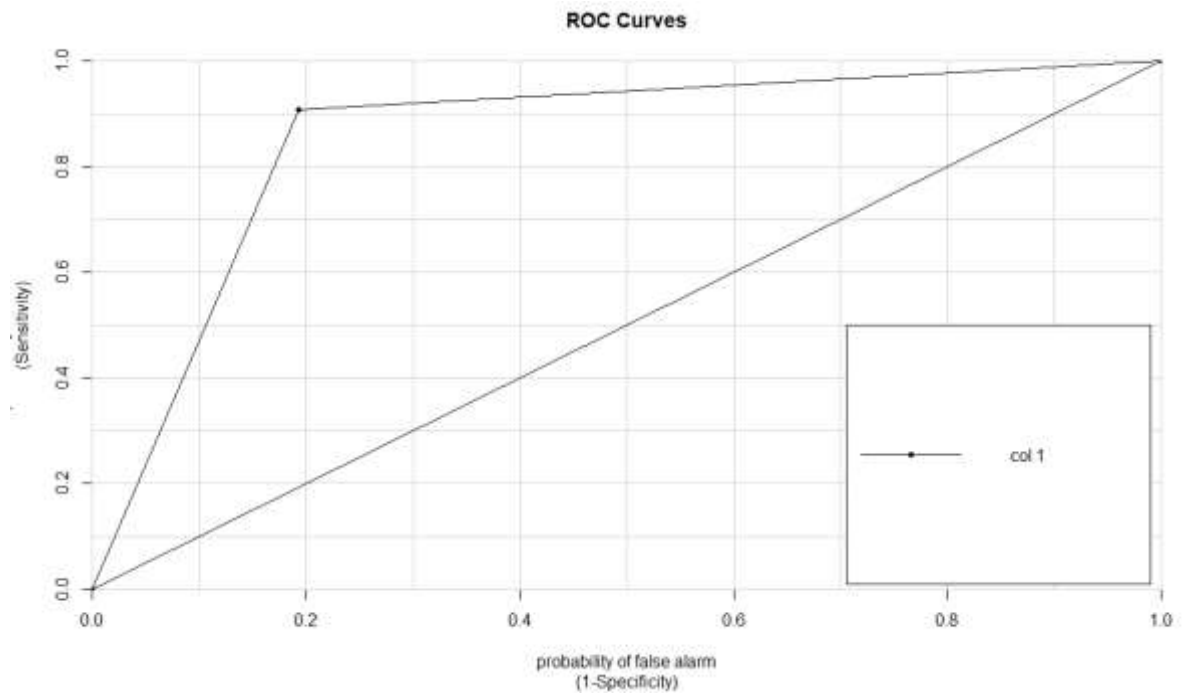
```
T_2014 = PAGO TIEMPO:
:...T_2016 = NO PAGO: NO (57.5/18.7)
:   T_2016 in {PAGO DESTIEMPO,PAGO TIEMPO}: SI (1170.3/480.9)
T_2014 in {NO PAGO,PAGO DESTIEMPO}:
:...T_2016 in {PAGO DESTIEMPO,PAGO TIEMPO}:
:   ...T_2015 = PAGO TIEMPO: SI (1532.7/751.6)
:   :   T_2015 = PAGO DESTIEMPO:
:   :   :   ...Edad <= 52: NO (1141.2/525)
:   :   :   :   Edad > 52: SI (359.6/140)
:   :   T_2015 = NO PAGO:
:   :   :   ...T_2013 in {PAGO DESTIEMPO,PAGO TIEMPO}: NO (22.3/3.2)
:   :   :   T_2013 = NO PAGO: [S1]
T_2016 = NO PAGO:
:...T_2010 in {PAGO DESTIEMPO,PAGO TIEMPO}: NO (420.9/187.3)
:   T_2010 = NO PAGO:
:   :   ...T_2015 = PAGO TIEMPO: NO (34.5)
:   :   :   T_2015 in {NO PAGO,PAGO DESTIEMPO}:
:   :   :   :   ...T_2011 in {PAGO DESTIEMPO,PAGO TIEMPO}: NO (21.2)
:   :   :   :   T_2011 = NO PAGO:
:   :   :   :   :   ...Numero_de_Hijos <= 0: NO (1929.1/860.2)
:   :   :   :   :   :   Numero_de_Hijos > 0: SI (116.1/47.7)
```

Diciembre

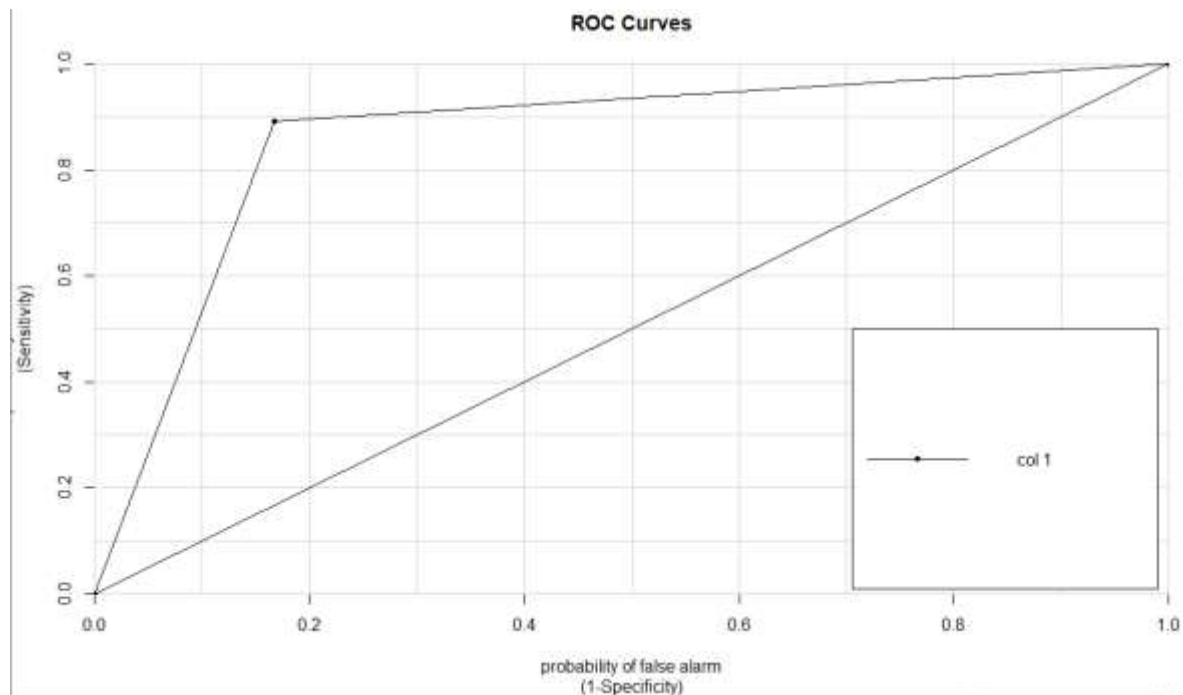
```
T_2016 in {PAGO DESTIEMPO,PAGO TIEMPO}: SI (5449.5/2675.4)
T_2016 = NO PAGO:
:...P_2005 > 5: NO (29.7/3.2)
:   P_2005 <= 5:
:   :   ...T_2012 = PAGO TIEMPO: NO (31.6/5.4)
:   :   :   T_2012 in {NO PAGO,PAGO DESTIEMPO}:
:   :   :   :   ...T_2015 = PAGO TIEMPO: NO (60.2/15.7)
:   :   :   :   :   T_2015 in {NO PAGO,PAGO DESTIEMPO}:
:   :   :   :   :   :   ...Estado_Civil = C: NO (43.7/20.1)
:   :   :   :   :   :   :   Estado_Civil = D:
:   :   :   :   :   :   :   :   ...T_2013 in {PAGO DESTIEMPO,PAGO TIEMPO}: NO (22.7)
:   :   :   :   :   :   :   :   :   T_2013 = NO PAGO:
:   :   :   :   :   :   :   :   :   :   :   ...T_2011 in {PAGO DESTIEMPO,
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   PAGO TIEMPO}: NO (7.5)
:   :   :   :   :   :   :   :   :   :   :   :   T_2011 = NO PAGO:
:   :   :   :   :   :   :   :   :   :   :   :   :   :   ...T_2014 in {NO PAGO,PAGO TIEMPO}: SI (1256.5/600.6)
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   T_2014 = PAGO DESTIEMPO: NO (4.9)
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   Estado_Civil = S:
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   ...Edad <= 31: SI (87.5/31.9)
:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   Edad > 31: [S1]
```

Anexo 14: CURVA ROC DE ADABOOST Y DECISIÓN TREE (ARBOL DE DECISION)

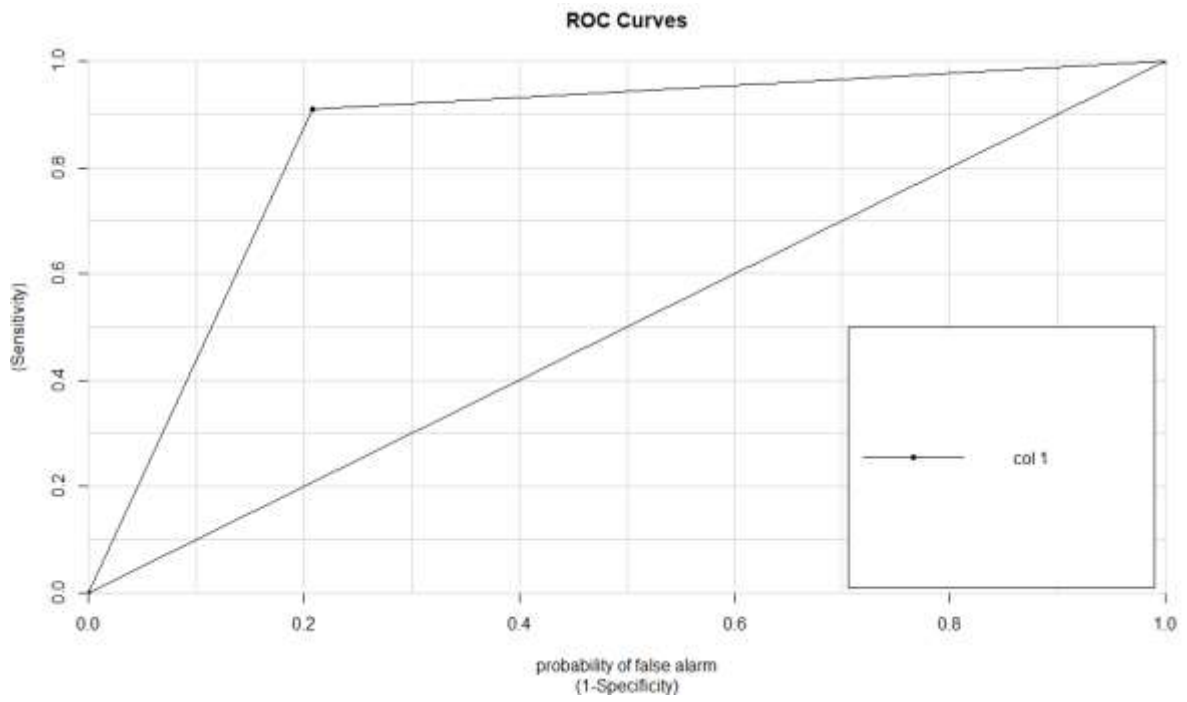
Enero Adaboost



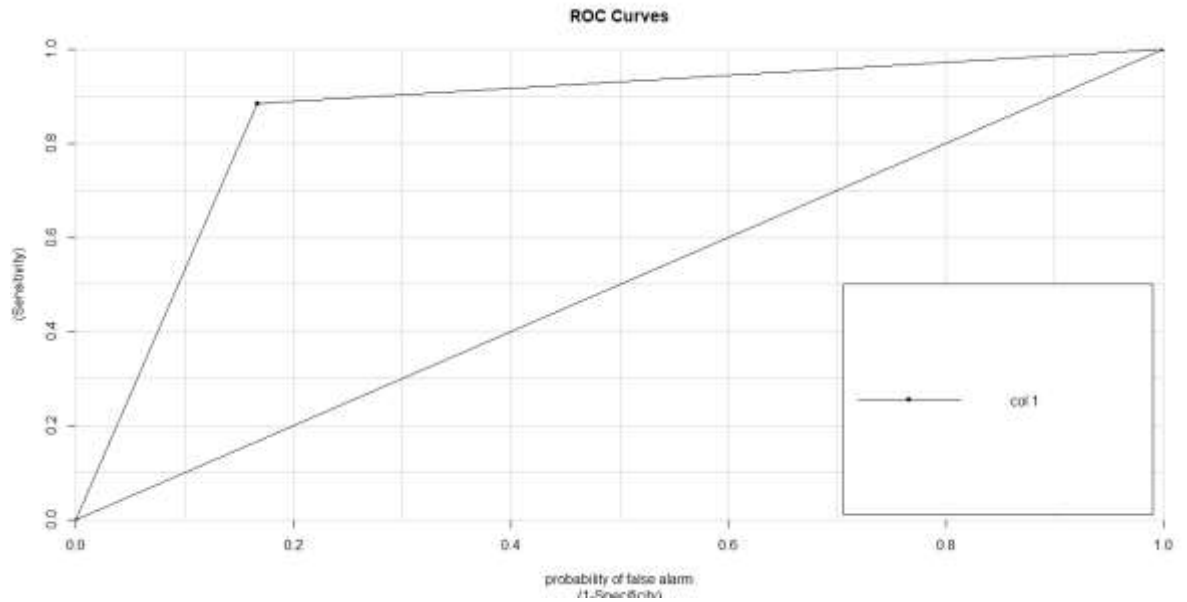
Enero Decision Tree



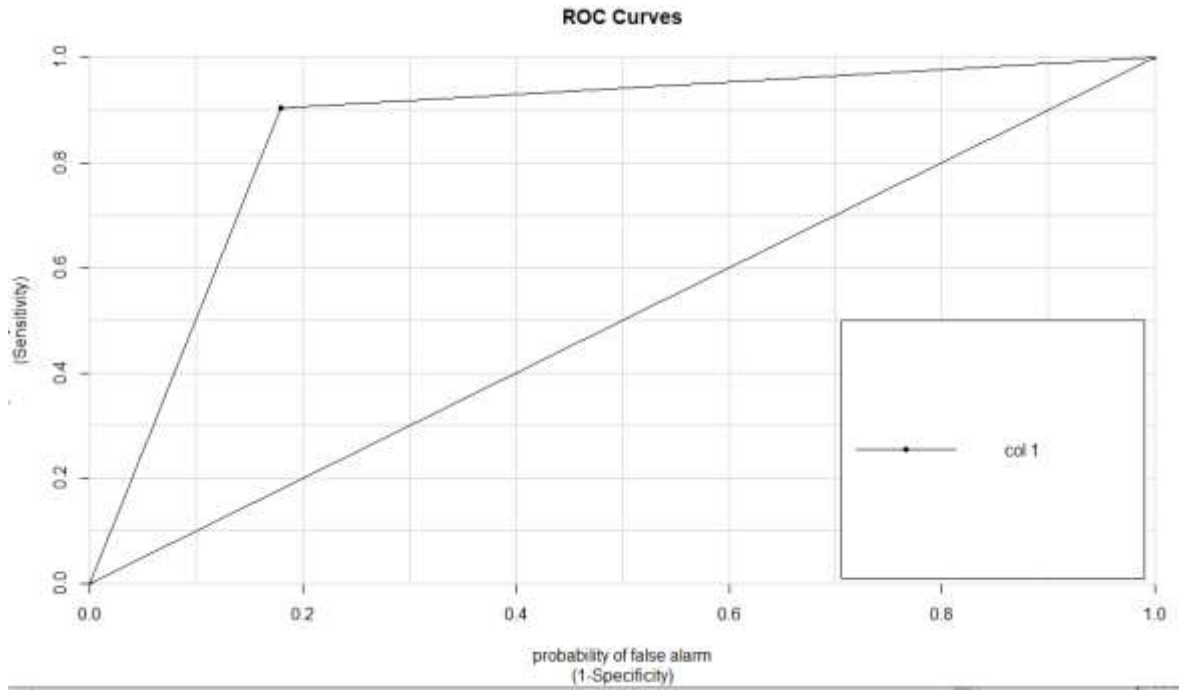
Febrero Adaboost



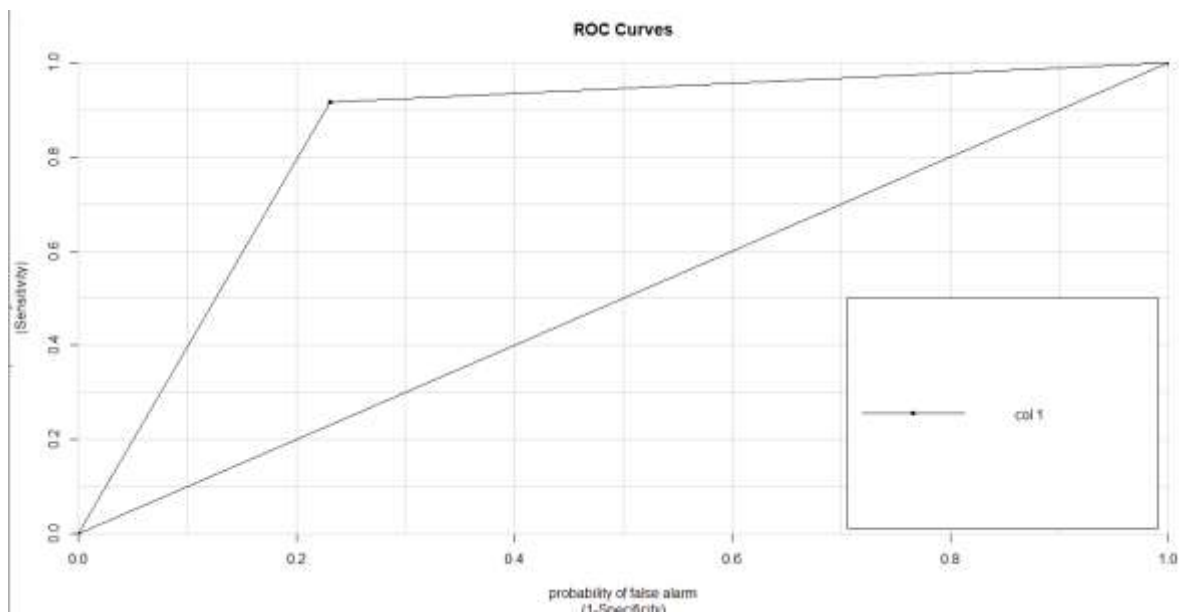
Febrero Decision Tree



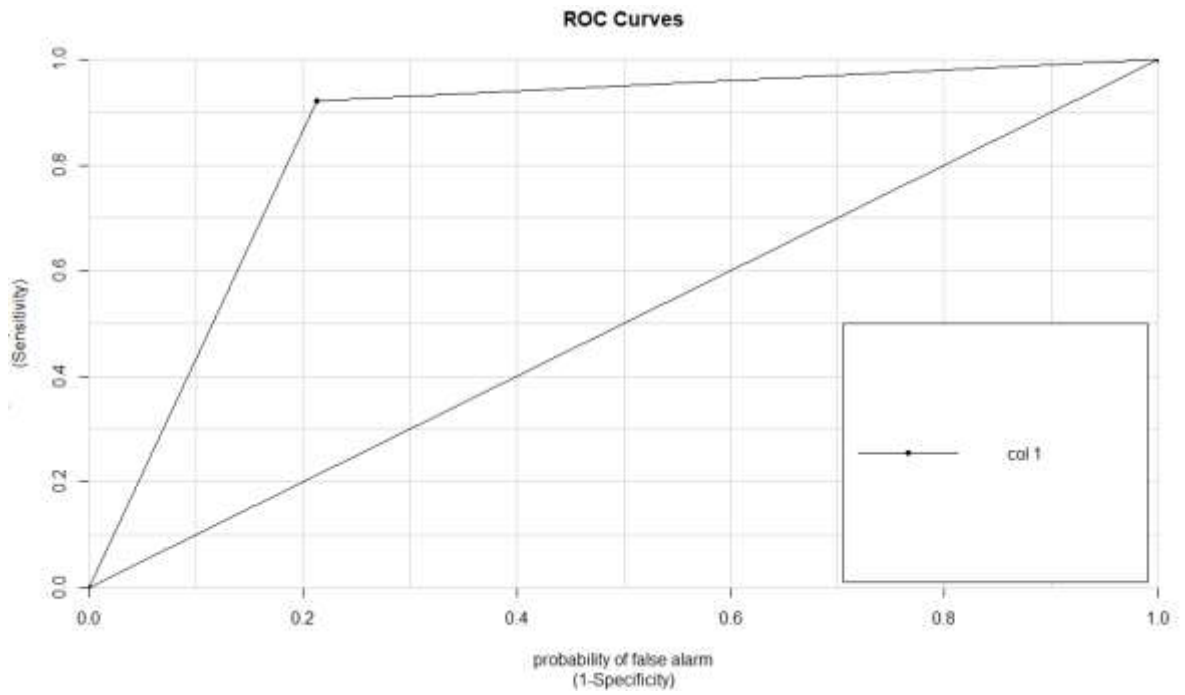
Marzo Adaboost



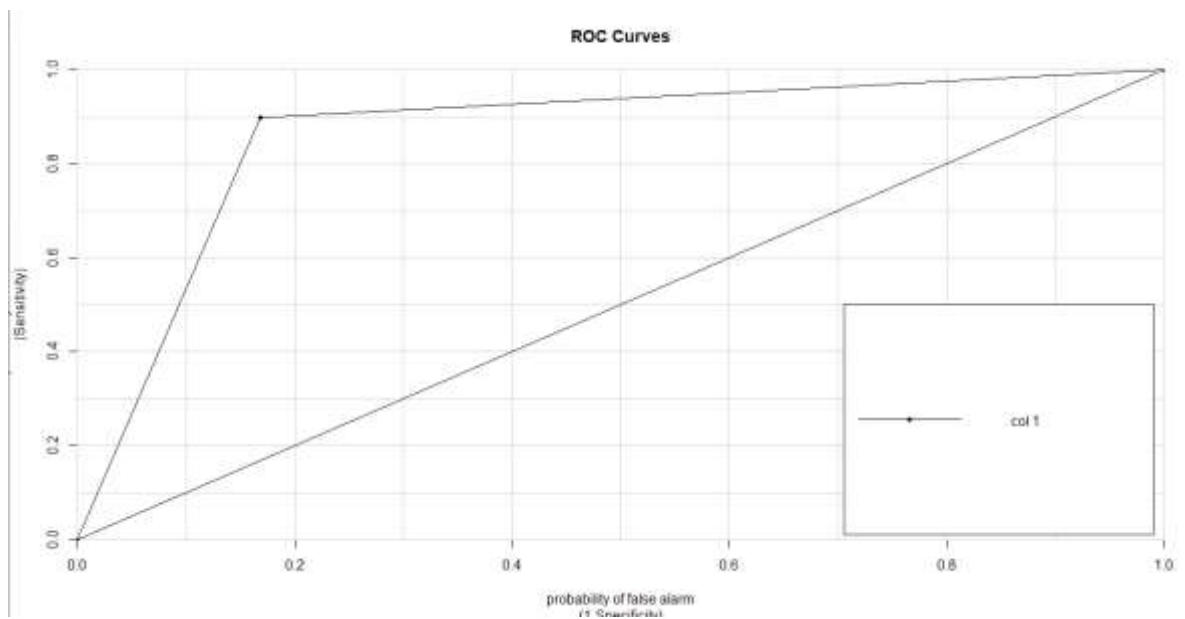
Marzo Decision Tree



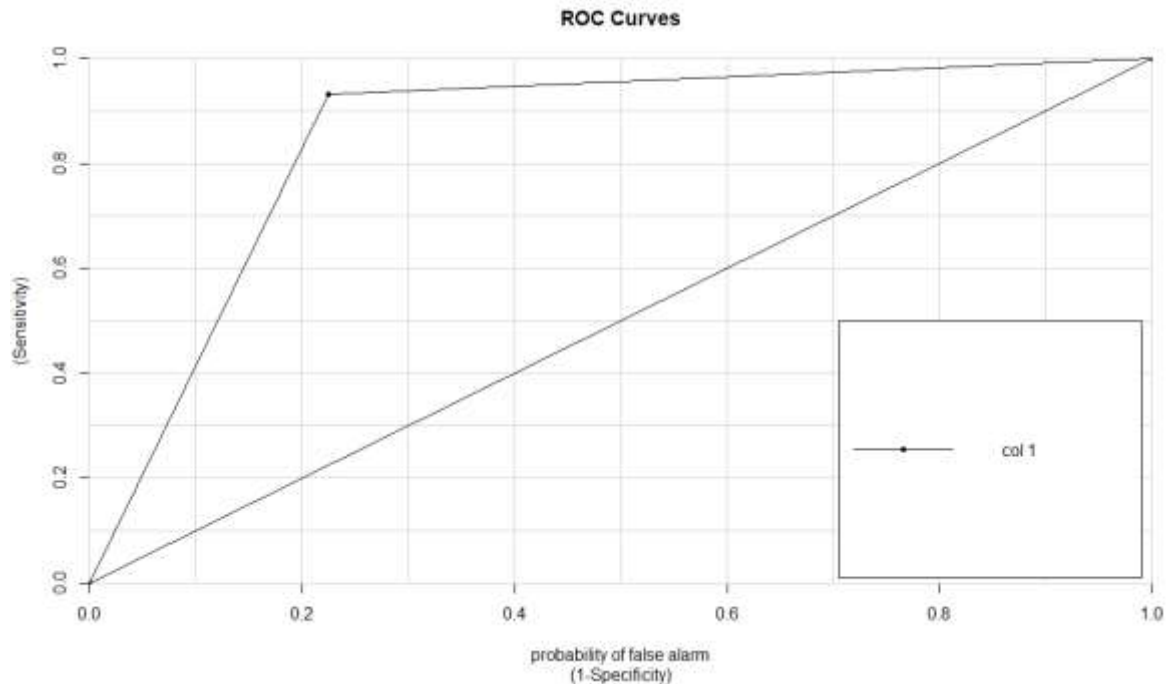
Abril Adaboost



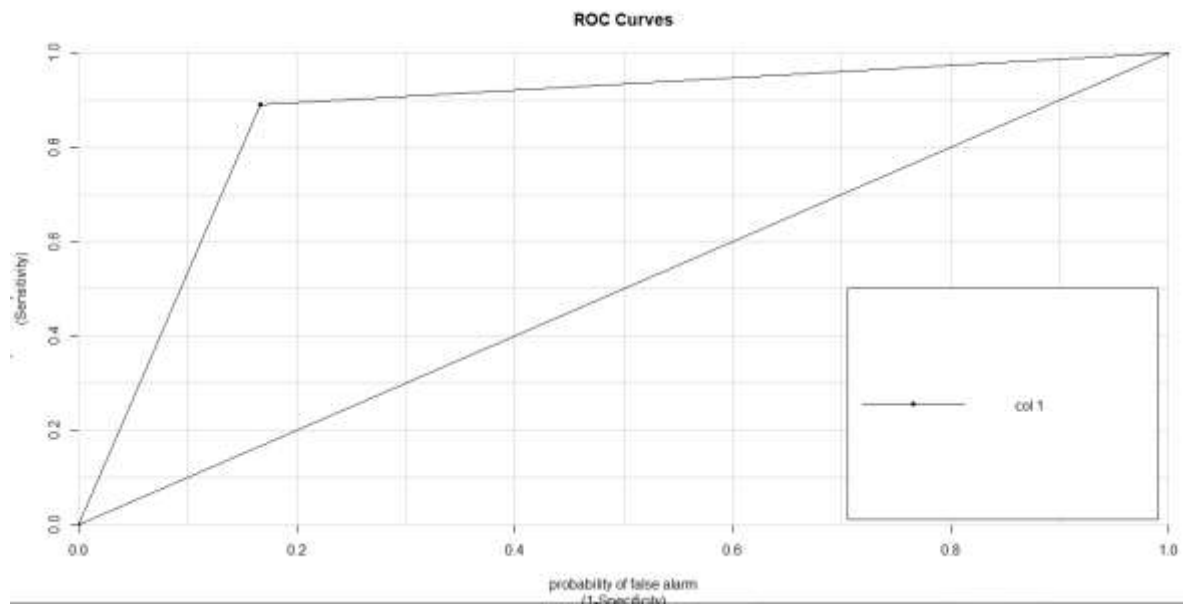
Abril Decision Tree



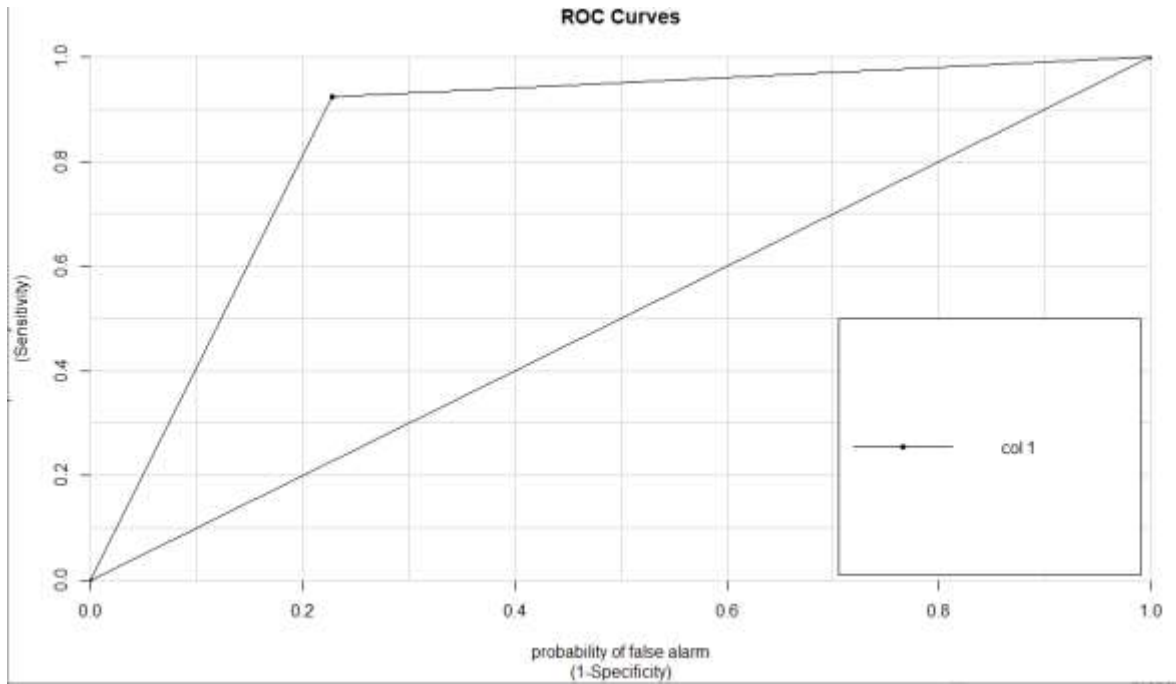
Mayo Adaboost



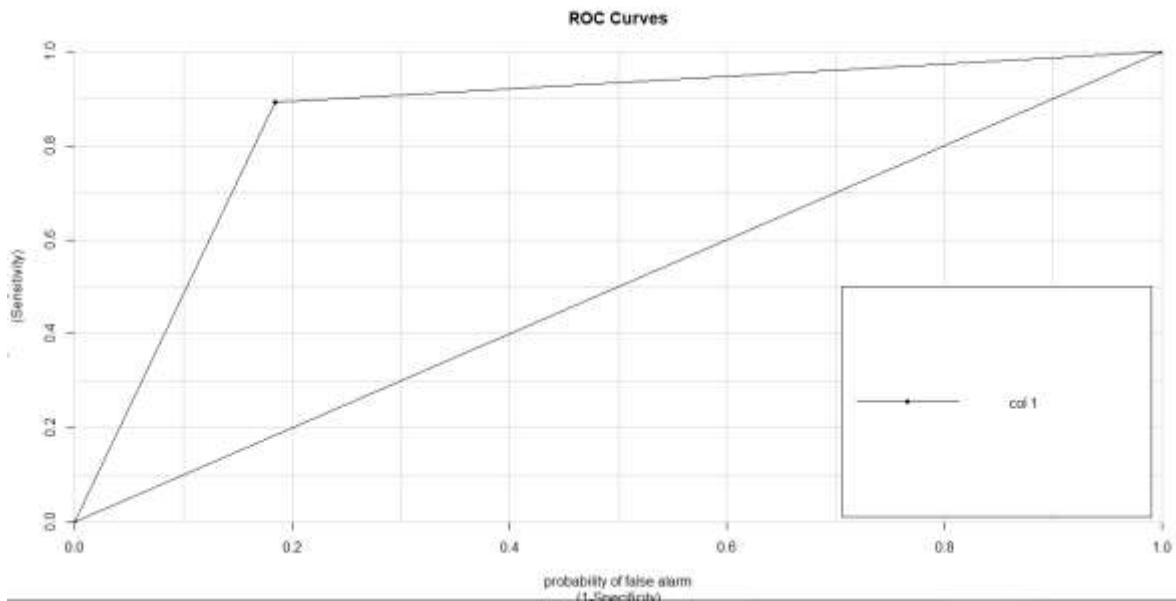
Mayo Decision Tree



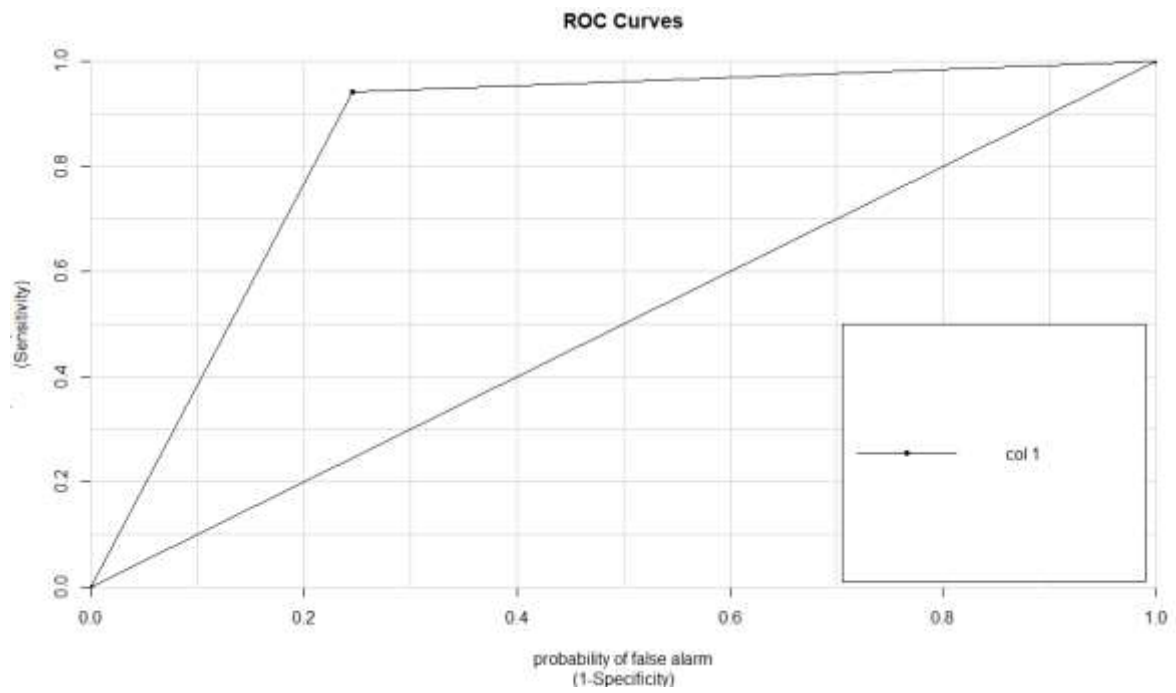
Junio Adaboost



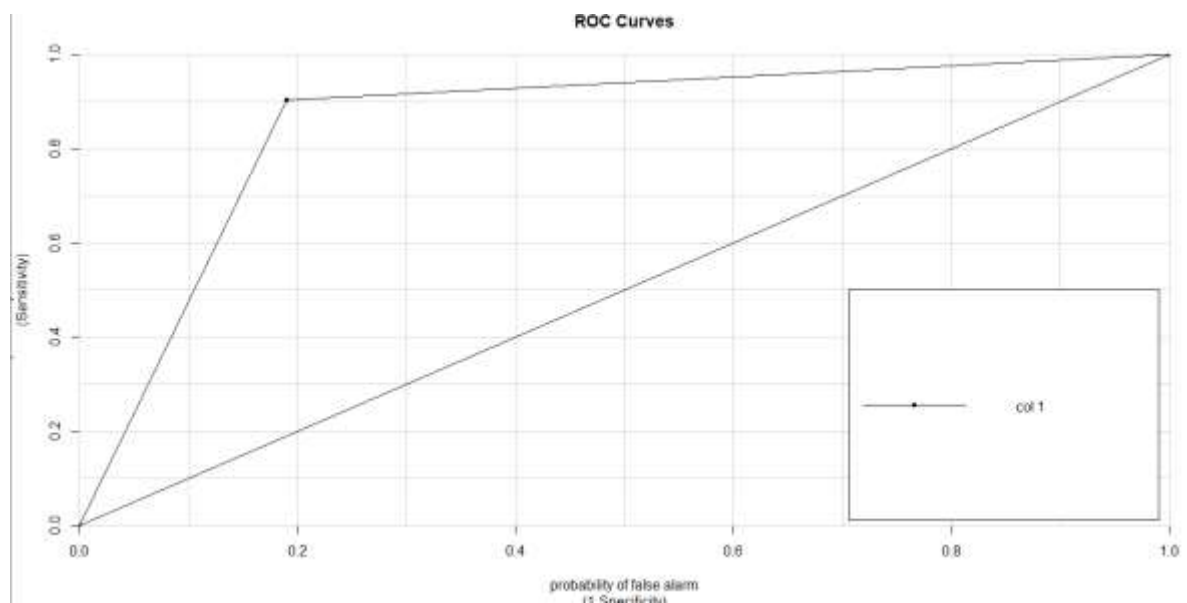
Junio Decision Tree



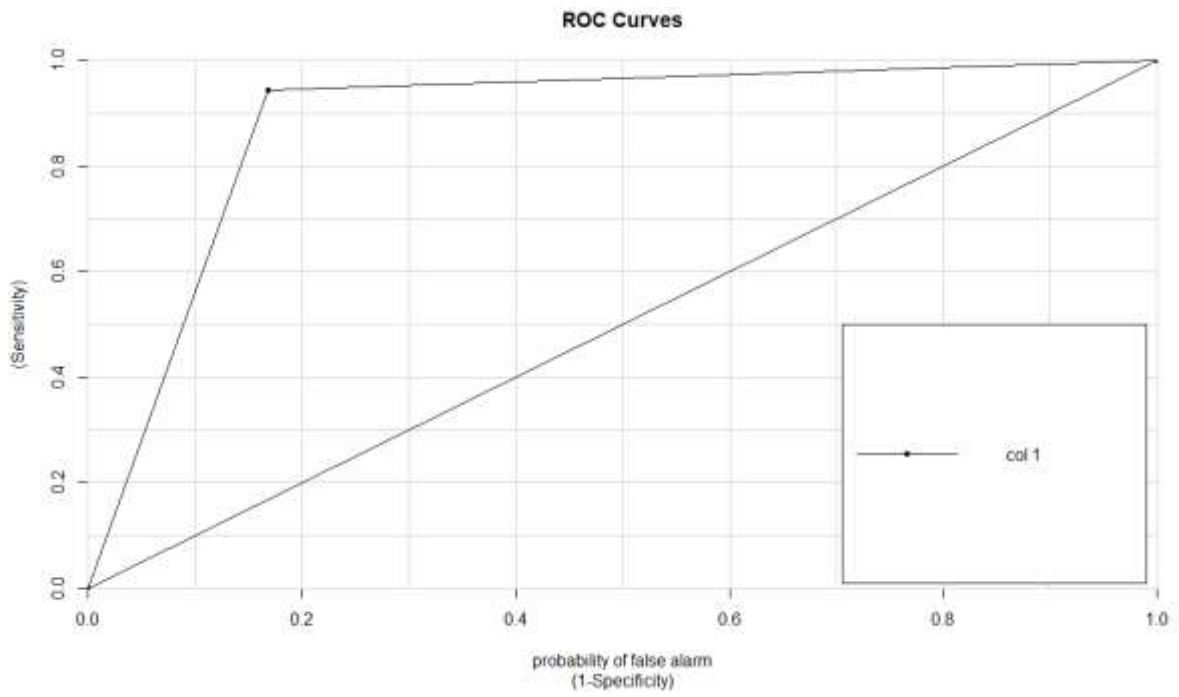
Julio Adaboost



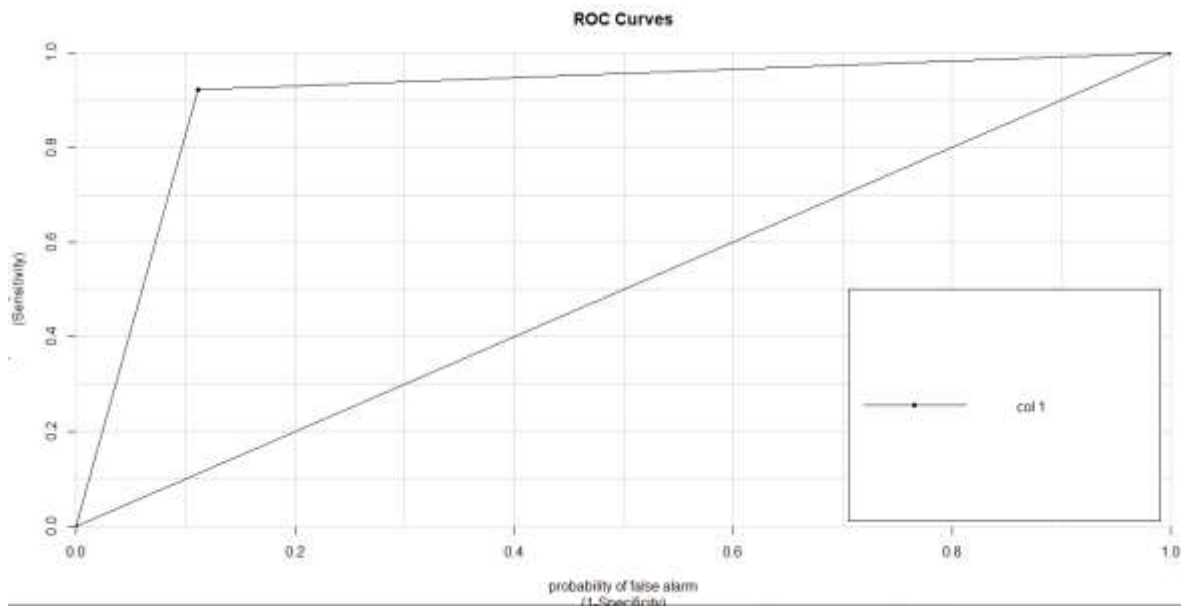
Julio Decision Tree



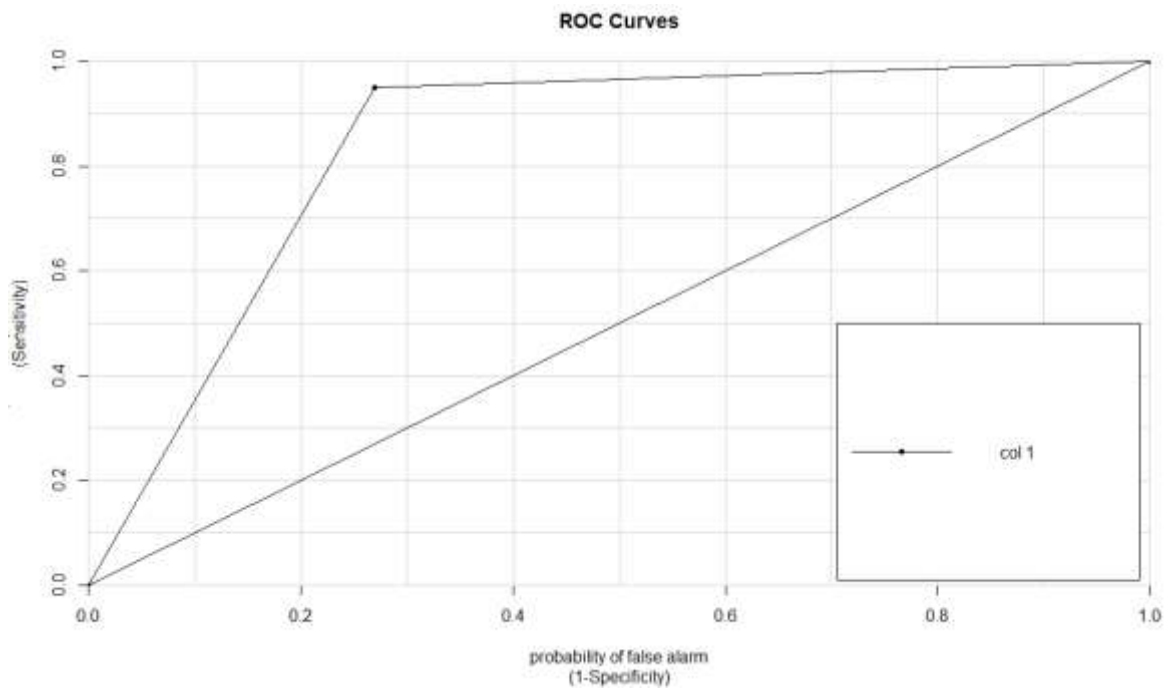
Agosto Adaboost



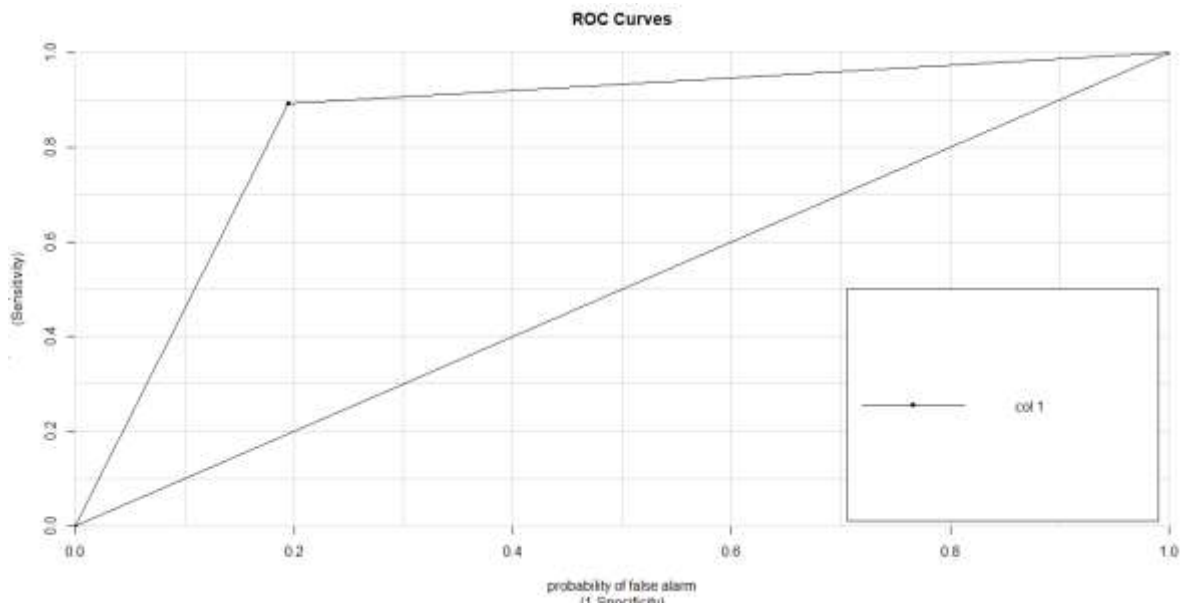
Agosto Decision Tree



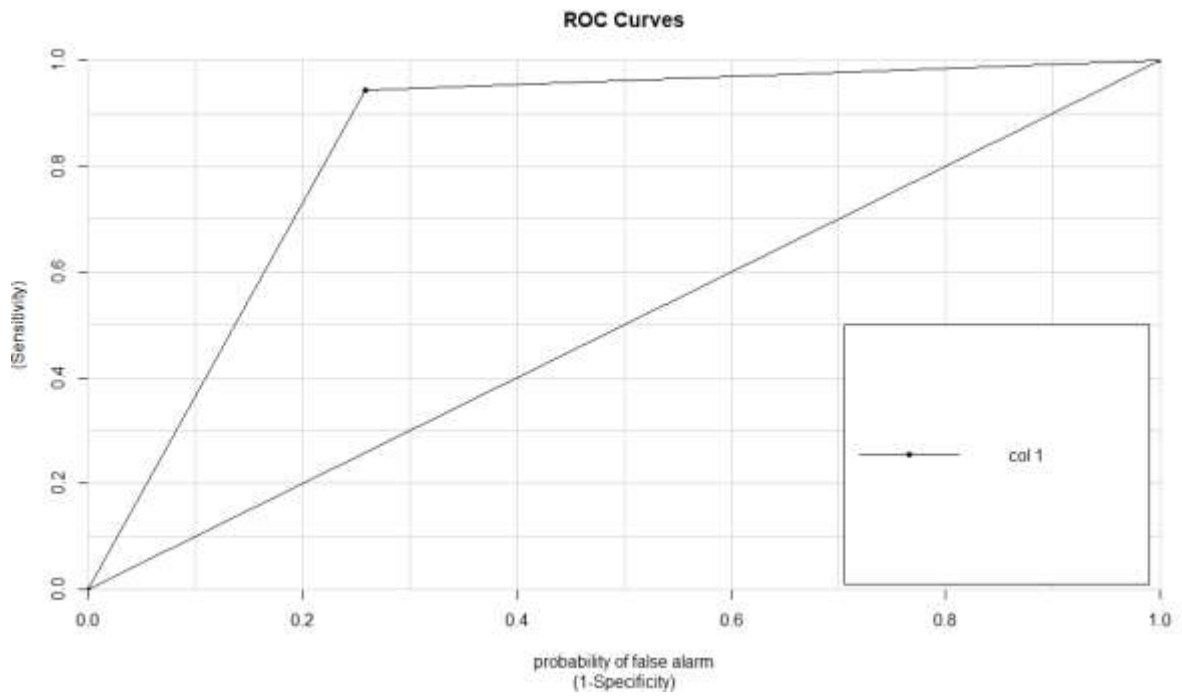
Septiembre Adaboost



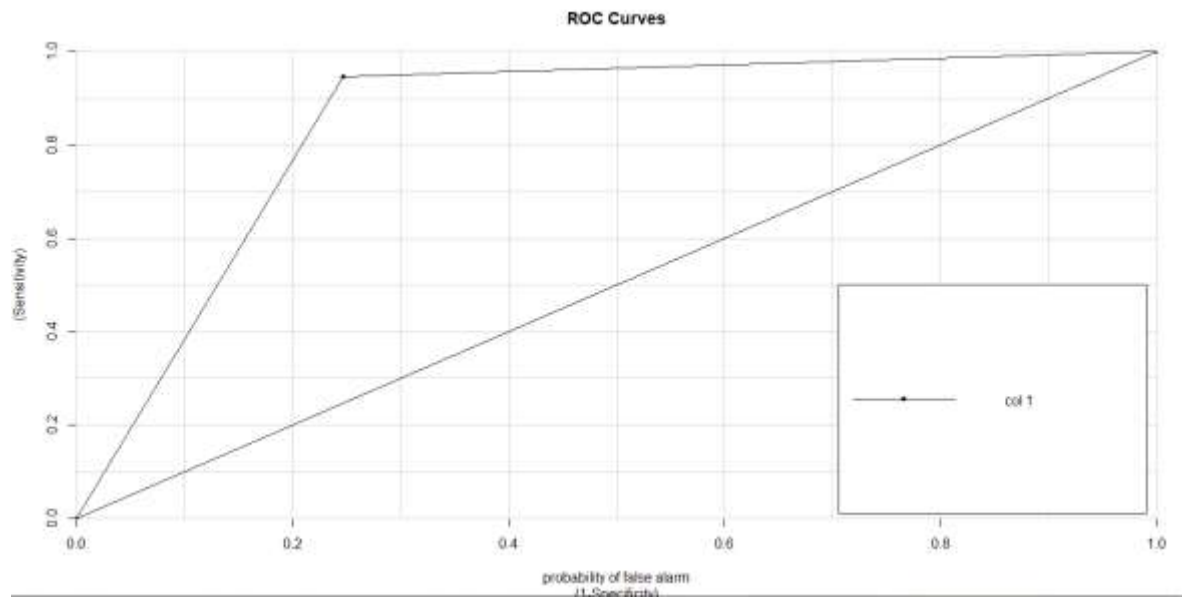
Septiembre Decision Tree



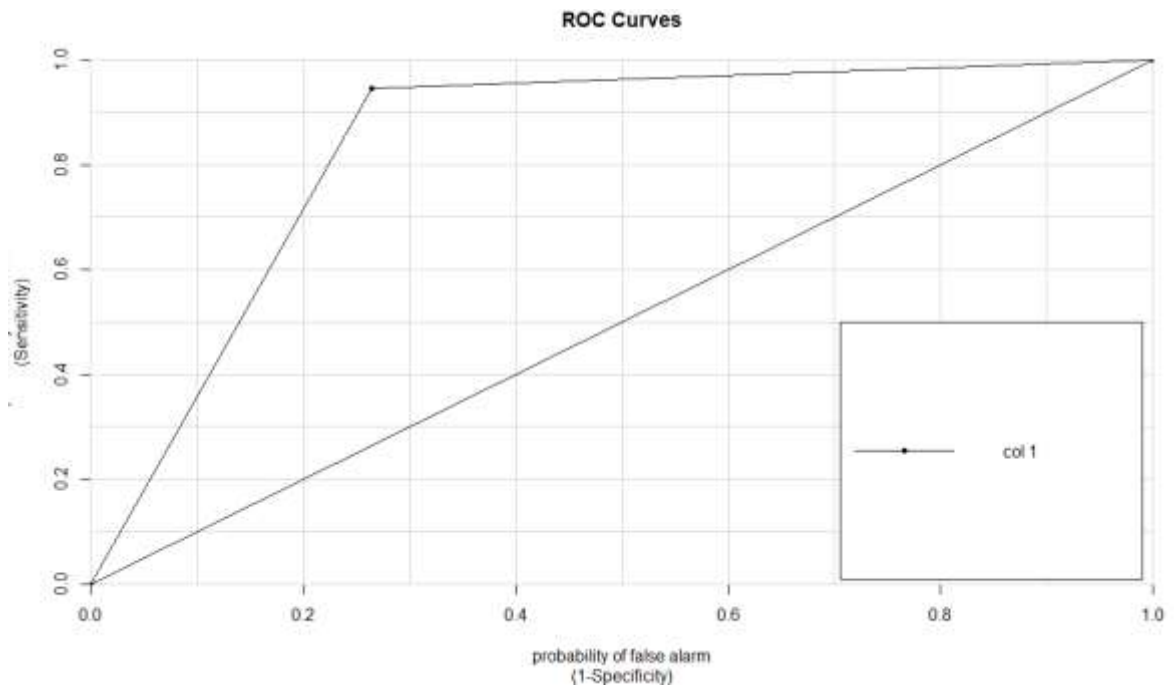
Octubre Adaboost



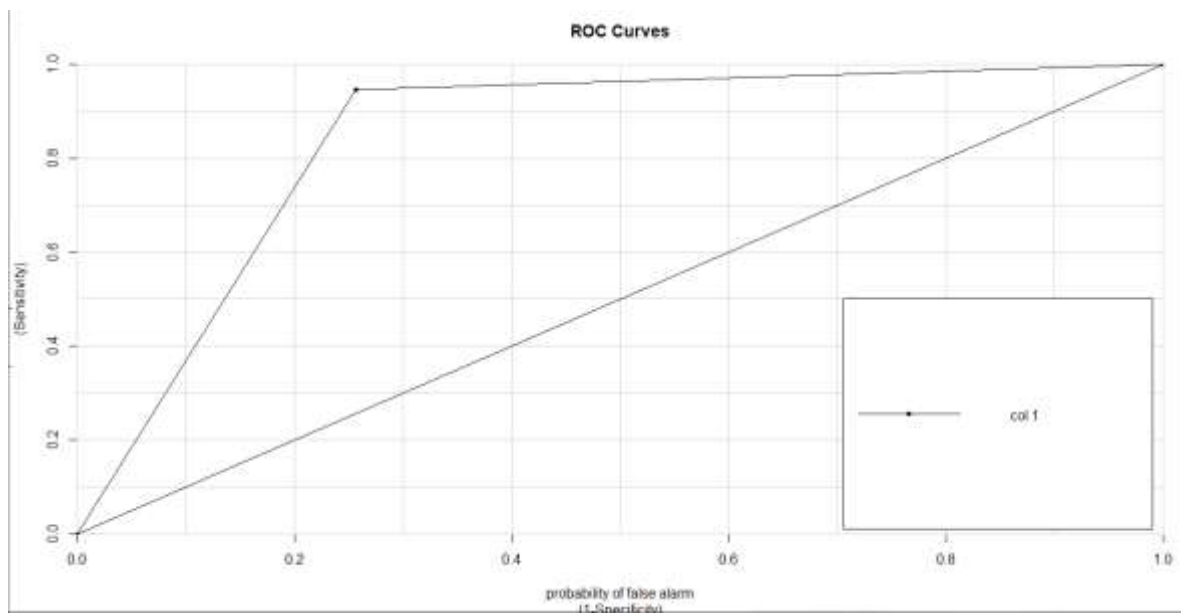
Octubre Decision Tree



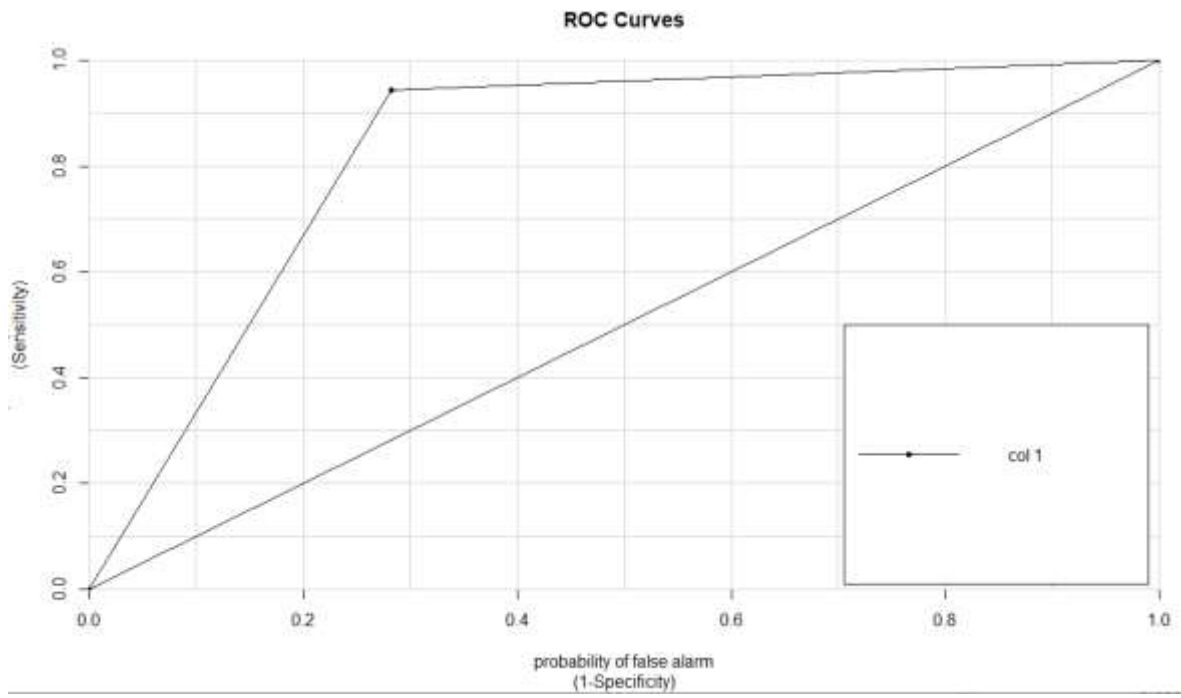
Noviembre Adaboost



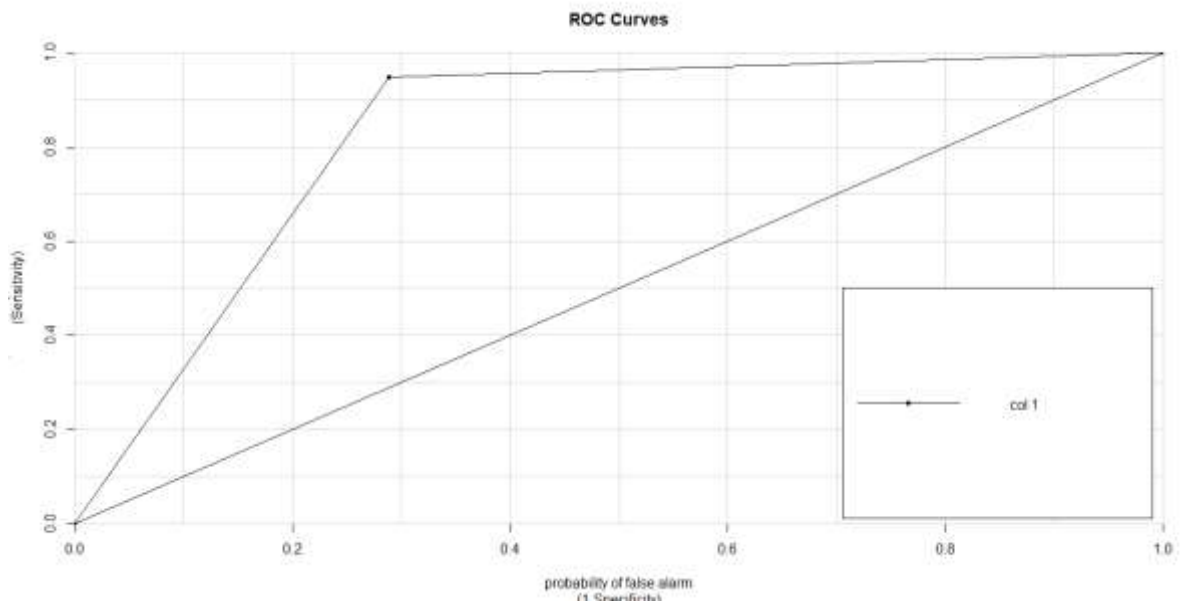
Noviembre Decision Tree



Diciembre Adaboost



Diciembre Decision Tree



ANEXO 15: Pasando datos de base de datos Origen (Colegio_Ingenieros) a base de datos destino (CIP).

-----para tabla CAPITULO

```
merge [CIP].[dbo].[Capitulo] as cap2
using (select * from [Colegio_Ingenieros].[dbo].[tbl_capitulo]) as cap1
on cap1.CodCap collate Modern_Spanish_CI_AS = cap2.CodCap collate
Modern_Spanish_CI_AS
when not matched then
insert(CodCap,NomCap,Orden_Cap)
values(CodCap,NomCap,orden_cap);
```

----- para tabla ESPECIALIDAD

```
merge [CIP].[dbo].[Especialidad] as esp2
using (select * from [Colegio_Ingenieros].[dbo].[tbl_Especialidad] where CodCap !=
'8' ) as esp1
on esp1.CodEsp collate Modern_Spanish_CI_AS = esp2.CodEsp collate
Modern_Spanish_CI_AS
when not matched then
insert(CodEsp,NomEsp,CodCap)
values(CodEsp,NomEsp,CodCap);
```

-----para tabla COBRO

```
merge [CIP].[dbo].[Cobro] as cob2
using (select * from [Colegio_Ingenieros].[dbo].[tbl01cdv]) as cob1
on cob1.codcdv collate Modern_Spanish_CI_AS = cob2.CodCdv collate
Modern_Spanish_CI_AS
when not matched then
insert (CodCdv,NomCdv,diaven,CtaCol,CtaCoE,CtaCrL,CtaCrE)
values (codcdv,nomcdv,diaven,CtaCol,CtaCoE,CtaCrL,CtaCrE);
```

-----para tabla CAJERO_ATENCIONCLIENTE

```
merge [CIP].[dbo].[Cajero_AtencionCliente] as caj2
using (select * from [Colegio_Ingenieros].[dbo].[tbl01ven]) as caj1
on caj1.codven collate Modern_Spanish_CI_AS = caj2.CodVen collate
Modern_Spanish_CI_AS
when not matched then
insert (CodVen,NomVen)
values (codven,nomven);
```

-----para tabla COLEGIADO

```
merge [CIP].[dbo].[Colegiado] as col2
using(select distinct
(c.nroCIP),c.codcli,c.nomcli,c.nrodni,c.ruccli,c.dircli,c.telcli,c.TelCel,c.email,c.
fecnac,c.feccol,
c.estciv,p.nompai,cat.nomcat,e.CodEsp,h.estado, DATEDIFF(YEAR,c.FecNac,GETDATE()) as
Edad from Colegio_Ingenieros.dbo.mst01cli as c
inner join Colegio_Ingenieros.dbo.tbl01pai as p on c.codpai = p.codpai
inner join Colegio_Ingenieros.dbo.tbl01cac as cat on c.codcat = cat.codcat
left join Colegio_Ingenieros.dbo.tbl_especialidad as e on c.codesp = e.CodEsp
left join Colegio_Ingenieros.dbo.habil2204201 as h on c.nroCIP = h.cip
```



```

where c.nroCIP not like '' and c.nroCIP <> '72288' and c.nroCIP <> '94001' and
c.nroCIP != '.' and c.NroCIP != 'G-001') as col1
on col1.nroCIP collate Modern_Spanish_CI_AS = col2.NroCIP collate
Modern_Spanish_CI_AS
when not matched then
insert(NroCIP,CodCli,NomCli,NroDni,DirCli,RucCli,TelCli,TelCel,Email,FecNac,Edad,Fec
Col,EstCivil,NomPai,NomCat,CodEsp,Estado_Habilidad)
values(nroCIP,codcli,nomcli,nrodni,dircli,ruccli,telcli,telcel,email,feclac,Edad,fecl
col,estciv,nompai,nomcat, codesp, estado);

```

-----para tabla FAMILIAR

```

merge [CIP].[dbo].[Familiar] as fam2
using(SELECT a.Codn, a.NomCon,a.FecNac,
DATEDIFF(YEAR,a.FecNac,GETDATE()) as Edad,
a.NroDni, c.nroCIP, b.NomPar
FROM Colegio_Ingenieros.dbo.Dtl01Con AS a
INNER JOIN Colegio_Ingenieros.dbo.mst01cli AS c ON a.Codn = c.codcli
INNER JOIN Colegio_Ingenieros.dbo.Tbl_Parentesco AS b ON a.CodPar =
b.CodPar
right JOIN CIP.dbo.Colegiado as cc on c.nroCIP collate Modern_Spanish_CI_AS
= cc.NroCIP collate Modern_Spanish_CI_AS
where cc.NroCIP collate Modern_Spanish_CI_AS = c.nroCIP collate
Modern_Spanish_CI_AS
)as fam1
on fam1.NomPar collate Modern_Spanish_CI_AS = fam2.NomPar collate
Modern_Spanish_CI_AS
when not matched then
insert (Familiar,FechaNac,Edad,NumDni,NroCIP,NomPar)
values (NomCon,FecNac,Edad,NroDni,nroCIP,NomPar);

```

-----para tabla HISTORIAL DE PAGO

-----TABLA PAGOS

-----PROCESO DE LIMPIEZA Y CONSISTENCIA

```

PROCESO-----1
SELECT * INTO Filtro_Padre FROM Colegio_Ingenieros.dbo.mst01ccc

PROCESO-----2

SELECT * into Filtro_Hijo from Filtro_Padre where glosa like 'CUOTA MENSUAL%'

delete from Filtro_Padre where glosa like 'CUOTA MENSUAL%'

PROCESO-----3

insert into Filtro_Hijo select * from Filtro_Padre where glosa like 'MIGRACION CUOTA
ORDINARIA%'

delete from Filtro_Padre where glosa like 'MIGRACION CUOTA ORDINARIA%'

PROCESO-----4

insert into Filtro_Hijo select * from Filtro_Padre where glosa like 'CUOTA
ORDINARIA%'

delete from Filtro_Padre where glosa like 'CUOTA ORDINARIA%'

```

PROCESO-----5

```
insert into Filtro_Hijo select * from Filtro_Padre where glosa like 'Migracion'
```

```
delete from Filtro_Padre where glosa like '%Migracion%'
```

PROCESO-----6

```
insert into Filtro_Hijo select * from Filtro_Padre where monto >17 and monto < 21  
order by monto
```

```
delete from Filtro_Padre where monto >17 and monto < 21
```

PROCESO-----7

```
insert into Filtro_Hijo select * from Filtro_Padre where glosa like 'SALDO%' and  
monto >0 and monto < 21
```

```
delete from Filtro_Padre where glosa like 'SALDO%' and monto >0 and monto < 21
```

-----PROCESO DE ELIMINACION DE DATOS IRRELEVANTES

```
delete from Filtro_Padre where glosa like 'ALQ%'  
delete from Filtro_Padre where glosa like ''  
delete from Filtro_Padre where glosa like '%ANTICIPO CERTIF. HABILIDAD%'  
delete from Filtro_Padre where glosa like '%COMISION%'  
delete from Filtro_Padre where glosa like '%ANTICIPO%'  
delete from Filtro_Padre where glosa like '%ARTICULOS%'  
delete from Filtro_Padre where glosa like '%INGRESOS%'  
delete from Filtro_Padre where glosa like '%AUTOCAD%'  
delete from Filtro_Padre where glosa like '%ARCG%'  
delete from Filtro_Padre where glosa like '%ASCRI%'  
delete from Filtro_Padre where glosa like '%Y RENO%'  
delete from Filtro_Padre where glosa like '%ADSC%'  
delete from Filtro_Padre where glosa like '%Carga Manual%'  
delete from Filtro_Padre where glosa like '%CERTIFICADO%'  
delete from Filtro_Padre where glosa like '%INSCRIPCION%'  
delete from Filtro_Padre where glosa like 'SALDO%'  
delete from Filtro_Padre where glosa like '%TRAMITE%'  
delete from Filtro_Padre where glosa like '%SOLAPE%'  
delete from Filtro_Padre where glosa like '%INGRESO%'  
delete from Filtro_Padre where glosa like '%POLO%'  
delete from Filtro_Padre where glosa like '%PROC%'  
delete from Filtro_Padre where glosa like '%POWE%'  
delete from Filtro_Padre where glosa like '%RECAR%'  
delete from Filtro_Padre where glosa like '%RIEG%'  
delete from Filtro_Padre where glosa like '%SAP%'  
delete from Filtro_Padre where glosa like '%TALLER%'  
delete from Filtro_Padre where glosa like '%TRAM.%'  
delete from Filtro_Padre where glosa like '%GORROS%'  
delete from Filtro_Padre where glosa like '%CURSO%'  
delete from Filtro_Padre where glosa like '%INSCR%'  
delete from Filtro_Padre where glosa like '%LLAVER%'  
delete from Filtro_Padre where glosa like '%MULTA%'  
delete from Filtro_Padre where glosa like '%PERITA%'  
delete from Filtro_Padre where glosa like '%CANCHA%'  
delete from Filtro_Padre where glosa like '%CERTIFI%'  
delete from Filtro_Padre where glosa like '%CHALECO%'  
delete from Filtro_Padre where glosa like '%COLEGIATURA%'  
delete from Filtro_Padre where glosa like '%CERT%'  
delete from Filtro_Padre where glosa like '%DUPLI%'
```

```

delete from Filtro_Padre where glosa like '%DPLOM%'
delete from Filtro_Padre where glosa like '%INSC.%'
delete from Filtro_Padre where glosa like '%INSTALADOR%'
delete from Filtro_Padre where glosa like '%Carga%'
delete from Filtro_Padre where glosa like '%AUD%'
delete from Filtro_Padre where glosa like '%DIPLO%'
delete from Filtro_Padre where glosa like '%DISE%'
delete from Filtro_Padre where glosa like '%TRAM%'
delete from Filtro_Padre where glosa like '%OFF%'
delete from Filtro_Padre where glosa like '%EXCEL%'
delete from Filtro_Padre where glosa like '%EMIT%'
delete from Filtro_Padre where glosa like '%LEY%'
delete from Filtro_Padre where glosa like '%PROJ%'
delete from Filtro_Padre where glosa like '%PAGO%'

```

-----VISTA (AYUDA AL PROCESO DE PASAR DATOS)

```

create view V_PagosM as
select
RTRIM(SUBSTRING(glosa,CHARINDEX(':',glosa)+2, LEN(glosa))) AS NroCIP,
MONTH(c.fecha) as mes,
CONVERT(date,c.fecha,101) as fecha,
CONVERT(date,c.fven,101) as fven,
CONVERT(date,c.uabo,101) as uabo,
c.monto,
pago = case
    when (c.saldo = 0) then c.monto
    when c.saldo = c.monto then 0
    else c.monto-c.saldo
end,
c.saldo,
dias = case
    when saldo = 0 then convert(varchar,(DATEDIFF(dd,c.fven,c.uabo)))
    else '-'
end,
c.codcdv,
c.codven,
c.glosa as DESCRIPCION
from Filtro_Hijo as c
WHERE glosa LIKE '%CUOTA MENSUAL%' and glosa != 'Cuota Mensual (09-2013)/'

```

----- para tabla PAGOS

```

merge [CIP].[dbo].[Pagos] as pag2
using(select
v.NroCIP,v.mes,v.fecha,v.fven,v.uabo,v.monto,v.pago,v.saldo,v.dias,v.codcdv,v.codven
from V_PagosM as v
right join colegiado as c on v.NroCIP collate Modern_Spanish_CI_AS= c.NroCip
collate Modern_Spanish_CI_AS
where c.NroCip collate Modern_Spanish_CI_AS = v.nroCip collate
Modern_Spanish_CI_AS
)as pag1
on pag1.NroCIP collate Modern_Spanish_CI_AS = pag2.NroCIP collate
Modern_Spanish_CI_AS
when not matched then
insert
(NroCIP,Mes_Pago,FechaApertura,FechaVencimiento,FechaPago,monto_programado,monto_pag
ado,Saldo,dias,CodCdv,CodVen)
values ( NroCIP,mes,fecha,fven,uabo,monto,pago,saldo,dias,codcdv,codven);

```

-----VISTA 2 (PARA DATOS FALTANTES)

```
create view V_PagosM2 as
select
i.nroCIP as NroCIP,
MONTH(c.fecha) as mes,
CONVERT(date,c.fecha,101) as fecha,
CONVERT(date,c.fven,101) as fven,
CONVERT(date,c.uabo,101) as uabo,
c.monto,
pago = case
    when (c.saldo = 0) then c.monto
    when c.saldo = c.monto then 0
    else c.monto-c.saldo
end,
c.saldo,
dias = case
    when saldo = 0 then convert(varchar,(DATEDIFF(dd,c.fven,c.uabo)))
    else '-'
end,
c.codcdv,
c.codven,
i.nroCIP as CIP,
c.glosa as DESCRIPCION
from Filtro_Hijo as c inner join [Colegio_Ingenieros].dbo.mst01cli i
on c.codcli = i.codcli
WHERE glosa = 'Cuota Mensual (09-2013)'
```

-----TABLA PAGOS (DATOS FALTANTES)

```
INSERT INTO Pagos
(NroCIP,Mes_Pago,FechaApertura,FechaVencimiento,FechaPago,monto_programado,monto_pago,Saldo,dias,CodCdv,CodVen)
SELECT NroCIP,mes,fecha,fven,uabo,monto,pago,saldo,dias,codcdv,codven FROM V_PagosM2
```

-----ACTUALIZAR CAMPO DE NUMERO DE HIJOS DE COLEGIADOS

```
merge [CIP].[dbo].[Colegiado] as col2
using (select NroCIP,count(NomPar) as N_Hijos
from Familiar where NomPar like 'HIJO(A)' group by NroCIP ) as col1
on col2.NroCIP collate Modern_Spanish_CI_AS = col1.NroCIP collate
Modern_Spanish_CI_AS
when matched then
update set
                col2.Nro_Hijos = col1.N_Hijos
;
```

-----ACTUALIZAR CAMPO GENERO EN COLEGIADO

```
--CAMBIAR de FLOAT a CHAR(7) en la tabla NROCIP de BASE DATOS: COLEGIADOGENERO
--CAMBIAR de VARCHAR(255) a CHAR(1) en la tabla SEXO de BASE DATOS: COLEGIADOGENERO
```

```
merge [CIP].[dbo].[Colegiado] as col2
using (select * from ColegiadoGenero.dbo.GENERO) as gen1
on col2.NroCIP collate Modern_Spanish_CI_AS = gen1.NroCIP collate
Modern_Spanish_CI_AS
when matched then
update set
                col2.Genero = gen1.SEXO
;
```

-----ACTUALIZAR DNI DE TABLA EXCEL_COLEGIADO A COLEGIADO

```
merge [CIP].[dbo].[Colegiado] as col2
using (select * from Excel_Colegiados.dbo.Seguro where convert(varchar,dni) not like
'NULL') as dni
on col2.Nomcli collate Modern_Spanish_CI_AS = dni.Nombre collate
Modern_Spanish_CI_AS
when matched then
update set
    col2.NroDni = dni.[N° CIP]
;
```

-----ACTUALIZAR TIPO DE SEGURO DE TABLA EXCEL_COLEGIADO A COLEGIADO

```
merge [CIP].[dbo].[Colegiado] as col2
using (select * from Excel_Colegiados.dbo.Seguro) as dni
on col2.Nomcli collate Modern_Spanish_CI_AS = dni.Nombre collate
Modern_Spanish_CI_AS
when matched then
update set
    col2.Tipo_Seguro = dni.[TIPO DE SEGURO]
;
```

-----ACTUALIZAR CATEGORIA DESDE EXCEL_CIP a CIP

```
merge [CIP].[dbo].[Colegiado] as col2
using (select * from EXCEL_CIP.dbo.Colegiados) as col1
on col2.NroCIP collate Modern_Spanish_CI_AS = col1.CIP collate Modern_Spanish_CI_AS
when matched then
update set
    col2.NomCat = col1.Categoria
;
```

-----ACTUALIZAR HABILIDAD DESDE EXCEL_CIP a CIP

```
merge [CIP].[dbo].[Colegiado] as col2
using (select * from EXCEL_CIP.dbo.Colegiados) as col1
on col2.NroCIP collate Modern_Spanish_CI_AS = col1.CIP collate Modern_Spanish_CI_AS
when matched then
update set
    col2.Estado_habilidad = col1.Habilidad
;
```

-----SANEANDO DATA DE TIPO DE SEGURO DE TABLA COLEGIADO

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'SIN REGISTROS'
where Tipo_Seguro like 'NO'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'SIN REGISTROS'
where Tipo_Seguro is null
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'SIN REGISTROS'
where Tipo_Seguro like '%REGIS%'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'PACIFICO SEGUROS'
where Tipo_Seguro like 'PASIFICO SEGUROS'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'ESSALUD'
where Tipo_Seguro like 'EESALUD'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'SIN REGISTROS'
where Tipo_Seguro like 'NO TIENE'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'PACIFICO SEGUROS'
where Tipo_Seguro like 'PACIFICIO SEGUROS'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'PACIFICO SEGUROS'
where Tipo_Seguro like 'PACÍFICO SEGUROS'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'PACIFICO EPS'
where Tipo_Seguro like 'PACÍFICO EPS'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'RIMAC EPS'
where Tipo_Seguro like 'RIMAC EPS'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'RIMAC EPS'
where Tipo_Seguro like 'RÍMAC EPS'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'RIMAC SEGUROS'
where Tipo_Seguro like 'RÍMAC SEGUROS'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'SANITAS PERU EPS'
where Tipo_Seguro like 'SANITAS PERÚ EPS'
```

```
update CIP.dbo.Colegiado set
Tipo_Seguro = 'SIN REGISTROS'
where Tipo_Seguro like 'DNI INCOMPLETO'
```

```
update CIP.dbo.Colegiado set
Genero = 'M'
where NroCIP like ' 186179'
```

```

update CIP.dbo.Colegiado set
NroCIP = '186179'
where NroCIP like ' 186179'

Update CIP.dbo.Colegiado set
CodEsp = '186179'
where NroCIP like ' 186179'

--SUSTITUYENDO NULL

UPDATE Colegiado SET Nro_Hijos=0 WHERE Nro_Hijos IS NULL

UPDATE Colegiado set Edad= '118' where Edad = '*'

UPDATE Colegiado set Estado_habilidad= 'Sin Referencia' where Estado_habilidad IS
NULL

UPDATE Colegiado SET EstCivil='-' WHERE EstCivil = ''

UPDATE Pagos SET CodVen='V0000' WHERE CodVen is null

UPDATE Pagos SET CodCdv='02' WHERE CodCdv is null

UPDATE Pagos set FechaPago = NULL where Monto_Pagado = 0

UPDATE Pagos set dias = DATEDIFF(dd,FechaVencimiento,FechaPago) where saldo!=0

update Pagos set dias = '-' where dias is null

update Pagos set FechaApertura = '01/05/2007'
where FechaApertura = '0007-05-01' and nrocip like '46930'

update Pagos set FechaApertura = '01/06/2007'
where FechaApertura = '0007-06-01' and nrocip like '46930'

update Pagos set FechaApertura = '01/05/2015'
where FechaApertura = '0015-05-01' and nrocip like '5549'

update Pagos set FechaApertura = '01/06/2015'
where FechaApertura = '0015-06-01' and nrocip like '5549'

update Pagos set FechaApertura = '01/07/2015'
where FechaApertura = '0015-07-01' and nrocip like '5549'

update Pagos set FechaApertura = '01/08/2015'
where FechaApertura = '0015-08-01' and nrocip like '5549'

update Pagos set FechaApertura = '01/09/2015'
where FechaApertura = '0015-09-01' and nrocip like '5549'

update Pagos set FechaApertura = '01/10/2015'
where FechaApertura = '0015-10-01' and nrocip like '5549'

update Pagos set FechaApertura = '01/11/2015'
where FechaApertura = '0015-11-01' and nrocip like '5549'

update Pagos set FechaApertura = '01/12/2015'
where FechaApertura = '0015-12-01' and nrocip like '5549'
update Pagos set dias = 0 where dias like '-'
update Pagos set dias = -1 where dias = 0 and FechaPago is not null

```