

 | UNIVERSIDAD
SEÑOR DE SIPÁN

**FACULTAD DE INGENIERIA, ARQUITECTURA Y
URBANISMO**

ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

TESIS

**TECNICAS DE MINERIA DE DATOS EN EL DISEÑO
DE APLICACIONES PARA MEJORAR EL ANALISIS
DE LA GESTION DE RECURSOS HUMANOS DEL
PROYECTO ESPECIAL ALTOMAYO**

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

Autor:

Bach. Chero Montalbán José Gabriel

Asesor Metodológico:

Mg. Oliver Vasquez Leyva

Línea de Investigación:

Infraestructura, tecnología y medio ambiente

Pimentel – Perú

2020

DEDICATORIA

Dedicado a mis padres que con
esfuerzo me sacaron adelante.

AGRADECIMIENTOS

A todas las personas que contribuyeron de alguna forma u otra para que este proyecto pueda salir adelante.

INDICE

DEDICATORIA	2
AGRADECIMIENTOS	3
RESUMEN	8
ABSTRACT	9
INTRODUCCIÓN	10
CAPÍTULO I: PROBLEMA DE INVESTIGACIÓN	11
1.1 Situación problemática	11
1.2 Formulación del problema	12
1.3 Delimitación de la investigación	12
1.4 Justificación e importancia de la Investigación	12
1.5 Limitaciones de la investigación	13
1.6 Objetivos de la investigación	13
1.6.1. Objetivo general	13
1.6.2 Objetivos específicos	13
CAPÍTULO II: MARCO TEORICO	14
2.1. Antecedentes de estudio	14
2.1.1. INTERNACIONAL	14
2.1.2 NACIONALES:	15
2.1.3 LOCAL	16
2.2. Estado del arte	16
2.3. Base teórica científica	18
2.3.1. Tecnología Web	18
2.4. Definición de la terminología	27
2.4.1. CRISP – DM:	27
2.4.2. Minería de datos:	27
2.4.3. Semma:	27
2.4.4. Tecnología web:	27
CAPITULO III MARCO METODOLOGICO	28
3.1. Tipo y Diseño de investigación	28
3.1.1 Tipo de la investigacion	28
3.1.2 Diseño de la investigacion	28
3.2. Población y muestra	28
3.2.1 Población	28
3.2.2 Muestra	28
3.3. Hipótesis	29
3.4. Variable	30
3.4.1 Variable independiente	30
3.4.2 Variable dependiente	30
3.6. Aborde Metodologico, técnicas e instrumentos de recolección de datos	31



3.6.1.	Aborde Metodológico.....	31
3.6.2.	Técnicas de recolección de datos.....	31
3.6.3.	Instrumentos de recolección de datos	31
3.7.	Procedimientos para la recolección de datos.....	32
3.8.	Análisis estadístico e interpretación de datos.....	32
3.9.	Principios éticos	32
	Criterios	32
	Características científicas del criterio	32
	Confiabilidad	32
	Validación.....	32
3.10.	Criterios de rigor científico	33
	Criterios	33
	Características éticas del criterio.....	33
CAPITULO IV: ANALISIS E INTERPRETACION DE LOS RESULTADOS		34
4.1.	Resultados en tablas y gráficos de la propuesta.....	34
4.2.	Discusión de resultados.....	40
CAPITULO V: PROPUESTA DE LA INVESTIGACION		42
5.1.	Generalidades de la propuesta.....	42
5.2.	Esquematzación por fases y sub procedimientos	42
5.3.	Desarrollo de fases	46
	Fase Modelado de datos.....	46
5.3.1.	Comprensión de Negocio.....	46
5.3.2.	Comprensión de datos.....	46
5.3.3.	Selección de datos.....	50
5.3.4.	Transformación de datos.....	51
5.3.5.	Modelado	57
5.3.6.	Evaluación	60
5.3.7.	Fase Modulo Web de Análisis de Algoritmos	60
CAPITULO VI – CONCLUSIONES Y RECOMENDACIONES.....		66
6.1.	Conclusiones	66
	Recomendaciones	67
	REFERENCIAS.....	68
	ANEXOS_.....	70



ÍNDICE DE FIGURAS

Figura 1: Proceso de Minería de Datos.....	22
Figura 2: Clasificación de las Tecnicas de Minería de Datos	22
Figura 3: Arbol de decisiones para evaluar el riesgo de un cliente.....	24
Figura 4: Red Neuronal con pesos asociados a cada nodo	25
Figura 5: Fases de modelo CRISP – DM.....	26
Figura 6: Índice de error entre algoritmos empleados	35
Figura 7: Índice de tiempo de entrenamiento entre algoritmos.....	36
Figura 8: Numero de meses mínimo para generar entrenamiento.....	37
Figura 9: Comparación eficiencia de algoritmos en pruebas	38
Figura 10: Comparacion tiempo de ejecucion del modelo.....	39
Figura 11: Flujograma de desarrollo de la propuesta.....	45
Figura 12: Vista en POSTGRESQL de base de datos.....	47
Figura 13: Vista de esquemas en POSTGRESQL de base de datos	48
Figura 14: Vista de entidades en POSTGRESQL de bases de datos	49
Figura 15:Codigo de seleccion de datos de las diferentes base de datos	50
Figura 16: Datos mostrados en R listo para ser transformados.....	51
Figura 17:Codigo para realizar la transformacion de datos	51
Figura 18: Visualizacion de datos transformados y mostrados en R	52
Figura 19: Conexion a base de datos creada "dataset"	52
Figura 20: Configuracion en R para conexion a POSTGRESQL	53
Figura 21: Test conexion a POSTGRESQL	53
Figura 22:Codigo a multiples conexiones en R	54
Figura 23:Codigo para creacion de nueva BD "dataset" e integracion de tablas	55
Figura 24: Esquema mostrando nueva BD "DataSet"	55
Figura 25:Codigo para generacion de datos del "dataset".....	56
Figura 26:Codigo para la transformacion de datos y mostrado en R	56
Figura 27: Dataset creada y lista para el test de entramiento y el test de prueba	57
Figura 28: Instalando libreria para los algoritmos clasificafores	58
Figura 29: Instalando libreria para los algoritmos clasificadores.....	58
Figura 30: Aplicando el algoritmo arbol de decision.....	59
Figura 31: Generando grafico en algoritmo arbol de clasificación	59
Figura 32: Instalando libreria para clasificador red bayesiana.....	60
Figura 33: Resultados aplicando clasificador red bayesiana.....	60
Figura 34: Arquitectura de la solucion web	62
Figura 35: Capaz de estructura y tecnologia web a construir	63
Figura 36: Primeras pruebas de grafica web.....	64
Figura 37: Grafica prueba arbol de decisiones.....	65
Figura 38: Territorio - ubicacion PEAM.....	70
Figura 39: Base de datos PEAM.....	70
Figura 40: Solicitud para inicio del trabajo de investigacion.....	71
Figura 41: Estructura organica PEAM.....	72



INDICE DE TABLAS

Tabla 1: Numero de población.....	28
Tabla 2: Operacionalizacion Independiente.....	30
Tabla 3: Operacionalizacion Dependiente.....	30
Tabla 4: Planteamiento de soluciones:.....	32
Tabla 5: Criterios de rigor cientificos.....	33
Tabla 6: Indice de error de algoritmos evaluados.....	34
Tabla 7: Tiempo de entrenamiento para cada algoritmo evaluado.....	35
Tabla 8: Mínimo de meses para generar entrenamiento y pronostico.....	37
Tabla 9: Eficiencia del modelo.....	38
Tabla 10: Tiempo de ejecucion de modelo.....	39
Tabla 11: Esquematizacion por fases y sub procedimientos.....	42
Tabla 12: Herramientas a utilizar.....	43
Tabla 13: Metodologia CRISP-DM SEMMA.....	44
Tabla 14: Tipos de algoritmos.....	44
Tabla 15: Herramientas a utilizar.....	44
Tabla 16: Base de datos utilizadas.....	47
Tabla 17: Esquemas de base de datos.....	48
Tabla 18: Historial de usuario.....	62

RESUMEN

El presente trabajo de investigación tiene como objetivo principal, la aplicación de la minería de datos analizando de manera comparativa las técnicas de clasificación, para ello se debe determinar el algoritmo que mejor se adecua a los datos estadísticos sobre inseguridad ciudadana, violencia y delitos en el país que maneja el Observatorio Nacional del Ministerio del Interior.

En la actualidad, el análisis de información, con el fin de extraer conocimiento que permita otorgar ventajas competitivas para mitigar los riesgos en una empresa u organización se ha vuelto muy importante. Los sistemas que usualmente se usan para la gestión de información en una empresa son los que, en una escala piramidal, se encuentran en un nivel superior a los sistemas transaccionales, ya sean conocidos como los sistemas de toma de decisiones, sistemas de información gerencial, sistemas de información ejecutivos.

Palabras clave: Minería de datos, Árbol de decisiones, Red bayesiana, R Project, Clasificadores, Toma de decisiones, clasificación.

ABSTRACT

The main objective of this research work is the application of data mining by comparing the classification techniques in a comparative way, for this the algorithm that best suits the statistical data on citizen security, violence and crime in the country that manages the National Observatory of the National Ministry of Interior.

At present, the analysis of information, in order to extract knowledge that allows to gain competitive advantages to mitigate the risks in a company or organization, has become very important. The systems that are usually used for information management in a company are those that, on a pyramid scale, are at a higher level than the transactional systems, and are known as decision-making systems, management information systems, executive information systems.

Key Words: Data mining, Decioniontree, Bayesiannetwork, R Project, Classifiers, Decisionmaking, Classification



INTRODUCCIÓN

La minería de datos es una herramienta que usa técnicas matemático – computacionales para determinar la detección de patrones sobre la exploración de bases de datos que permitan establecer las probabilidades para predecir comportamientos de un determinado fenómeno.

La minería de datos es útil en las áreas de recursos humanos para poder seleccionar a sus trabajadores con mejor rendimiento laboral. Los datos recopilados ayudan a contratar trabajadores.

En el mundo algunos elementos del RRHH en que puede ser usada la minería de datos son: Evaluación de competencias de los empleados, las metas y objetivos anuales de empleados, análisis de datos de mayor productividad, definir grupos diferenciados de empleados.

Dentro de la minería de datos existen diversas técnicas aplicables a la realidad de negocio según el fenómeno a analizar, desde análisis de pronósticos, a probabilidades se utilizan técnicas como regresión, series temporales, clasificación, asociación entre otras.

Uno de los problemas que se presentan al utilizar técnicas de minería de datos es determinar el algoritmo a utilizar para la fase de entrenamiento del modelo, que es una fase importante en el cual se explota los datos históricos con el fin de hallar los patrones de comportamiento.



CAPÍTULO I: PROBLEMA DE INVESTIGACIÓN

1.1 Situación problemática

En la actualidad, el análisis de información, con el fin de extraer conocimiento que permita otorgar ventajas competitivas para mitigar los riesgos en una empresa u organización se ha vuelto muy importante. Los sistemas que usualmente se usan para la gestión de información en una empresa son los que, en una escala piramidal, se encuentran en un nivel superior a los sistemas transaccionales, ya sean conocidos como los sistemas de toma de decisiones, sistemas de información gerencial, sistemas de información ejecutivos etc.

Sin embargo, estas herramientas permiten medir con indicadores, valores consolidados de los reportes de extracción de los sistemas transaccionales, una tendencia reciente es utilizar herramientas especializadas a partir de un análisis avanzado, que comprende el uso de complejos algoritmos computacionales, basados en estadística y matemática, que explotan la información de los sistemas transaccionales con el fin de identificar patrones de comportamiento en los datos históricos, patrones que definen reglas con lo cual se comprende un determinado fenómeno, a lo narrado anteriormente se le denomina Minería de datos.

La minería de datos es una herramienta que usa técnicas matemático – computacionales para determinar la detección de patrones sobre la exploración de bases de datos que permitan establecer las probabilidades para predecir comportamientos de un determinado fenómeno.

La minería de datos es útil en las áreas de recursos humanos para poder seleccionar a sus trabajadores con mejor rendimiento laboral. Los datos recopilados ayudan a contratar trabajadores.

En el mundo algunos elementos del RRHH en que puede ser usada la minería de datos son: Evaluación de competencias de los empleados, las metas y objetivos anuales de empleados, análisis de datos de mayor productividad, definir grupos diferenciados de empleados.

En la minería de datos existen varias técnicas aplicables a la realidad de negocio según el fenómeno a analizar, desde análisis de pronósticos, a probabilidades se utilizan técnicas como regresión, series temporales, clasificación, asociación entre otras.



Uno de los problemas que se presentan al utilizar técnicas de minería de datos es determinar el algoritmo a utilizar para la fase de entrenamiento del modelo, que es una fase importante en el cual se explota los datos históricos con el fin de hallar los patrones de comportamiento. Es necesario determinar esto, ya que en el escenario actual se plantea utilizar minería de datos para analizar el histórico generado en la gestión de recursos humanos del PROYECTO ESPECIAL ALTOMAYO.

El PEAM actualmente es un ente descentralizado que forma parte del Gobierno Regional de San Martín, por lo tanto ejecuta muy dependientemente del Gobierno regional de San Martín, dentro del organigrama considera varias oficinas y cuyo personal dependiendo de los mecanismos y características de las tareas encomendadas y según el rubro de la oficina, sus actividades y sus permisos varían, en la investigación se han considerado aspecto como los permisos que tiene los trabajadores: Permisos Internos y Permisos Externos.

Los permisos Internos son aquellos que el colaborador realiza por motivos laborales y permisos externos son aquellos que el colaborador realiza ajeno a la empresa.

1.2 Formulación del problema

¿Qué técnica de minería de datos permitirá una mejor exploración de análisis de datos en la gestión de recursos humanos del proyecto Altomayo?

1.3 Delimitación de la investigación

La presente investigación se limita en el ámbito del PROYECTO ESPECIAL ALTOMAYO en la ciudad de Moyobamba, provincia de San Martín.

1.4 Justificación e importancia de la Investigación

Científica:

La investigación se centra en realizar un análisis utilizando herramientas especializadas a partir de un análisis avanzado, que comprende el uso de complejos algoritmos computacionales, basados en estadística y matemática, que explotan la información de los sistemas transaccionales con la finalidad de determinar las características de comportamiento



en datos históricos, patrones que definen reglas con lo cual se comprende un determinado fenómeno, a lo narrado anteriormente se le denomina Minería de datos.

Institucional

Permitir a mejorar el nivel de atención en los distintos servicios profundizando en factores que no sean perceptibles en el de los datos de recursos humanos de la empresa Altomayo.

Social

En el ámbito social se busca resolver mediante la comparación de algoritmos computacionales, matemáticos, cuya finalidad es obtener patrones o reglas en el fenómeno a analizar, y brindar un mejor servicio adecuado a sus necesidades.

1.5 Limitaciones de la investigación

Los datos recopilada muestra deficiencias en su estructura documentaria (procesos).

1.6 Objetivos de la investigación

1.6.1. Objetivo general

Identificar las técnicas de minería de datos para mejorar el análisis de la gestión de recursos humanos del PROYECTO ESPECIAL ALTOMAYO

1.6.2 Objetivos específicos

- a) Examinar la base de datos de recursos humanos de la entidad
- b) Distinguir las técnicas que se ajusten a los hechos reales de la información obtenida.
- c) Planificar e implementar un aplicativo que permita dar respuestas más rápido en el análisis de gestión de recurso humanos utilizando técnicas de minería de datos
- d) Realizar un comparativo de los algoritmos de clasificación del modelo diseñado

CAPÍTULO II: MARCO TEORICO

2.1. Antecedentes de estudio

2.1.1. INTERNACIONAL.

TESIS “*DATA MINING: A TOOL FOR KNOWLEDGE MANAGEMENT IN HUMAN RESOURCE*”

(Sadath, 2013) nos dice que para ser una empresa competitiva hay que mantener y ganar reputación en su respectivo mercado o industria. Gestión de Recursos Humanos (HRM) juega un papel principal en la determinación de esta competitividad y la eficacia para una mejor supervivencia. La Gestión de recursos humanos generalmente muestra las políticas, prácticas y sistemas que influyen en el comportamiento de los empleados, la actitud y el rendimiento.

Las empresas consideran HRM como "personas prácticas". Por lo tanto, se convierte en la responsabilidad de la gestión de recursos humanos a la mía a los mejores talentos de la derecha

tiempo, entrenarlos, observar su rendimiento, y recompensarlos en última instancia, mantenerlos felices en una empresa. Es simplemente debido a la razón de que todas las estrategias de una organización son directa o indirectamente relacionadas con el talento de la misma. Para obtener y mantener una ventaja competitiva, la gestión del conocimiento (en desarrollo, compartir y aplicar el conocimiento) dentro de la organización se convierte en esencial. Pero entonces, ¿cómo está conectado a HRM Conocimiento Management (KM) se convierte en una cuestión muy relevante. Cuando los empleados se evalúan de su actuación, diferentes métodos se pueden utilizar para la minería del mejor conocimiento de ellos. El presente trabajo es un intento de estudiar y comprender el potencial de la minería de datos (DM) técnicas para decisiones inteligentes automatizadas a partir de la base de datos de los empleados para las predicciones de los empleados y el rendimiento de la aplicación de las mejores estrategias de gestión, por lo tanto el logro de un sistema de recursos humanos estable y brillante negocio.

TESIS “*APLICACIÓN DE TECNICAS DE MINERIA DE DATOS PARA MEJORAR EL PROCESO DE CONTROL DE GESTION EN ENTEL*”

Esta tesis tiene como prioridad disminuir el tiempo de identificación de las guías de servicios ENTEL, para ello se realizó un diseño multidimensional, técnicas de minería de datos y simplificación de procesos, y así



tener como resultado la información más oportunamente. Con el diseño de minería de datos se diseñaron otras métricas para identificar y procesar a los clientes y servicios, para obtener valores más reales de ingresos perdidos en cada segmento, y adaptar diferentes estrategias para hacer el cobro a sus clientes. (CLEMENTE ANTONIO MARTÍNEZ ÁLVAREZ, 2012)

TESIS “APLICACIÓN DE MINERÍA DE DATOS PARA LA EXPLORACION Y DETECCION DE PATRONES DELICTIVOS EN ARGENTINA” El autor tiene como objetivo implementar minería de datos para distinguir la información de los criminal en Argentina y verificar su eficiencia y valor agregado. Por lo tanto se necesita la identificación y detección de tipos de homicidios dolosos hechos en Argentina en el periodo del 2005. (Ignacio Perversi, 2007)

La minería de datos se basa sobre grandes volúmenes de información que provienen de distintos sistemas, con el fin de utilizar los mismos para extraer conocimiento en beneficio de un determinado objetivo, utilizando adecuadas técnicas. (Herman Jose, Ramirez Jose, Ferri Cesar, 2004)

2.1.2 NACIONALES:

TESIS “MINERÍA DE DATOS APLICADOS A LAS VENTAS CON TARJETA DE CRÉDITO CLÁSICA REALIZADOS EN LAS TIENDAS SAGA FALABELLA EN LA CIUDAD DE LIMA” el desarrollo de esta tesis tiene como objetivo identificar los patrones que se van presentando y la naturaleza de los datos adquiridos los pedidos con tarjeta de crédito de la ciudad de Lima.

Con el objetivo de distinguir hábitos de clientes para satisfacer su demanda, aumentar la eficiencia administrativa de los creadores de productos que se relacionan a trato de pedidos y mejorar la rentabilidad de pedidos realizadas con tarjeta de crédito en las tiendas (Hoher Willy Siccha Vega, 2012)

TESIS “TÉCNICAS DE MINERÍA DE DATOS APLICADAS A LA CONSTRUCCIÓN DE MODELOS DE SCORE CREDITICIO” (Ramírez A., 2007) Menciona que la diferencia entre técnicas como las redes neuronales, máquinas vectoriales de soporte, árboles de decisiones, análisis discriminante y regresión logística; son las utilizadas de score crediticio que permite identificar características con las características de los clientes, concluyendo que en el margen del uso de técnicas habitual, la Regresión Lineal es la más precisa en la

predicción de la muestra. Pero las Redes Neuronales dan más presión en cada una de las muestras que los creadores realizan.

2.1.3 LOCAL

TESIS “IMPLEMENTACIÓN DE UNA APLICACIÓN WEB UTILIZANDO MINERÍA DE DATOS PARA MEJORAR LA GESTIÓN DE FACTURACIÓN EN LA EMPRESA PEXPORT S.A.C”

La tesis planteada, parte del problema de facturación del consumo eléctrico por debajo del real o en exceso, generado por la toma de lectura en campo errada por parte de la empresa PEXPORT S.A.C. y el parámetro de consumo genérico manejado por la concesionaria; ante esto el objetivo principal es implementar un sistema web, utilizando técnicas de minería de datos, contribuya a mejorar la gestión del proceso de facturación de la empresa PEXPORT S.A.C. reflejándose en una mejora de sus resultados económicos y operativos, por ende, esto repercute, no solo en mejorar la facturación del servicio de energía eléctrica a los clientes finales de Electronorte S.A., sino también, en el confort de los colaboradores de la empresa, automatizando procesos muy importantes para garantizar calidad en el servicio.

En relación a esta investigación, este trabajo brinda información útil con la metodología de uso de modelos de minería de datos, como es el caso de CRISP-DM (Díaz, Jorge Alfredo Salas, 2013)

2.2. Estado del arte

Hoy en día, la conducción de información es fundamental en toda entidad, especialmente aun para sustentar procesos del negocio que manejan recursos de información, aumentando la labor de aclarar el ingreso a los datos (**Logreira, 2011**)

En los últimos años ha cobrado un gran interés y avance, mediante el avance de técnicas que utiliza para el análisis de los datos; convirtiéndose en un procedimiento global de búsqueda de conocimiento potencialmente útil utilizando un gran volumen de datos.

La minería de datos es la consecuencia de una evolución, el término se introdujo en los años 1990 a raíz de tecnologías de información, algoritmos y avance de software. El desarrollo de la minería de datos sea originado en tres conceptos: “la estadística”, ya que sin ella, no se



podría utilizar la minería de datos; “inteligencia artificial” y, finalmente, el “aprendizaje automático”, que se describiría como el resultado de estadísticas e inteligencia artificial.(Aspiazu, 2010).

La mejor metodología en proyectos de minería de datos es el modelado de procesos jerárquicos consistentes las metodologías CRISP – DM y SEMMA. Ya permiten estructurar el proyecto de Datamining en áreas detalladas en cuatro niveles de abstracción. (Daedalus, 2002).

La minería de datos utiliza un grupo de técnicas, que dan como resultado un modelo sobre el conjunto de datos y una enumeración de patrones sobre la información dada. Dentro de los ejemplos importantes dentro de la minería, son las reglas de asociación, que permiten, encontrar relaciones no visibles entre elementos, y de esta manera plantear reglas que ayuden a perfeccionar la toma de decisiones. (Mannila, 2002).

La minería de datos, es imprescindible disminuir el número de datos, obteniendo la data que solo se necesita, para disminuir el rendimiento computacional y humano. La demás data será redundante, esto se basa en un axioma fundamental: “la hipótesis más simple”. (Marcano Aular & Talavera Pereira, 2007)

Hoy en día, Con la minería de datos se trabaja en muchos problemas, especialmente aquellos que están relacionadas con el medio ambiente, el cambio climático, la salud. Incluso, existen investigaciones recientes desarrolladas para analizar la parte genética de la sangre para detectar el VIH. (Poncelet, 2014).

En la minería de datos, tomamos y procesamos los documentos con finalidad que aflore la hipótesis indicada. Lo apropiado sería que los datos indiquen o describan el por qué de esa configuración y comportamiento.

Con datamining las entidades tienen una manera de visualizar sus información, esperar mejoras con el resultado de toda la diversidad de obstáculos como: Planeación económica, inteligencia empresarial, finanzas, análisis de mercados y análisis de perfiles de clientes, problemas ambientales o de la salud. Bien planteada, se transforma en estrategia que aumentará la escala de competitividad. En los negocios La toma de decisiones indicada se determina según la velocidad con que se identifica y procesa los datos valiosos



2.3. Base teórica científica

Solución Web / Analítica

2.3.1. Tecnología Web

2.3.1.1. Lenguajes de Programación

a. HTML5

Según (Gauchat, 2012) es una comunicación estática que ejecuta programas de sitios web (acrónimo en inglés HyperTextMarkupLanguage, en español Lenguaje de Marcas Hipertextuales).

Ventajas:

- a) Fácil va permite describir hipertexto
- b) Texto mostrado como procedimiento y agradable.
- c) Lenguaje fácil aprendizaje.

Desventajas:

- a) Lenguaje estático.
- b) La traducción de los navegadores son diferente.
- c) El diseño demora.
- d) Las etiquetas son limitadas.

b. Javascript

(Maza, 2012) JavaScript es un lenguaje para las aplicaciones cliente/servidor mediante internet. Se diseñó como un lenguaje para realizar los scripts que pudieran desarrollarse en archivos HTML y dar dinamismo a las páginas web.

Ventajas:

- a) Plataforma segura y confiable
- b) Scripts el tamaño es limitada por motivos de seguridad
- c) El código se lleva a cabo con el cliente

Desventajas:

- a) Codificación visualizado por los usuarios
- b) La codificación se puede descargarse



c. PHP

Lenguaje de script que se realiza del lado del servidor. El producto de este procedimiento se encaja en la página HTML, que se remite al navegador. La técnica permitirá hacer páginas Web dinámicas con contenido que puede resultar total o parcialmente en el instante de la llamada de la página, según (Heurtel, 2011).

Ventajas:

- a) Rápido de entender.
- b) veloz.
- c) Se ejecuta en varias plataformas: Linux, Windows, entre otros.
- d) Tiene facultad de conexión con los manejadores de base de datos : MySQL, PostgreSQL, Oracle, MS SQL Server, entre otras.

Desventajas:

- a) Es imprescindible instalar un servidor
- b) Es lento cuando tiene muchas solicitudes de peticiones
- c) La originalidad de la programación puede visualizarse ya sea en HTML y PHP.

2.3.1.2. Base de Datos

a. PostgreSQL

(Eslava, 2013) Es el motor de bases de datos relacionales (RDBMS) que comprueba la originalidad referencial con la utilidad como base de datos, es un poco lenta a otros motores.

Tiene eficaz escalabilidad y versatilidad, sobresale por la rigidez que ofrece. Tiene una dilatada trayectoria de avance comunitario y su fama es inmensa.

Características

- a) Llaves ajenas (foreignkeys)
- b) Vistas (views)
- c) Reglas (Rules)
- d) Funciones/procedimientos almacenados
- e) Numerosos tipos de datos, posibilidades de definir nuevos tipos
- f) Puede guardar audio y videos
- g) Herencia de tablas (Inheritance)



- h) Tablespaces
- i) Replicación asíncrona

c. SQL Server

2.3.1.3. Servicios Web

Es la unión de tecnologías con facultad de interactuar en la Web, con la finalidad de dar servicios

Utiliza la Web como plataforma, los usuarios pueden solicitar un servicio que algún proveedor tenga en la red. Para esto deben existir unos procesos de comunicación estándares entre diferentes aplicaciones. Estos mecanismos tienen que interactuar para dar la información de forma dinámica al usuario. (Mills, 2006)

A. Metodologías de desarrollo de Software

a. Extreme Programming

Formulada por kentbecka programación extrema y la acogida de los nuevos métodos de desarrollo como pretende hacer en el proyecto, y aplicarlo de forma dinámica mientras el ciclo de vida del software.

Características

Avance interactivo e incremental: cortos progreso, consecutivas

Test individuales seguidas: Frecuentes repetidas y automatizadas, incluyendo pruebas de regresión.

Las labores de desarrollo se realizan por individuos del mismo contexto laboral.

Continuamente la unión del personal de desarrollo con el proveedor o el cliente

Rectificación de todas la fallas antes de implementar nuevas funciones.

Propiedades del código compartida: consiste en que todo el equipo de trabajo puede corregir y extender en cualquier parte del aplicativo

Simplicidad en el código La programación extrema es más sencillo hacer algo simple, que realizar algo complicado y quizás nunca utilizarlo.



b. SCRUM

Es una plataforma de desarrollo rápido determinado por:

Con una habilidad de avance, en el sitio de planificación y ejecución finalizada del producto.

Prioriza el control del resultado para que interactue con las personas en grupos auto-organizados.

Roles Principales

ProductOwner: se encarga que el SCRUM trabaje de manera optima relacionado al negocio o historias del cliente, y las coloca en el ProductBacklog.

ScrumMaster (o Facilitador): tiene como prioridad quitar los obstáculos que pueden dificultar que el área de trabajo no llegue a su meta trazada ya que se pueden auto-organizar

Equipo de desarrollo: consiste en la entrega del producto. Esta compuesto por 3 a 9 personas con diferentes destreza para finalizar el trabajo

c. Minería de Datos

La minería de datos o exploración de datos (es la etapa de análisis de "KnowledgeDiscovery in Databases" o KDD) es una parte de la ciencia computacional que esta referida a la exportación de patron en volumen abundante.

La labor primordial es analizar automáticamente o semi-automatico volúmenes masivos de información con el fin de separar rasgos importantes.

Estos rasgos pueden ser hasta desconocidos su importancia por el usuario, se puede obtener resultado con mas precisión

c.1 Proceso de Minería de datos

Consiste en las siguientes etapas:

Selección de grupo de datos: esta etapa consiste en seleccionar los datos que queremos analizarlos y preparar con la finalidad de emplear la técnica de minería de datos esta etapa tiene también como nombre pre-procesamiento de datos.



Procesos de minería de datos

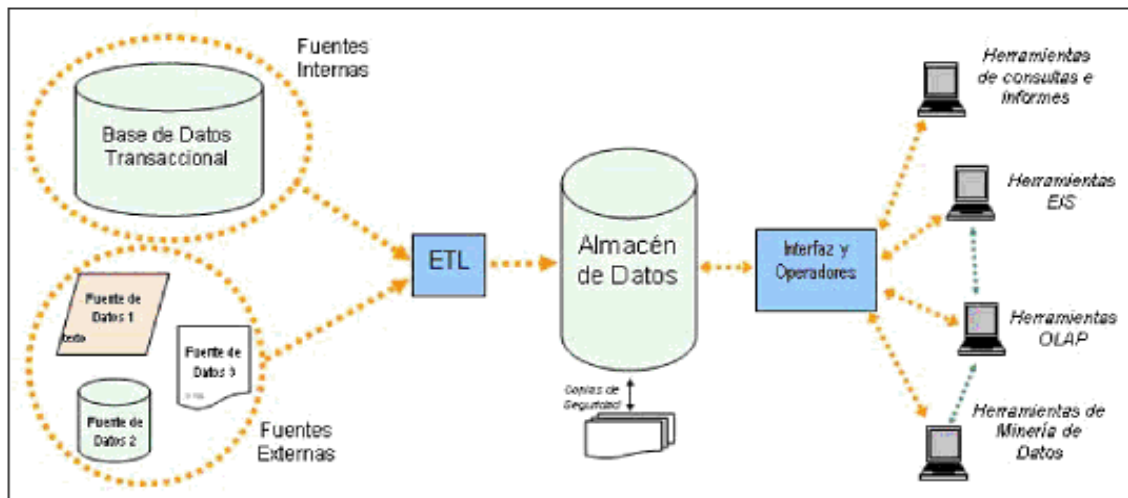


Figura 1: Proceso de Minería de Datos

Fuente: Cesar Perez Lopez(2007) "Minería de Datos, tecnicas y herramientas" Pag 3

c.2 Técnicas de Minería de Datos (Martínez, 2012)

Se diferencia técnicas predictivas, donde la variable son dependientes e independientes.

Técnica descriptivas las variables inicialmente tienen el mismo estatus

Clasificación de técnicas de minería de datos



Figura 2: Clasificación de las Técnicas de Minería de Datos

Fuente: Cesar Perez Lopez(2007) "Minería de Datos: Tecnicas y Herramientas" Pag 5



Análisis de Clúster

El análisis de clústeres (o clustering), con el objetivo de dividir un grupo de información en pequeños grupos o clusters, teniendo en cuenta que tengan como objetivo los mismo rasgos similares cada cluster identifica un grupo de objetos con propiedades o atributos

Etapas:

- a) Selección o extracción de atributos: Se toma las propiedades mas iniciala los que se desea realizar clustering.
- b) Selección del algoritmo: teniendo en cuenta el punto de vista de igualdad conveniente
- c) Validación de los clústeres: en base al algoritmo los parámetros que se emplean, se obtienen la porción y organcion de clusteres diferentes
- d) Interpretación de los resultados: se realiza con los conocedores del producto para ser entregados a los clientes convencionales con vistas que sea entendible

Reglas de Asociación

Instrumento descriptivo, su finalidad principal es identificar el vinculo elocuente que hay en los datos, teniendo la posibilidad que hay entre objetos.

Clasificación

Permite encontrar un elemento de un sector de datos, de una condición predifina. Ósea indica a que conjunto pertenece. Identifica las ragos parecido que este relación a una porción de datos. Puede aprovecharse para ver el comportamiento de nuevas peticiones

Regresión

Le da un valor real a los elementos a travez de valores existentes para predicir datos futuros. La regresión utiliza técnicas estadísticas siendo la regresión lineal, algunos obstáculos en la actualidad no sirven con proyecciones recto

c.3 Modelos y algoritmos de minería de datos

c.3.1. K- means: Es un algoritmo de clustering particional, mas reconocidos y empleados cuando las variables con de característica cuantitativo.trabaja iterativamente

Árboles de decisión

Son los procedimientos inductivos de enseñanza revisados, en los que se hace fracción consecutiva de un grupo de datos, a través de la posición de selección, estando bien establecidas de forma jerárquica, con la finalidad de aumentar el espacio entre las clases de datos obtenidos en cada uno de los intercambios.

Es una manera de presentar un conjunto de reglas que dirigen hacia una clase de datos, son utilizados para estudiar y ejercer predicciones.

Poseen la siguiente estructura:

Nodos: sirven para determinar los atributos de un grupo de datos. El nodo principal determina a los datos

Ramas: son las que separan uno de otros, y las que limitan las variables

Hojas: son grupos de información obtenidos por la fracción que realiza el algoritmo.

Arbol de decision para evaluar el riesgo de un cliente



Figura 3: Arbol de decisiones para evaluar el riesgo de un cliente

Fuente: Clemente Antonio Martinez alvarez (2012) "Aplicación de técnicas de minería de datos para mejorar el proceso de control de gestión de ENTEL" Pag 48



Redes Neuronales Artificiales

Son los sistemas de proceso de datos. identifica voz, datos, ya que podemos configurar diferentes comportamientos entrenando o programando puede realizar operaciones matemáticas, están enlazadas entre nodos y emite una señal cuando estos se desconectan.

Red neuronal con pesos asociados a cada nodo

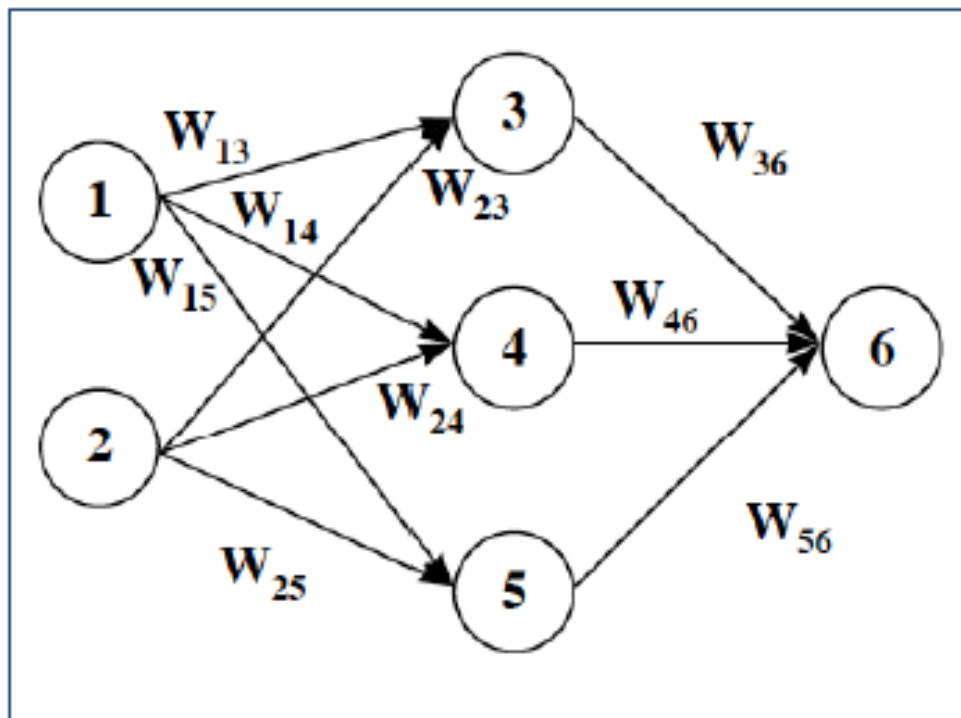


Figura 4: Red Neuronal con pesos asociados a cada nodo

Fuente: Patricio Garcia Baez (2013) "Introduccion a las redes Neuronales y su aplicación a la investigación astro física" Pag 15

Support Vector Machines (SVM)

Es un procedimiento de minería de datos sólido y acertado, es más usado para decifrar obstáculos de selección y regresión. Su finalidad es buscar la sobresaliente función para separar un grupo de datos

La finalidad es encontrar un clasificador lo más aptable, incrementando la facilidad del modelo y desminuyendo los obstáculos que se presenten en los datos obtenidos para el entrenamiento



d. Metodologías para la aplicación de Minería de datos

d.1 CRISP – DM (Cross Industry Standard Process for Data Mining)

CRISP-ordena la implementación de un plan de Data Mining en un conjunto de pasos, con labor complejas e individuales que llevaran a dar con la finalidad cumplir con el objetivo. Estas etapas trabajan de forma cíclica e iteraria, pudiendo retorna de cualquiera etapa.

Fases de modelo CRISP - DM

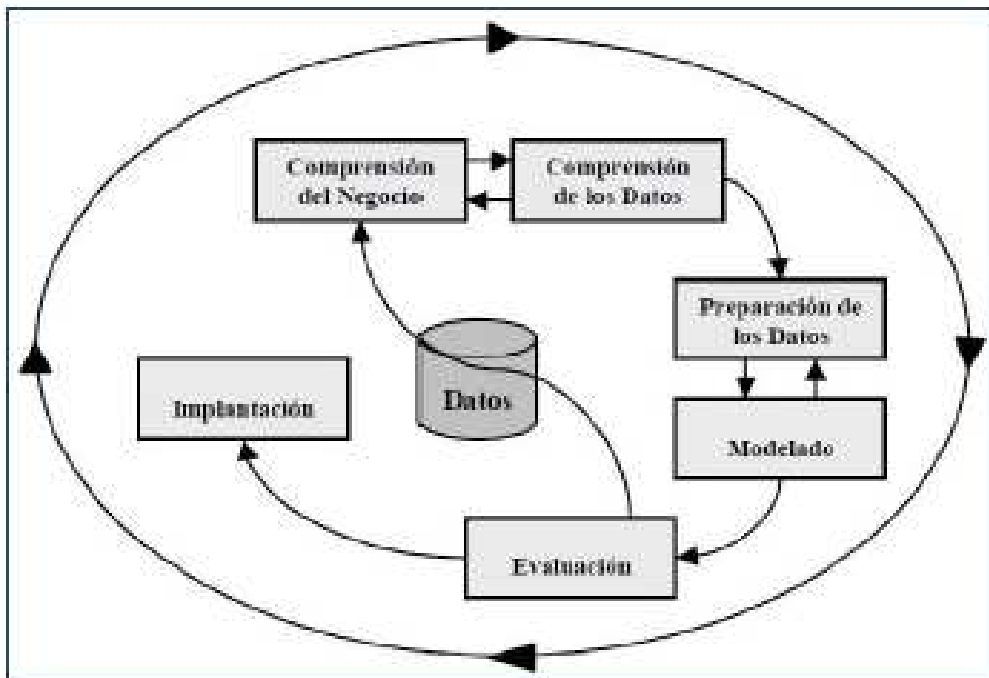


Figura 5: Fases de modelo CRISP – DM

Fuente: The CRISP - DM consortium (2013) “CRISP – DM 1.0 Step by Step Data Mining guide” Pag 12

Comprensión del negocio: reconocer el comportamiento del estudio antecedentes a utilizar.

Preparación de los datos: se identifican la información a utilizar, después siguen por la etapa de limpia estructuración, integración y formatos

Modelamiento y evaluación: Se indetifica la técnica que mas se adapte, armando el modelo, para después someterse a multiples test y apreciación

Despliegue del proyecto: Implantación del modelo y seguimiento de su funcionamiento.



d.2 SEMMA (Kuna, 2014)

Su nombre se debe a los fases metodología semeja para los pasos de investigación de datos, son: muestreo (sample) exploración (explore) modificación (modify) modelado (model) valoración (assess)

SEMMA es creada para emplearla en la herramienta de Data Mining, es usada principalmente para los que utilizan herramienta de software y aplicación es reducido.

Esta metodología es usada fundamentalmente por los que usan esta herramienta software y su aplicación es limitada.

2.4. Definición de la terminología

2.4.1. CRISP – DM: Cross Industry Standard Process for Data Mining. Es un tipo de pauta de desarrollo de minera de datos que plantea rasgos común que adaptan las personas conocedoras en minería de datos

2.4.2. Minería de datos: es una fase que busca encontrar rasgos importantes de considerables grupos de información. Mediante los procedimientos de la inteligencia artificial, aprendizaje automático, estadístico y sistema de base de datos

2.4.3. Semma: Muestra, Explorar, Modificar, Modelo y Evaluar. Consiste en una cinta de procedimientos creadas por SAS institute inc. Uno de los grandes creadores de estadística y de inteligencia de negocios de software.

2.4.4. Tecnología web: un grupo de mejoras y encargo que van a acceder crear y juntar tareas de formas inteligente mediante la creación de un sitio web, blog, aplicativos móvil. etc



CAPITULO III MARCO METODOLOGICO

3.1. Tipo y Diseño de investigación

3.1.1 Tipo de la investigacion

Descriptiva porque cumple el objetivo general mediante la implementación de una metodología con fases.

3.1.2 Diseño de la investigacion

Cuasa Experimental que permitirá el desarrollo de una aplicación en las fases enfocadas y conseguiremos la propuesta mejor y seguidamente serán probadas las metas alcanzadas.

3.2. Población y muestra

3.2.1 Población

Como población se usan los registros generados por los empleados en el PROYECTO ESPECIAL ALTOMAYO

Tabla 1:
Numero de población

Cantidad de trabajadores	
PROYECTO ESPECIAL ALTOMAYO	254
TOTAL	254

Nota: Fuente Propia

3.2.2 Muestra

Al tener información sobre los registros de los 254 trabajadores, en las bases de datos entregadas para la investigación, se procederá a realizar la exploración sobre todos los registros de movimientos generados por lo que no se aplica muestra.

3.3. Hipótesis

Los modelos de clasificación de minería de datos me permitirán realizar un análisis de datos de la gestión de recursos humanos del proyecto Altomayo

3.4. Variable

3.4.1 Variable independiente

Análisis de las técnicas de minerías de datos

Algoritmos Clasificadores (Red bayesiana, Arbol de Decisiones)

3.4.2 Variable dependiente

Mejorar análisis de gestión de recursos humanos.

3.5. Operacionalizacion

Tabla 2:
Operacionalizacion Independiente

Dimensión	Indicador	Pregunta	Ecuación	Tipo	UM	Categoría	Ítem
Estimación	Eficiencia del modelo	Tasa de error	Numero de errores antes del modelo / Numero de errores con modelo	Cuantitativo	%	MALO: 60 a 70 % REGULAR: 70 a 80% BUENO: 80 a 90%	1
Tiempo	Tiempo de ejecución del modelo	Procesamiento	Promedio de tiempo entre todas las pruebas de procesamiento modelo	Cuantitativo	Seg.	MALO: más de 1 minuto REGULAR: 30 seg a 60 seg BUENO: 1 a 30 seg	2

Fuente Propia

Tabla 3:
Operacionalizacion Dependiente

Confiabilidad	Índice de error de algoritmos evaluados
Tiempo	Índice de tiempo de entrenamiento para cada algoritmo
Longitud	Numero de meses mínimo soportado para generar entrenamiento y pronósticos

Fuente propia



3.6. Aborde Metodologico, técnicas e instrumentos de recolección de datos

3.6.1. Aborde Metologico

El método que se usa es el inductivo, pues se parte de una situación específica hacia lo general; es decir, a través de técnicas se logra captar una determinada situación problemática en la empresa, la misma que será descrita y analizada durante el proceso de investigación, para dar como resultado un producto.

3.6.2. Técnicas de recolección de datos

3.6.2.1 La observación científica:

Es básicamente el registro visual sacado de la información de la Empresa, según la programación realizada, obteniendo el reconocimiento general de éstas, tal como su realidad actual, identificación de la situación problemática, objeto de estudio en el proyecto de investigación.

3.6.3. Instrumentos de recolección de datos

3.6.3.1. Entrevistas

Se refiere a la información dada por los individuos que están comprometidos en las áreas en estudio; con el fin de tener toda la información en los procesos identificados como problemáticos.

3.6.3.2. Planteamiento de soluciones

Son las técnicas propias de las metodologías usadas para el desarrollo de la investigación. En este caso, se utilizó la metodología de Ralph Kimball para la construcción del Datamart, la metodología CRISP-DM para la aplicación de técnicas de minería de datos y la creación de los modelos; y la metodología de desarrollo ágil XP, para la construcción del sistema web.



3.7. Procedimientos para la recolección de datos

Extracción de datos de los sistemas de gestión de planilla y personal del PROYECTO ESPECIAL ALTOMAYO

3.8. Análisis estadístico e interpretación de datos

Tabulación de datos.

- a) Tablas y muestrascálculo.
- b) Ordenar las interrogantes.

Análisis de datos

- a) Analizado por señalizador.
- b) Uso de señalizadores de preferencia no central.
- c) Empleo de indicadores de medida de dispersión: Rango, varianza, desviación típica, coeficiente de variación.
- d) Empleo de plataflormar SPSS y MS Excel.

3.9. Principios éticos

Tabla 4:

Planteamiento de soluciones:

Criterios	Características científicas del criterio
Confiabilidad	Se estimaran aproximaciones estadísticas que ayuden a determinar el nivel de consistencia interna de las herramientas de acumulación de información
Validación	Se certifica las herramientas de acumulación de información y la oferta de solución mediante el criterio de personas expertas.

Nota: Fuente Propia



3.10. Criterios de rigor científico

Tabla 5:
Criterios de rigor científicos

Criterios	Características éticas del criterio
Confidencialidad	Se asegura la identificación de la empresa, individuos que esten incluidos como testigos de la investigación
Objetividad	El estudio del caso encontrada se situara en juicio tecnológicos e imparciales.
Originalidad	Se buscaran la similtud de los datos mostrados, con la finalidad de mostrar la exitencia de un copia intelectual.
Veracidad	Los datos mostrados será existente, protegiendo los datos de la empresa
Derechos laborales	La oferta de mejorar priorizara el respeto a las obligaciones del trabajador en la empresa de estudios

Nota: Fuente Propia



CAPITULO IV: ANALISIS E INTERPRETACION DE LOS RESULTADOS

4.1. Resultados en tablas y gráficos de la propuesta

Indice de error de algoritmos evaluados

El índice de error se calcula a partir de la fase de entrenamiento en la cual se entrena el modelo con datos históricos reales contra los obtenidos por el modelo.

Tabla 6:
Indice de error de algoritmos evaluados

Test	Árbol de Decisiones	Redes Bayesianas
1	2.6	2.8
2	2.4	2.7
3	2.56	2.89
4	2.45	2.69
5	2.37	2.45
6	2.48	2.51
7	2.56	2.63
8	2.47	2.51
9	2.89	3.12
10	2.94	2.99
Total	2.572	2.729

Nota: fuente Propia

Como resultado, el promedio del error generado por el algoritmo de árbol de decisiones es 2.5 %, en el caso de las redes bayesianas el error se calcula en 2.72



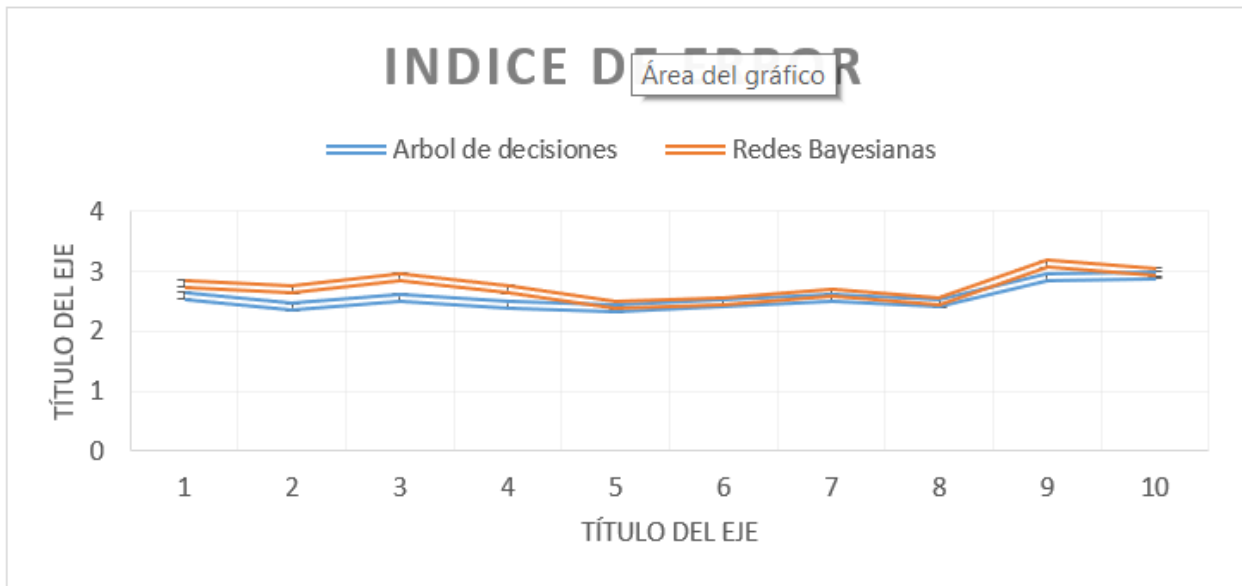


Figura 6: Índice de error entre algoritmos empleados
Fuente: Elaboración propia

Índice de tiempo de entrenamiento para cada algoritmo evaluado

El índice del tiempo de entrenamiento se calcula a partir de la aplicación de la función propia del algoritmo en el script del modelador, el tiempo se obtiene en función al log consultado del entorno R.

Tabla 7:
Tiempo de entrenamiento para cada algoritmo evaluado

Lote de registros	Árbol de decisiones	Redes Bayesianas
100	3	2
500	4	3
1000	6	5
1500	6	6
2000	8	7
5000	12	8
10000	16	10
Total	7.86	5.86

Nota: Fuente Propia



Se obtiene en promedio un tiempo de 7.86 segundos para el árbol de decisiones y un total de 5.86 segundos para las redes bayesianas.

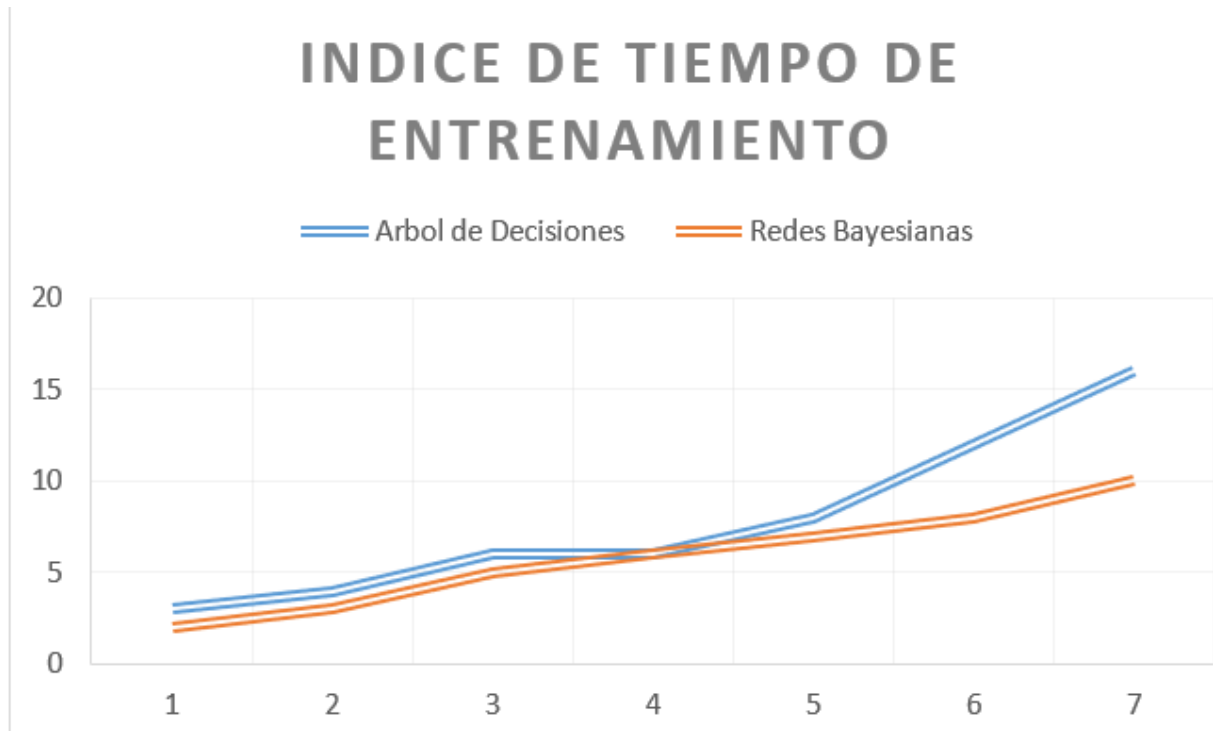


Figura 7: Índice de tiempo de entrenamiento entre algoritmos
Fuente: Elaboración propia

Numero de meses mínimo para generar entrenamiento y pronostico

El número de meses mínimo para generar el entrenamiento está en función al tamaño de la base de datos o cantidad de datos para alimentar en el modelo, esto también está en función a las muestras proporcionales a la cantidad de registros por año.



Tabla 8:
Mínimo de meses para generar entrenamiento y pronostico

Numero de meses	Árbol de decisiones	Redes Bayesianas
12	63%	58%
24	72%	71%
36	86%	82%
48	92%	88%
60	97%	89%
Total	82%	78%

Nota: Fuente Propia

Está claro que mientras menor sea el volumen de registros para entrenamiento, afectara la eficiencia del modelo, El árbol de decisiones logra un 82 % promedio y la red bayesiana obtiene un 78 % como se denota en la tabla presentada y en el grafico a continuación:

EFICIENCIA POR NUMERO DE MESES

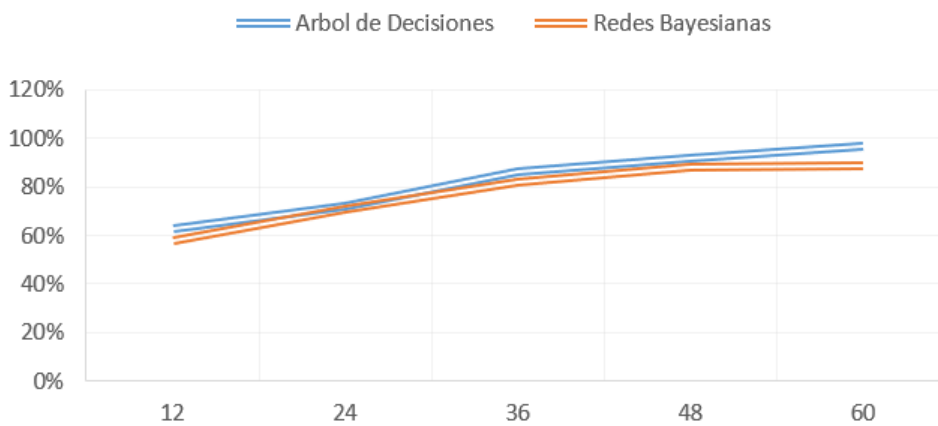


Figura 8: Numero de meses mínimo para generar entrenamiento
Fuente: Elaboracion propia



Eficiencia del modelo

La eficiencia del modelo se calcula de todos los parámetros utilizados para la validación en las pruebas de entrenamiento y testeo, no solo interesa el volumen sino los parámetros propios de cada algoritmo a utilizar.

Tabla 9:
Eficiencia del modelo

Casos de prueba	Árbol de decisiones	Redes Bayesianas
1	84.5	78.9
2	87.6	81.2
3	86.3	79.8
4	85.4	78.7
5	84.6	76.5
6	83.6	78.5
Total	85.33	78.93

Nota: Fuente Propia

EFICIENCIA DE ALGORITMOS

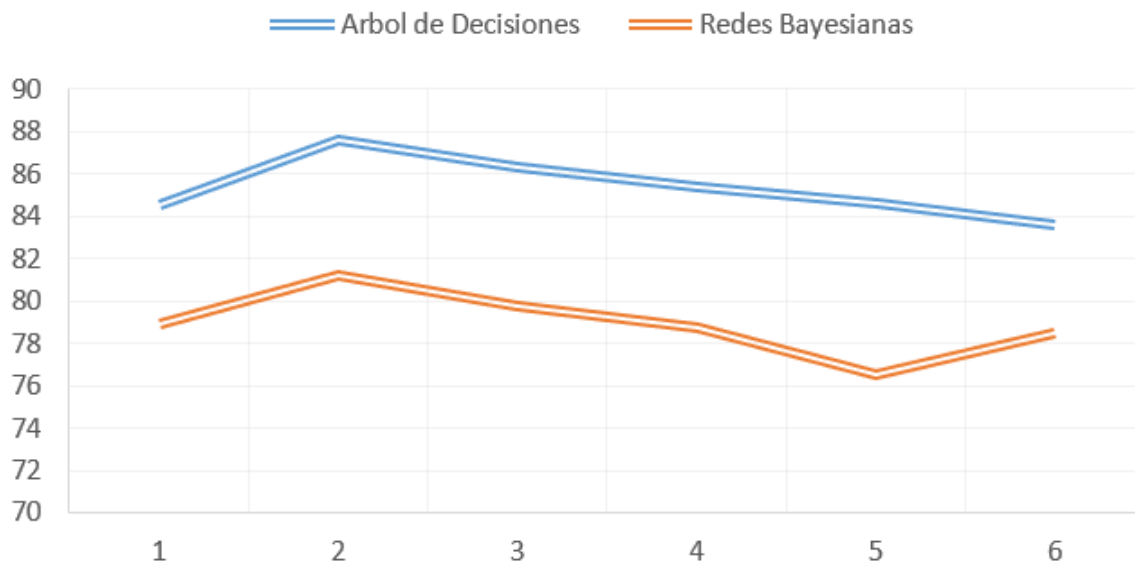


Figura 9: Comparación eficiencia de algoritmos en pruebas
Fuente: Elaboracion Propia



Tiempo de ejecución del modelo

En este punto se calcula el tiempo de ejecución del modelo a partir de la contabilización de tiempo en segundos desde el inicio del proceso hasta la generación de resultados.

Tabla 10:
Tiempo de ejecucion de modelo

Test	Árbol de Decisiones	Redes Bayesianas
1	12	9
2	11	8
3	14	7
4	13	8
5	12	7
6	10	8
Total	12	7.83

Nota: El tiempo es medido en m/s Fuente Propia

TIEMPO EJECUCION MODELO

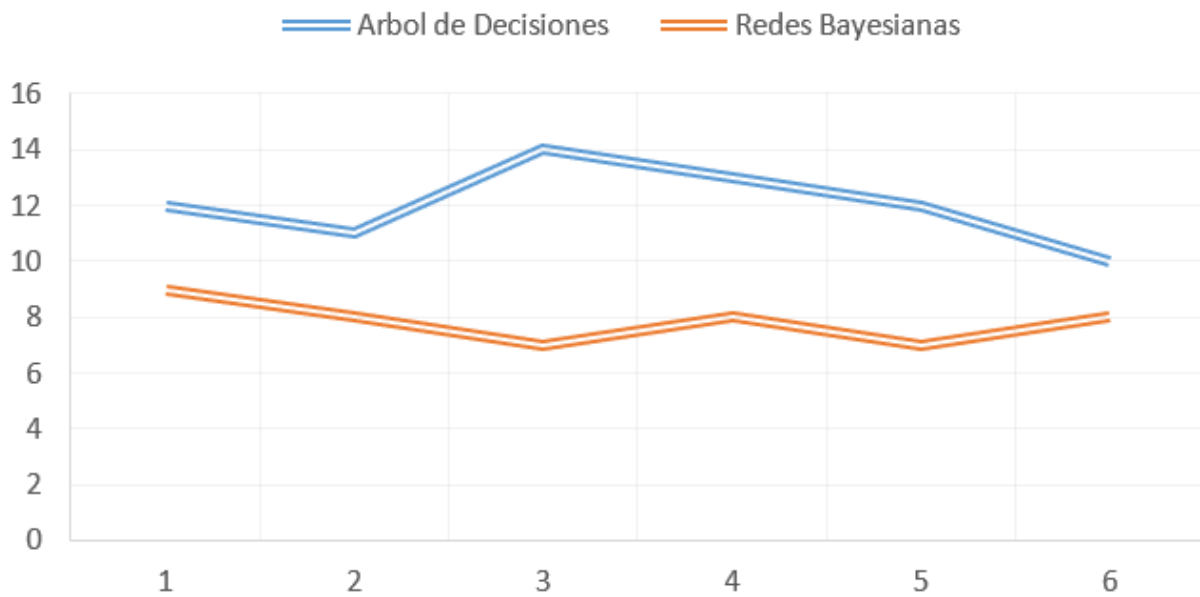


Figura 10: Comparacion tiempo de ejecucion del modelo
Fuente: Elaboracion Propia



4.2. Discusión de resultados

Indice de error de algoritmos evaluados

Si se da error, o margen de error generados por los algoritmos se ha obtenido que las redes bayesianas generan un margen de error mayor al de los árboles de decisiones, las cifras exactas son 2.72 para las redes bayesianas y 2.57 % para los arboles de decisiones, esta brecha a sido probada en escenarios de laboratorio con muestras reducidas de datos, se espera que en la fase de pruebas final se extienda.

Indice de tiempo de entrenamiento para cada algoritmo evaluado

En el caso del entrenamiento las redes bayesianas obtienes 5.86 segundos para generar el análisis de datos, mientras los arboles de decisiones 7.86 segundos, esto se puede deber a la complejidad de ramas que puede tener un árbol de decisiones, que genera más tiempo de procesamiento, aunque este valor puede ser mitigado por el uso de parámetros que definan los cortes a las ramas que se van a analizar.

Numero de meses mínimo para generar entrenamiento y pronostico

En el número de meses mínimo para generar el entrenamiento está clara la tendencia a seguir por el volumen de registros históricos a ingresar, considerar la calidad de datos de los mismos, obteniendo que para periodos de 12 meses los valores son bajos, pero el modelo incrementa eficiencia cuando los datos son con mayor tiempo histórico, como resultado los arboles de decisión obtienen un 82 % mientras que las redes bayesianas un 78%.

Eficiencia del modelo

En la eficiencia del modelo se ha obtenido un 85.33 % para los arboles de decisiones, mientras las redes bayesianas obtienen un 78.93 %, esto se debe a la distribucion de variables categóricas que se tiene, siendo de mejor performance los arboles de decisiones, las redes bayesianas obtienen una mayor performance cuando se trata de variables binomiales.

Tiempo de ejecución del modelo

El tratamiento de datos, aplicación de las reglas obtenidas por el algoritmo de aprendizaje han concluido en la fase de pruebas que los arboles de decisiones obtengan un promedio de 12 segundos contra los 7.83 segundos de las redes bayesianas, esto se debe a la naturaleza de cada algoritmo.

CAPITULO V: PROPUESTA DE LA INVESTIGACION

5.1. Generalidades de la propuesta

La propuesta consiste en desarrollar un modelo analítico que mediante técnicas de clasificación identifique patrones de comportamiento en el análisis de la información (datos) del fenómeno de recursos humanos en una organización del estado.

Identificando los sets de datos a explorar se realizará una construcción de un modelo de minería de datos que usa la metodología CRISP DM para garantizar la correcta elaboración del mismo.

Una vez obtenido el modelo se procederá a evaluar diversos algoritmos clasificadores y su comportamiento en el fenómeno a analizar, para determinar la performance del mismo.

Se diseñará una interfaz web de análisis donde se podrá realizar operaciones de la fase de modelado para el entrenamiento de datos con respecto a la evaluación de los algoritmos clasificadores.

5.2. Esquematización por fases y sub procedimientos

Tabla 11:
Esquematización por fases y sub procedimientos

Fase	Metodología	Sub Procedimiento
Modelado	CRISP DM	- Negocio
		- Datos
		- Selección
		- Transformación
		- Modelado
		- Evaluación
Modulo Web Analítico	XP	- Análisis
		- Diseño
		- Codificación
		- Pruebas

Nota: Esquematización por fases y sus procedimientos



Métricas a implementar

Según la ISO de calidad 9126 estos son los criterios seleccionados que más afectan al desarrollo de la propuesta, según su relevancia.

a) Modulo web Analítico:

Herramientas a Utilizar

En cuanto a las herramientas para tratar el modelo se determinaron las siguientes:

Tabla 12:
Herramientas a utilizar

Herramienta	Licencia	Tipo	Manipulación Algoritmos	Consola	Conexión a BD
R project	Libre	Lenguaje	SI	SI	SI
Rapidminer	Libre	Proceso	NO	Parcial	SI
Microsoft BI	Propietario	Proceso	NO	NO	SI
Excel	Propietario	Hoja de calculo	NO	NO	Limitado
Pentaho	Libre	Proceso	NO	NO	SI
Weka	Libre	Proceso	NO	NO	SI

Nota: Fuente propia

De la tabla se trabaja con R porque es un lenguaje que dara flexibilidad a la hora de realizar el modelado, al igual que permite de manera transparente estudiar los algoritmos en las librerías que aporta la comunidad, en algunos casos se puede modificar el código de algunas de ellas para optimizar los resultados



Tabla 13:
Metodología CRISP-DM SEMMA

Metodologías de Desarrollo de Modelo de Minería de datos	CRISP-DM	SEMMA
Libre elección de las herramientas	2	0
Todas las fases que pueden relacionar	2	0
Procesos de Inteligencia de Negocios	2	0
Comercial – Licencias – Privativa	0	2
Técnicas de ETL	2	2
Módulo de referencia para el usuario	2	0
TOTAL	10	4

Fuente propia

Tabla 14:
Tipos de algoritmos

Algoritmo	Tipo de clase	Tipo de dato	Consola	Conexión a BD
Arboles Clasificación	Polinomial	Real	SI	SI
Redes Bayesianas	Binomial	Boolean	Parcial	SI
Redes Neuronales	Polinomial	Real	NO	SI
Regresión lineal	Polinomial	Real	NO	Limitado
SVM	Polinomial	Numérica	NO	SI

Nota: Fuente propia

Tabla 15:
Herramientas a utilizar

Fase	Herramientas
Modelado	- PostgreSQL
	- R Project
Modulo Web Analítico	- PHP
	- PostgreSQL
	- Sublime Text
	- Zend Server
Plataforma	Windows Server 2012 R2

Nota: fuente Propia



Flujograma de desarrollo

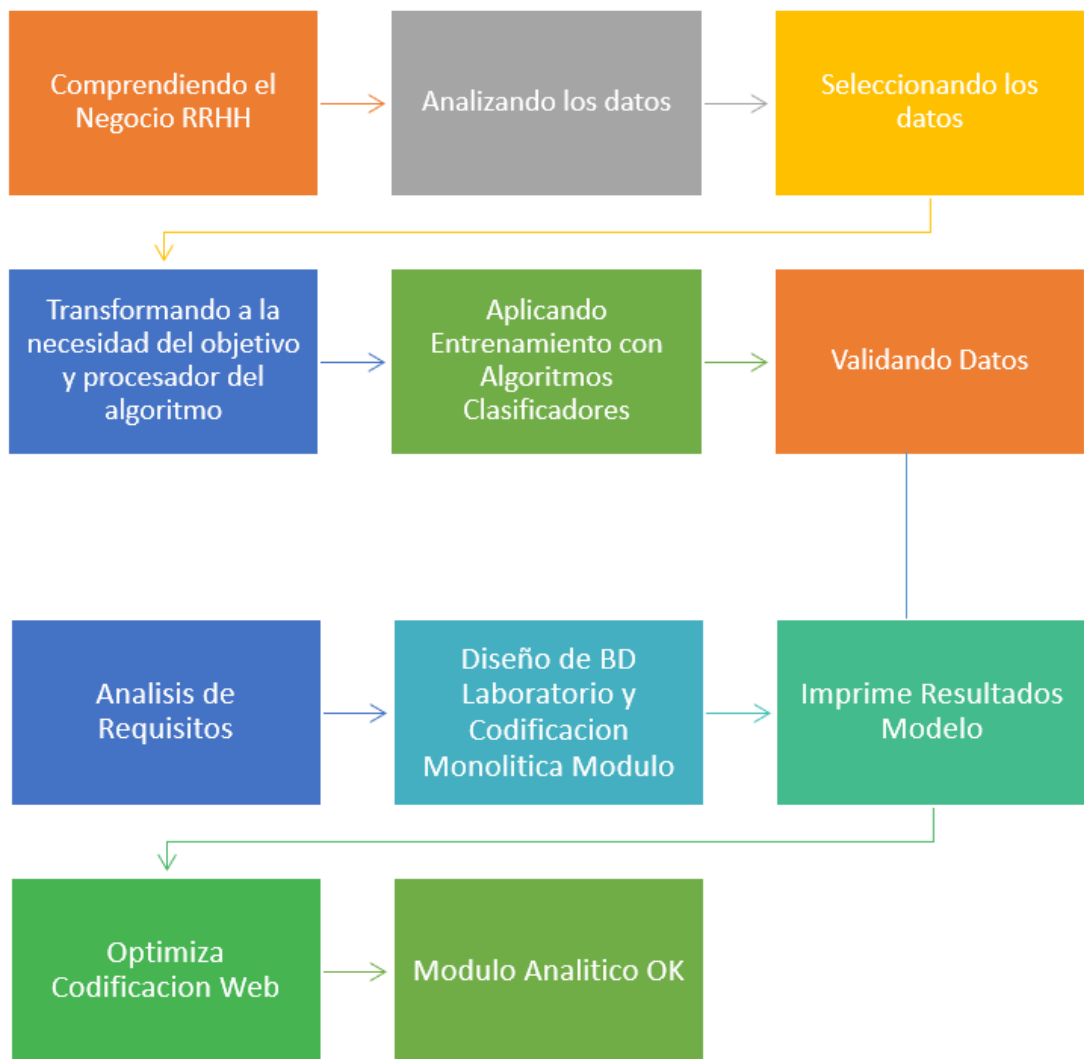


Figura 11: Flujograma de desarrollo de la propuesta
Fuente propia



5.3. Desarrollo de fases

Fase Modelado de datos

5.3.1. Comprensión de Negocio

Cuando se habla de Organización del Estado, se refiere a entidades gubernamentales del Perú que desempeñan actividades de servicios públicos, siendo los movimientos que realizan sus empleados en cuanto a eventos como asistencia, permisos, licencias que realizan miles de empleados del sector público el foco de interés para indagar sobre algún comportamiento oculto dentro de la información registrada de las ocurrencias mencionadas en cuestión de personas y cualidades, fechas, tiempo, eventos, etc.

Por lo que el negocio, es decir una organización del estado desea saber, en función al itinerario cíclico que realiza cada día o semana, saber que probabilidades se tienen para un determinado fenómeno, en este caso por ejemplo se realizara un modelo que determine sobre las categorías de los servicios una descripción exploratoria de los datos.

Siendo el objetivo del modelo:

Determinar patrones de comportamiento a través de la clasificación de datos en función a determinados estadios.

5.3.2. Comprensión de datos

Los datos proporcionados para esta ocasión pertenecen a una organización de tipo proyecto en el Proyecto Especial Altomayo de San Martín, esta base de datos se mantendrá anónima por respeto a la privacidad de la organización que ha donado su información.

La base de datos se encuentra en PostgreSQL con un backup generado para 2 años de información histórica de su base de datos transaccional.

Con un tamaño aproximado de 5 GB se ha procedido a hacer la restauración de la base de datos.

Aquí empieza lo interesante de la investigación, se han entregado 03 bases de datos correspondientes al sistema transaccional de la organización.



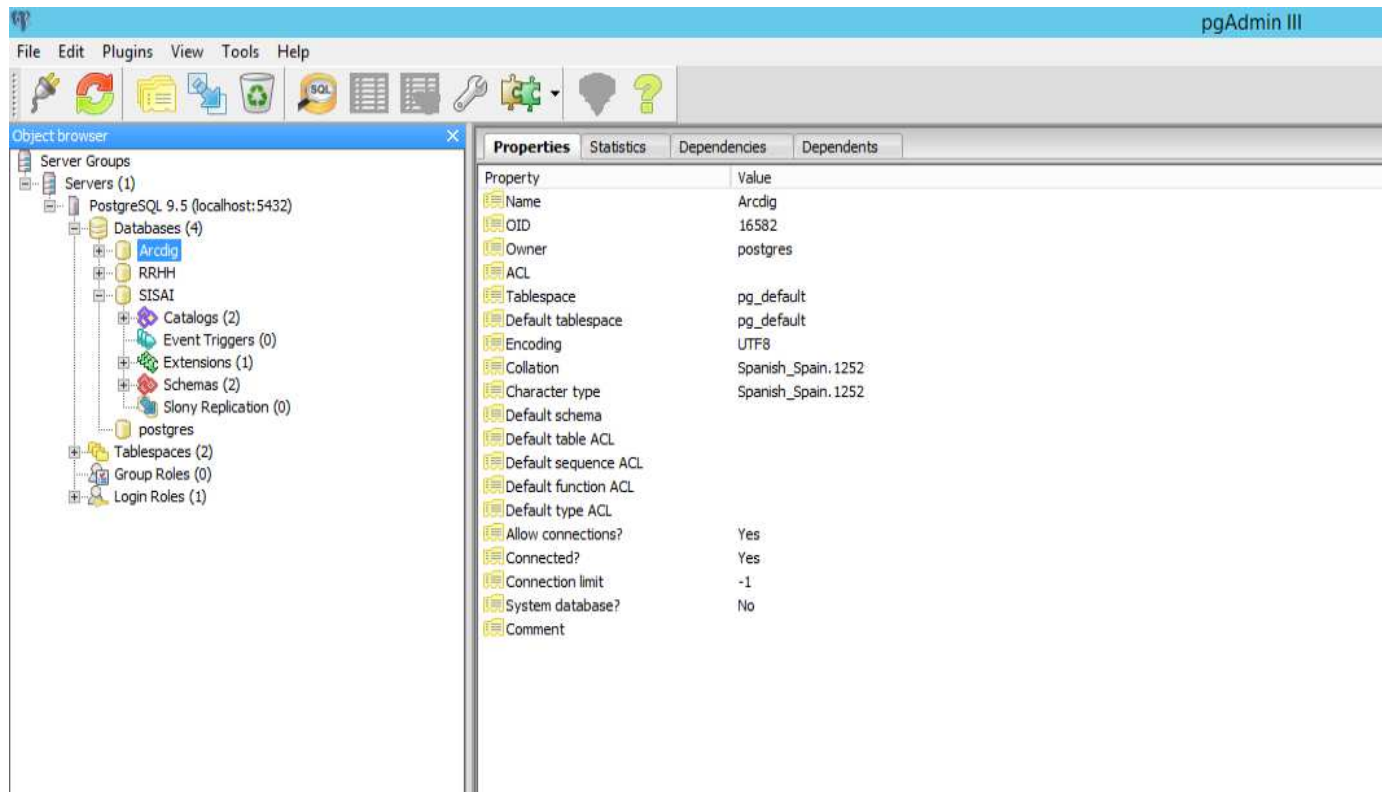


Figura 12: Vista en POSTGRESQL de base de datos
Fuente: Base de datos de la institución

Tabla 16:
Base de datos utilizadas

Base de datos	
ArcDig	Base de datos reservada
RRHH	Base de datos de Recursos Humanos
SISAI	Base de datos con datos de Personal

Nota: Fuente propia



Para esta investigación se trabajará con la base de datos RRHH y SISAI con respecto al módulo de Recursos Humanos y administración para datos del personal.

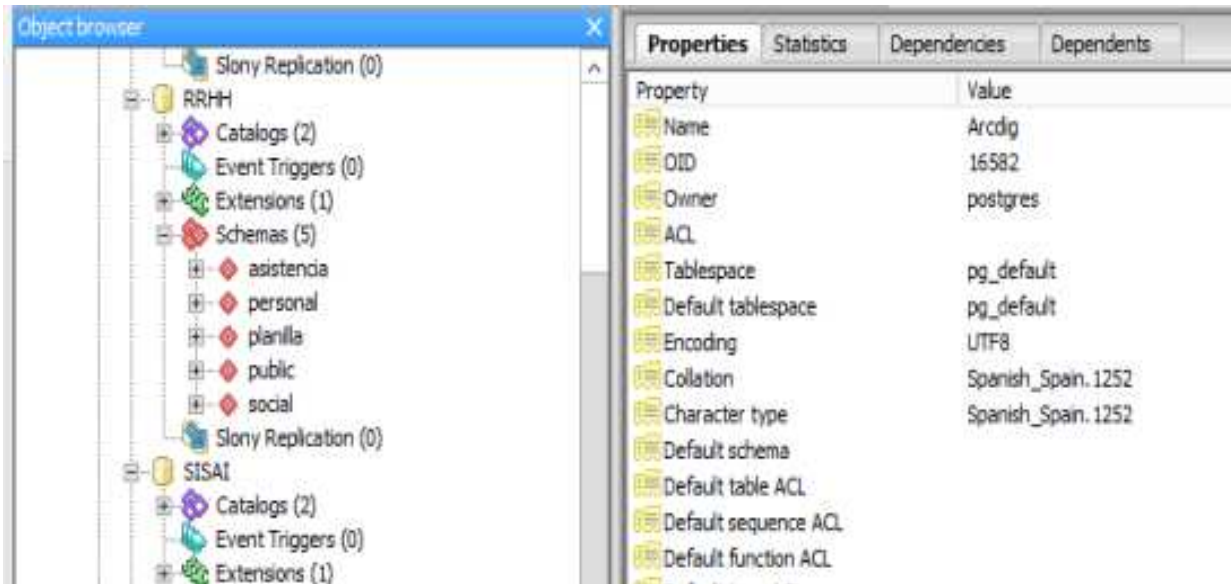


Figura 13: Vista de esquemas en POSTGRESQL de base de datos
Fuente: Base de datos de la institución

Tabla 17:
Esquemas de base de datos

RRHH DatabasesSchemas	
Asistencia	Información sobre asistencia
Personal	Datos de personal
Planilla	Planilla
Public	Publica
Social	Comunidad

Nota: fuente base de datos PEAM

Obviamente los esquemas de interés para esta investigación en las pruebas de clasificadores son las de personal, asistencia y planilla, para considerar variables en el minado de datos con características en los eventos de asistencia, permisos, licencia que puedan distinguirse de variables clasificadoras.



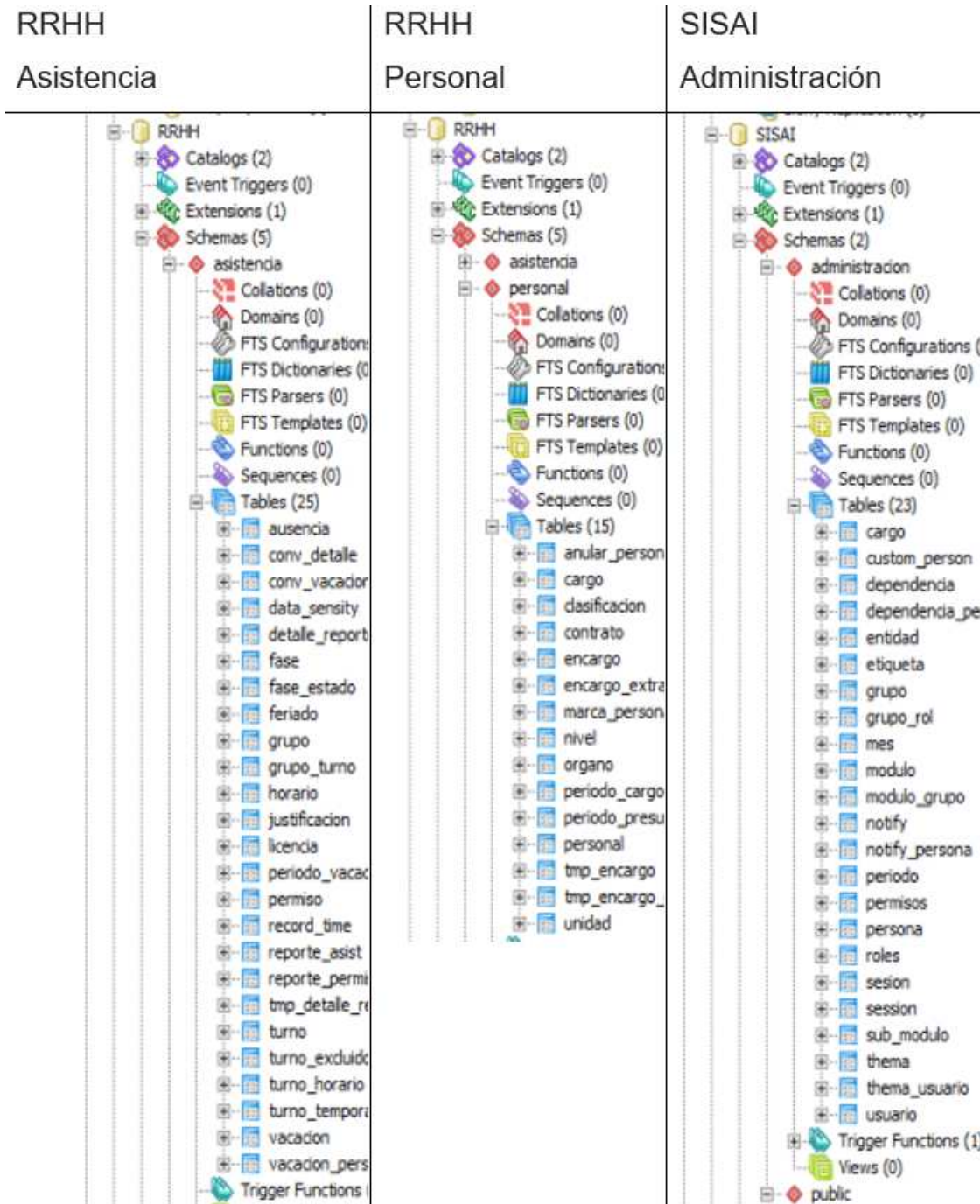


Figura 14: Vista de entidades en POSTGRESQL de bases de datos
Fuente: Base de datos de la institución



Para este caso se realizará una vista para consolidar la información de la base de datos a fin de obtener un Dataset para realizar el análisis de datos, ya que esta investigación no realizará el diseño de un datamart por cuestiones de tiempo y esfuerzo.

Dado que se realiza un estudio de algoritmos clasificadores se elaborará dicho procedimiento para generar un dataset histórico de datos.

5.3.3. Selección de datos

Se realizará un procedimiento ETL para el tratamiento de datos a fin de generar una vista resultante de análisis de datos por categorías con un destino de ocurrencia clase

```
destino <- dbConnect(PostgreSQL(),user="postgres",password="postgres",dbname="dataset")

t1 <- dbGetQuery(db1,"select * from personal.personal")
#dataframe
#head(t1)
#nrow(t1)
#ncol(t1)
t2 <- dbGetQuery(db2,"select * from administracion.persona")
t3 <- dbGetQuery(db2,"select * from administracion.dependencia")
#head(t3)
#nrow(t3)
#ncol(t3)
t4 <- dbGetQuery(db2,"select * from administracion.dependencia_persona")

t5 <- dbGetQuery(db1,"select * from asistencia.detalle_reporte
where horario like 'HORARIO PLANILLA MA%'")
#t6 <- dbGetQuery(db1,"select * from asistencia.detalle_reporte
#where horario like 'HORARIO PLANILLA MA%'")
t6 <- dbGetQuery(db1,"select *,case when fundamento like 'COO%' or fundamento like 'REU%' THEN 'INTRANE
ELSE 'EXTERNO' END AS tipo_permiso from asistencia.permiso ")

dbWriteTable(destino,"personaglobal",t1)
dbWriteTable(destino,"personaadmin",t2)
dbWriteTable(destino,"dependencia",t3)
dbWriteTable(destino,"dependencia_persona",t4)
dbWriteTable(destino,"reporte_asistencia",t5)
dbWriteTable(destino,"reporte_permiso",t6)
```

Figura 15: Código de selección de datos de las diferentes base de datos

Fuente: elaboración propia

Seleccionamos los datos personaglobal, personaadmin, dependencia, dependencia_persona, reporte_asistencia_reporte_permiso. se hace esta selección de datos de las diferentes base de datos para ser implementados en una nueva base de datos llamada dataset



5.3.4. Transformación de datos

El procedimiento de transformación parte desde el motor de tratamiento de datos en lenguaje R. Para esto se prepara el entorno utilizando las herramientas declaradas al inicio de este capítulo.

podemos encontrar datos que no pueden ser visualizados por el usuario convencional lo cual ese datos vamos a transformarlo.

como podemos visualizar el campo sexo y pers_estado esta identificado con numeros (0-1) y letras (true - false) lo cual proseguimos a transformalos

Configurando el entorno R

Instalación de Librería para conexión con base de datos PostgreSQL en R

	id	pers_global_id	pers_id	pers_estado	nombre	apellido_paterno	apellido_materno	sexo
1	000005	000005	57	TRUE	SILVIA	CARRANZA	RÃfÃOS	1
2	000041	000041	84	TRUE	AMADOR	UGARTE	ARBILDO	0
3	000083	000083	123	FALSE	LUDWIN	GÃfÃNGORA	LÃfÃPEZ	0
4	000049	000049	91	TRUE	GLORIA MARISOL	TELLO	ARBIETO	1
5	000024	000024	156	TRUE	ADÃfÃN	FABIAN	VERGARA	0
6	000043	000043	86	TRUE	WILSON	PINEDO	MORÃfÃN	0
7	000046	000046	19	TRUE	JOSEFA	MESÃfÃA	VÃfÃSQUEZ	1
8	000023	000023	155	TRUE	MARILYN	RIVERA	BRIONES	1
9	000035	000035	78	TRUE	MÃfÃXIMO DAVID	RAMOS	CHÃfÃVARRI	0
10	000020	000020	72	TRUE	SEBASTIAN	INOÃfÃCAN	GRANADOS	0
11	000077	000077	117	FALSE	JUAN MANUEL	IBAÃfÃEZ	GARCÃfÃA	0
12	000004	000004	56	FALSE	LUCY	MARINA	OLÃfÃRTEGUI	1
13	000188	000188	53	FALSE	JULIO CESAR	PISCO	TUESTA	0
14	000063	000063	103	FALSE	CARLOS JOSÃfÃ°	RIOFRÃfÃO	VELA	0
15	000010	000010	69	FALSE	LOLA	GÃfÃRATE	MELÃfÃ°NDEZ	1

Figura 16: Datos mostrados en R listo para ser transformados
Fuente: Elaboracion propia

```
#CONSULTA DE DATA SET DE ANALISIS PARA CLASIFICAR
q2 <- dbGetQuery(idb,"select z.id,sexoclase,estadocivil,nacionalidad,nombre,apellido_paterno,
apellido_materno,edad,
case when edad < 30 then 'joven' when edad > 31 and edad < 50 then 'adulto' else 'adulto mayor' end as tipo_edad,
depenombre,tipo_permiso from (
select id,case when sexo=1 then 'femenino' else 'masculino' end as sexoclase,
case when estado_civil=0 then 'casado' else 'soltero' end as estadocivil,
case when nacionalidad=1 then 'regional' else 'nacional'end as nacionalidad,
```

Figura 17: Codigo para realizar la transformacion de datos
Fuente: Elaboracion propia

en la siguiente imagen podemos visualizar los datos ya transformados en un lenguaje ya visualizados por el usuario convencional



	sexoclase	estado_civil	estado_civil	id	nombre	apellido_paterno	apellido_materno	edad
1	femenino	soltero	nacional	000115	SIRLEN	PÃfÃe°REZ	VÃfÃSQUEZ	37
2	femenino	casado	regional	000016	SUSANA	CARO	CHÃfÃVEZ	49
3	masculino	casado	regional	000086	ROLANDO	RIVA	VILLACORTA	39
4	masculino	casado	regional	000024	ADÃfÃN	FABIAN	VERGARA	48
5	femenino	casado	regional	000144	NATIVIDAD	RÃfÃOS	PUSCÃfÃN	52
6	femenino	soltero	nacional	000167	DARLYNG LUZ	LOPEZ	GONZALES	45
7	femenino	soltero	nacional	000023	MARILYN	RIVERA	BRIONES	55
8	masculino	soltero	nacional	000043	WILSON	PINEDO	MORÃfÃN	58
9	masculino	soltero	nacional	000035	MÃfÃXIMO DAVID	RAMOS	CHÃfÃVARRI	50
10	masculino	soltero	nacional	000020	SEBASTIAN	INOÃfÃCAN	GRANADOS	55
11	masculino	casado	regional	000077	JUAN MANUEL	IBAÃfÃEZ	GARCÃfÃA	44
12	femenino	soltero	nacional	000046	JOSEFA	MESÃfÃA	VÃfÃSQUEZ	66
13	femenino	soltero	nacional	000011	ELSA ELIZABETH	ZUMAETA	RUIZ	51
14	masculino	soltero	nacional	000032	ARMANDO	SOSA	VICTORIANO	54
15	femenino	casado	regional	000034	BETSY	CENTURIAÃfÃCAN	PIÃfÃCA	37
16	masculino	casado	regional	000198	LUIS ALBERTO	ZUMAETA	RUIZ	49

Figura 18: Visualizacion de datos transformados y mostrados en R
Fuente: Elaboracion propia

Configurando el entorno R

Instalación de Librería para conexión con base de datos PostgreSQL en R

```

1 # install.packages("RPostgreSQL")
2
3 #BIBLIOTECA DE CONEXION A BASE DE DATOS
4 library("RPostgreSQL")
5
6 #CONEXION A BASE DE DATOS DATASET
7 idb <- dbConnect(PostgreSQL(),user="postgres",password="postgres",dbname="dataset")

```

Figura 19: Conexion a base de datos creada "dataset"
Fuente: Elaboracion propia consola R



```
> install.packages("RPostgreSQL")
--- Please select a CRAN mirror for use in this session ---
also installing the dependency 'DBI'

probando la URL 'https://cloud.r-project.org/bin/windows/contrib/3.2/DBI_0.4-1.zip'
Content type 'application/zip' length 209778 bytes (204 KB)
downloaded 204 KB

probando la URL 'https://cloud.r-project.org/bin/windows/contrib/3.2/RPostgreSQL_0.4-1.zip'
Content type 'application/zip' length 448635 bytes (438 KB)
downloaded 438 KB

package 'DBI' successfully unpacked and MD5 sums checked
package 'RPostgreSQL' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
      C:\Users\Administrador\AppData\Local\Temp\RtmpgFSw6N\downloaded_packages
> |
```

Figura 20: Configuración en R para conexión a PostgreSQL
Fuente: R Project Console Propia

Una vez instalada la librería realizamos un test de prueba para conexión con nuestra base de datos.

```
> library("RPostgreSQL")
Loading required package: DBI
Warning messages:
1: package 'RPostgreSQL' was built under R version 3.2.5
2: package 'DBI' was built under R version 3.2.5
> db <- dbConnect(PostgreSQL(), user="postgres", password="postgres", dbname="RRHH")
> db
Error: objeto 'DB' no encontrado
> db
<PostgreSQLConnection>
> qt <- dbGetQuery(db,"select * from asistencia.licencia")
> qt
```

	id	año	id_m	id_p	fecha_peticion	fecha_inicio	fecha_final
1	037	2014	17	000011	2014-10-07 16:32:45	2014-09-22	2014-10-03
2	038	2014	26	000041	2014-10-07 16:49:03	2014-10-04	2014-10-11
3	013	2014	17	000051	2014-06-03 16:57:21	2014-05-21	2014-05-23
4	003	2015	17	000189	2015-02-03 10:54:41	2015-01-23	2015-01-23
5	039	2014	17	000034	2014-10-09 15:52:25	2014-09-15	2014-10-03
6	040	2014	17	000016	2014-10-21 15:59:13	2014-10-13	2014-10-15
7	041	2014	14	000084	2014-10-27 16:39:34	2014-10-23	2014-10-26
8	011	2014	17	000011	2014-06-03 16:50:20	2014-05-14	2014-05-23
9	008	2014	14	000011	2014-05-06 16:21:59	2014-04-21	2014-04-25
10	005	2014	14	000010	2014-05-06 15:03:05	2014-04-21	2014-04-25
11	014	2014	14	000192	2014-06-05 15:15:59	2014-06-02	2014-06-06
12	001	2014	35	000041	2014-04-29 11:54:45	2014-03-28	2014-07-15

Figura 21: Test conexión a PostgreSQL
Fuente: R Project console propia



Como se puede evidenciar en la imagen anterior, se puede realizar consultas a nivel de esquemas, como comúnmente se trabaja en PostgreSQL, dado a la librería RPostgreSQL que permite una conexión de tipo ODBC.

Una vez terminada la conexión establecida se procede a desarrollar el tratamiento de datos a partir de los resultantes de la selección de datos, dado que se evaluará un árbol de clasificación se debe determinar las variables a analizar y la clase objetivo del modelo.

Se construye a partir de las entidades a analizar un dataset básico conformado por las variables primarias de un empleado, así mismo se entrelaza entre las ocurrencias de “licencias” y “permisos” para determinar los aspectos más relevantes en cuanto a la productividad del personal.

Para lo cual se crea una base de datos repositorio para depositar los datos de la extracción:

```
> library("RPostgreSQL")
Loading required package: DBI
Warning messages:
1: package 'RPostgreSQL' was built under R version 3.2.5
2: package 'DBI' was built under R version 3.2.5
>
> db1 <- dbConnect(PostgreSQL(), user="postgres", password="postgres", dbname="RRHH")
> db2 <- dbConnect(PostgreSQL(), user="postgres", password="postgres", dbname="SISAI")
>
>
> idb <- dbConnect(PostgreSQL(), user="postgres", password="postgres", dbname="dataset")
>
> |
```

Figura 22: Código a múltiples conexiones en R
Fuente: R Project Console Propia

Esto quiere decir que se extraen datos de las dos bases de datos de la entidad y se volcaran en un dataset para el procesamiento del modelo a construir.




```
library("RPostgreSQL")

db1 <- dbConnect(PostgreSQL(),user="postgres",password="postgres",dbname="RRHH")
db2 <- dbConnect(PostgreSQL(),user="postgres",password="postgres",dbname="SISAI")

idb <- dbConnect(PostgreSQL(),user="postgres",password="postgres",dbname="dataset")

t1 <- dbGetQuery(db1,"select * from personal.personal")
t2 <- dbGetQuery(db2,"select * from administracion.persona")
t3 <- dbGetQuery(db2,"select * from administracion.dependencia")
t4 <- dbGetQuery(db2,"select * from administracion.dependencia_persona")

dbWriteTable(idb,"personalglobal",t1)
dbWriteTable(idb,"personaadmin",t2)
dbWriteTable(idb,"dependencia",t3)
dbWriteTable(idb,"dependencia_persona",t4)
```

Figura 23: Código para creación de nueva BD "dataset" e integración de tablas
Fuente: R Project Console Propia

En el gráfico anterior se muestra el script que permitirá volcar estas entidades teniendo finalmente nuestra base de datos para análisis

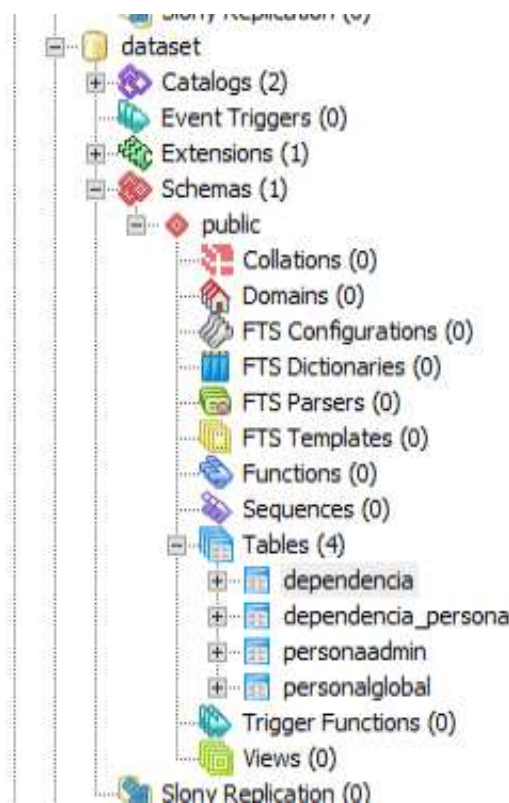


Figura 24: Esquema mostrando nueva BD "DataSet"
Fuente: PostgreSQL Propia



Finalmente se empiezan a concentrar los datos con la siguiente consulta:

```
> q1 <- dbGetQuery(idb,"select p.id,pe.pers_global_id,pe.pers_id,pe.pers_estado,p.nombre,p.apellido_paterno,p.apellido_materno,
+ p.sexo,p.estado_civil,p.fecha_nacimiento,
+ p.nacionalidad,dp.depe_id,d.depe_nombre from personalglobal p join
+ personaadmin pe on p.id = pe.pers_global_id
+ join dependencia_persona dp on pe.pers_id = dp.pers_id
+ join dependencia d on dp.depe_id = d.depe_id")
> q1
```

	id	pers_global_id	pers_id	pers_estado	nombre	apellido_paterno	apellido_materno	sexo
1	000005	000005	57	TRUE	SILVIA	CARRANZA	RÃfÃOS	1
2	000041	000041	84	TRUE	AMADOR	UGARTE	ARBILDO	0
3	000083	000083	123	FALSE	LUDWIN	GÃfÃNGORA	LÃfÃPEZ	0
4	000049	000049	91	TRUE	GLORIA MARISOL	TELLO	ARBieto	1
5	000024	000024	156	TRUE	ADÃfÃN	FABIAN	VERGARA	0
6	000043	000043	86	TRUE	WILSON	PINEDO	MORÃfÃN	0
7	000046	000046	19	TRUE	JOSEFA	MESÃfÃA	VÃfÃSQUEZ	1
8	000023	000023	155	TRUE	MARILYN	RIVERA	BRIONES	1
9	000035	000035	78	TRUE	MÃfÃXIMO DAVID	RAMOS	CHÃfÃVARRI	0
10	000020	000020	72	TRUE	SEBASTIAN	INOÃfÃCAN	GRANADOS	0
11	000077	000077	117	FALSE	JUAN MANUEL	IBAÃfÃEZ	GARCÃfÃA	0
12	000004	000004	56	FALSE	LUCY	MARINA	OLÃfÃRTEGUI	1
13	000188	000188	53	FALSE	JULIO CESAR	PISCO	TUESTA	0
14	000063	000063	103	FALSE	CARLOS JOSÃfÃ	RIOFRÃfÃO	VELA	0
15	000010	000010	69	FALSE	LUCIA	GÃfÃRTE	MPTÃfÃNDEZ	1

Figura 25: Codigo para generacion de datos del "dataset"
Fuente: R Project Console Propia

Sin embargo, aún existen campos que deben ser tratados, por ejemplos las columnas sexo, estado civil aún se conservan con sus propiedades de metadatos originales, para lo cual se mejora el script de extracción.

```
> q2 <- dbGetQuery(idb,"select case when sexo=1 then 'femenino' else 'masculino' end as sexoclase,
+ case when estado_civil=0 then 'casado' else 'soltero' end as estado_civil,
+ case when nacionalidad=1 then 'regional' else 'nacional' end as estado_civil,
+ id,nombre,apellido_paterno,apellido_materno,date_part('year',age(fecha_nacimiento)) as edad,depe_nombre from (
+ select p.id,pe.pers_global_id,pe.pers_id,pe.pers_estado,p.nombre,p.apellido_paterno,p.apellido_materno,
+ p.sexo,p.estado_civil,p.fecha_nacimiento,
+ p.nacionalidad,dp.depe_id,d.depe_nombre from personalglobal p join
+ personaadmin pe on p.id = pe.pers_global_id
+ join dependencia_persona dp on pe.pers_id = dp.pers_id
+ join dependencia d on dp.depe_id = d.depe_id) as q")
> q2
```

	sexoclase	estado_civil	estado_civil	id	nombre	apellido_paterno	apellido_materno	edad
1	femenino	soltero	nacional	000115	SIRLEN	PÃfÃREZ	VÃfÃSQUEZ	37
2	femenino	casado	regional	000016	SUSANA	CARO	CHÃfÃVEZ	49
3	masculino	casado	regional	000086	ROLANDO	RIVA	VILLACORTA	39
4	masculino	casado	regional	000024	ADÃfÃN	FABIAN	VERGARA	48
5	femenino	casado	regional	000144	NATIVIDAD	RÃfÃOS	PUSCÃfÃN	52
6	femenino	soltero	nacional	000167	DARLYNG LUZ	LOPEZ	GONZALES	45
7	femenino	soltero	nacional	000023	MARILYN	RIVERA	BRIONES	55
8	masculino	soltero	nacional	000043	WILSON	PINEDO	MORÃfÃN	58
9	masculino	soltero	nacional	000035	MÃfÃXIMO DAVID	RAMOS	CHÃfÃVARRI	50
10	masculino	soltero	nacional	000020	SEBASTIAN	INOÃfÃCAN	GRANADOS	55
11	masculino	casado	regional	000077	JUAN MANUEL	IBAÃfÃEZ	GARCÃfÃA	44
12	femenino	soltero	nacional	000046	JOSEFA	MESÃfÃA	VÃfÃSQUEZ	66
13	femenino	soltero	nacional	000011	ELSA ELIZABETH	ZUMAETA	RUIZ	51
14	masculino	soltero	nacional	000032	ARMANDO	SOSA	VICTORIANO	54
15	femenino	casado	regional	000034	BETSY	CENTURIÃfÃCAN	PIÃfÃA	37
16	masculino	casado	regional	000198	LUIS ALBERTO	ZUMAETA	RUIZ	49

Figura 26: Codigo para la transformacion de datos y mostrado en R
Fuente: R Project Console Propia

Con el algoritmo mejorado se obtiene un dataset con mayor detalle que servirá para el procesamiento del clasificador, para una prueba de concepto en la etapa de modelado se evaluará con este pequeño dataset antes de agregarle las otras variables de análisis.



5.3.5. Modelado

En la fase de modelado se procede a tomar el dataset generado para el análisis del modelo, el cual utilizara 2 algoritmos para evaluar la tasa de rendimiento del mismo.

Pasos para el modelado

Paso 1

Cargar datos de set de entrenamiento y test, en este caso se utilizará una muestra que se divide en set de entrenamiento para que el algoritmo genere las reglas y el set de prueba para medir la precisión de la predicción, el resultado serial algo como:

En este caso tenemos 2500 datos. Entonces 2000 serian de entrenamiento y 500 de prueba esto quiere decir que de un 100%. 75% será de entrenamiento y un 25% de test.

	estado_civil text	estado_civil text	id text	nombre text	apellido_pate text	apellido_mat text	edad double	depe_nombre text	sexoclase text
1	soltero	nacional	000115	SIRLEN	PÁ&REZ	VÁSQUEZ	37	PROYECTO ESPECIAL AL	femenino
2	casado	regional	000016	SUSANA	CARO	CHÁVEZ	49	GERENCIA	femenino
3	casado	regional	000086	ROLANDO	RIVA	VILLACORTA	39	GERENCIA	masculino
4	casado	regional	000024	ADÁN	FABIAN	VERGARA	48	DIRECCION DE MANEJO	masculino
5	casado	regional	000144	NATIVIDAD	RÃ&OS	PUSCÁN	52	DIRECCION DE MANEJO	femenino
6	soltero	nacional	000167	DARLYNG LUZ	LOPEZ	GONZALES	45	DIRECCION DE MANEJO	femenino
7	soltero	nacional	000023	MARILYN	RIVERA	BRIONES	55	DIRECCION DE MANEJO	femenino
8	soltero	nacional	000043	WILSON	PINEDO	MORÁN	58	DIRECCION DE MANEJO	masculino
9	soltero	nacional	000035	MÁXIMO DAVID	RAMOS	CHÁVARRI	50	DIRECCION DE MANEJO	masculino
10	soltero	nacional	000020	SEBASTIAN	INOÁ'AN	GRANADOS	55	DIRECCION DE MANEJO	masculino
11	casado	regional	000077	JUAN MANUEL	IBAÃ'EZ	GARCÃ&A	44	DIRECCION DE MANEJO	masculino
12	soltero	nacional	000046	JOSEFA	MESÃ&A	VÁSQUEZ	66	DIRECCION DE MANEJO	femenino
13	soltero	nacional	000011	ELSA ELIZABETH	ZUMAETA	RUIZ	51	DIRECCION DE MANEJO	femenino
14	soltero	nacional	000032	ARMANDO	SOSA	VICTORIANO	54	DIRECCION DE INFRAES	masculino
15	casado	regional	000034	BETSY	CENTIURIÃ&N	PIÃ&A	37	DIRECCION DE INFRAES	femenino
16	casado	regional	000198	LUIS ALBERTO	ZUMAETA	RUIZ	49	DIRECCION DE INFRAES	masculino
17	casado	regional	000029	JAVIER ALFONSO	CARO	PEREA	56	DIRECCION DE INFRAES	masculino
18	soltero	nacional	000049	GLORIA MARISOL	TELLO	ARBieto	47	DIRECCION DE INFRAES	femenino
19	soltero	nacional	000044	ULPIANO	VELA	MUÃ&OZ	49	DIRECCION DE INFRAES	masculino
20	soltero	nacional	000042	LEYDI	VELA	MACEDO	39	DIRECCION DE INFRAES	femenino
21	casado	regional	000018	ZOILA	PINEDO	PINEDO	59	DIRECCION DE DESARRO	femenino
22	soltero	nacional	000012	LEONARDO	BARDALEZ	VILLACORTA	53	DIRECCION DE DESARRO	masculino
23	casado	regional	000093	ALVARO	PRADA	GUADALUPE	36	DIRECCION DE DESARRO	masculino
24	casado	regional	000095	MIGUEL ANGEL	LEVEAU	TUESTA	53	DIRECCION DE DESARRO	masculino

SET DE ENTRENAMIENTO

SET DE PRUEBA

Figura 27: Dataset creada y lista para el test de entramiento y el test de prueba

Fuente: PostgreSQL Propia



Paso 2

Se crea el modelo clasificador, para ello se aplica el algoritmo a utilizar (Árbol de clasificación o Red Bayesiana) para que se generen las reglas de aprendizaje)

Para ello se procede a instalar las librerías correspondientes en R, en este caso se utilizarán:

```
> install.packages("C50")
--- Please select a CRAN mirror for use in this session ---
also installing the dependencies 'Formula', 'partykit'

probando la URL 'https://cloud.r-project.org/bin/windows/contrib/3.2/Formula_1.2-1.zip'
Content type 'application/zip' length 163520 bytes (159 KB)
downloaded 159 KB

probando la URL 'https://cloud.r-project.org/bin/windows/contrib/3.2/partykit_1.1-0.zip'
Content type 'application/zip' length 1229006 bytes (1.2 MB)
downloaded 1.2 MB

probando la URL 'https://cloud.r-project.org/bin/windows/contrib/3.2/C50_0.1.0-24.zip'
Content type 'application/zip' length 461265 bytes (450 KB)
downloaded 450 KB

package 'Formula' successfully unpacked and MD5 sums checked
package 'partykit' successfully unpacked and MD5 sums checked
package 'C50' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Administrador\AppData\Local\Temp\Rtmp4Sq9z7\downloaded_packages
```

Figura 28: Instalando librería para los algoritmos clasificadores

Fuente: R Project Console Propia

```
> install.packages("rpart")
probando la URL 'https://cloud.r-project.org/bin/windows/contrib/3.2/rpart_4.1-10.zip'
Content type 'application/zip' length 921857 bytes (900 KB)
downloaded 900 KB

package 'rpart' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Administrador\AppData\Local\Temp\Rtmp4Sq9z7\downloaded_packages
> library(rpart.plot)
Error in library(rpart.plot) : there is no package called 'rpart.plot'
> install.packages("rpart.plot")
probando la URL 'https://cloud.r-project.org/bin/windows/contrib/3.2/rpart.plot_2.0.1.zip'
Content type 'application/zip' length 704965 bytes (688 KB)
downloaded 688 KB

package 'rpart.plot' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Administrador\AppData\Local\Temp\Rtmp4Sq9z7\downloaded_packages
```

Figura 29: Instalando librería para los algoritmos clasificadores

Fuente: R Project Console Propia

Se instalan las librerías con los algoritmos para el entrenamiento a continuación se procede con el script de aprendizaje:

El primer algoritmo a aplicar para este modelo es el de árbol de clasificación



```
> arbol = rpart(sexoclase ~ estadocivil + nacionalidad + edad + depenombre, data = dq2)
> arbol
n= 63

node), split, n, loss, yval, (yprob)
 * denotes terminal node

1) root 63 22 masculino (0.3492063 0.6507937)
2) depenombre=CONTROL PATRIMONIAL,OFICINA DE ASESORIA JURIDICA,ORGANO DE CONTROL INSTITUCIONAL,PERSONAL,PROYECT
3) depenombre=ABASTECIMIENTO Y SERVICIOS AUXILIARES,AREA DE ESTUDIOS,CONTABILIDAD,DIRECCION DE DESARROLLO AGROP
6) depenombre=CONTABILIDAD,DIRECCION DE INFRAESTRUCTURA,DIRECCION DE MANEJO AMBIENTAL,GERENCIA 25 11 masculin
12) edad< 48.5 9 3 femenino (0.6666667 0.3333333) *
13) edad>=48.5 16 5 masculino (0.3125000 0.6875000) *
7) depenombre=ABASTECIMIENTO Y SERVICIOS AUXILIARES,AREA DE ESTUDIOS,DIRECCION DE DESARROLLO AGROPECUARIO,INF
> plotcp(arbol)
> |
```

Figura 30: Aplicando el algoritmo arbol de decision
Fuente: R Project Console Propia

Obteniendo las primeras pruebas e imágenes resultantes de la prueba de concepto del laboratorio aplicado.

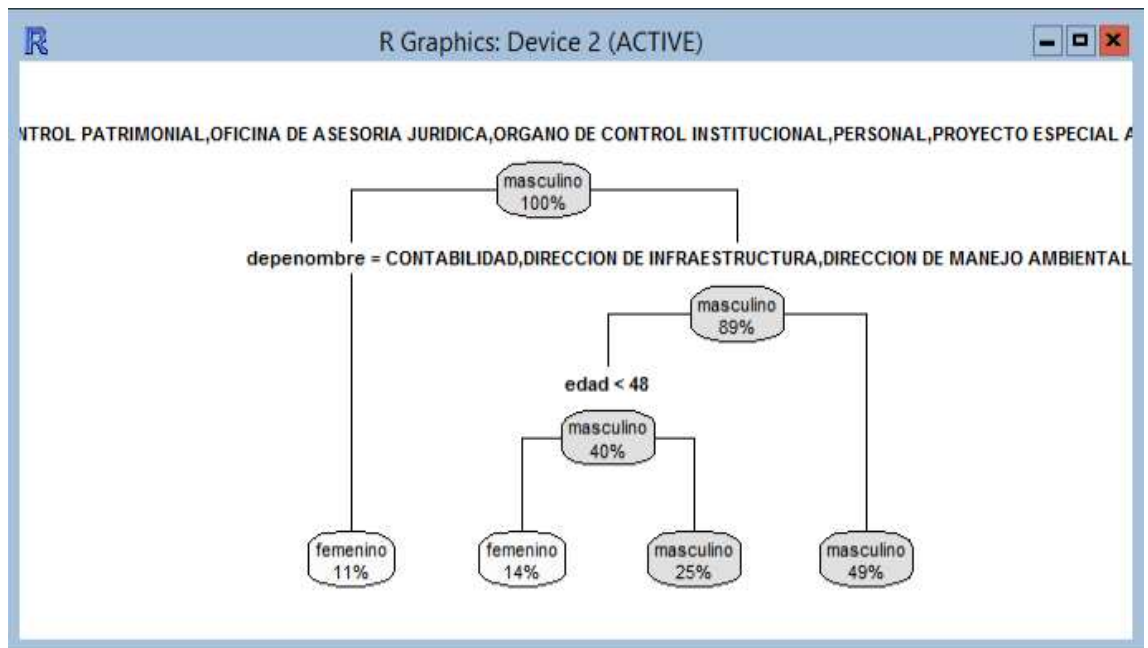


Figura 31: Generando grafico en algoritmo arbol de clasificación
Fuente: R Project Console Propia

En la segunda fase de este apartado se empleará el algoritmo de redes bayesianas para la clasificación usando la misma data de laboratorio instalando la librería e1071



```
> install.packages("e1071")
probando la URL 'https://cloud.r-project.org/bin/windows/contrib/3.2/e1071_1.6-7.zip'
Content type 'application/zip' length 806820 bytes (787 KB)
downloaded 787 KB

package 'e1071' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
  C:\Users\Administrador\AppData\Local\Temp\Rtmp4Sq9z7\downloaded_packages
> |
```

Figura 32: Instalando libreria para clasificador red bayesiana
Fuente: R Project Console Propia

```
> model <- naiveBayes(sexoclase ~ ., data = dq2)
> model

Naive Bayes Classifier for Discrete Predictors

Call:
naiveBayes.default(x = X, y = Y, laplace = laplace)

A-priori probabilities:
Y
  femenino masculino
0.3492063 0.6507937

Conditional probabilities:
id
Y
  femenino masculino
0.000002 0.000003 0.000004 0.000005 0.000006 0.000010 0.000011 0.000012 0.000013
femenino 0.04545455 0.00000000 0.04545455 0.04545455 0.04545455 0.04545455 0.04545455 0.00000000 0.00000000
masculino 0.00000000 0.02439024 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.02439024 0.02439024
id
Y
  femenino masculino
0.000015 0.000016 0.000018 0.000020 0.000021 0.000023 0.000024 0.000027 0.000029
femenino 0.00000000 0.04545455 0.04545455 0.00000000 0.00000000 0.04545455 0.00000000 0.00000000 0.00000000
masculino 0.02439024 0.00000000 0.00000000 0.02439024 0.02439024 0.00000000 0.02439024 0.02439024 0.02439024
id
Y
  femenino masculino
0.000030 0.000032 0.000033 0.000034 0.000035 0.000039 0.000040 0.000041 0.000042
femenino 0.00000000 0.00000000 0.00000000 0.04545455 0.00000000 0.00000000 0.00000000 0.00000000 0.04545455
masculino 0.02439024 0.02439024 0.02439024 0.00000000 0.02439024 0.02439024 0.02439024 0.02439024 0.00000000
id
Y
  femenino masculino
0.000043 0.000044 0.000046 0.000048 0.000049 0.000054 0.000056 0.000061 0.000063
femenino 0.00000000 0.00000000 0.04545455 0.04545455 0.04545455 0.00000000 0.00000000 0.04545455 0.00000000
masculino 0.02439024 0.02439024 0.00000000 0.00000000 0.00000000 0.02439024 0.02439024 0.00000000 0.02439024
id
Y
  femenino masculino
0.000073 0.000074 0.000077 0.000083 0.000085 0.000086 0.000088 0.000092 0.000093
femenino 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000 0.00000000
masculino 0.02439024 0.02439024 0.02439024 0.02439024 0.02439024 0.02439024 0.02439024 0.02439024 0.02439024
```

Figura 33: Resultados aplicando clasificador red bayesiana
Fuente: R Project Console Propia

Paso 3

5.3.6. Evaluación

Este ítem se describe en el capítulo IV de la propuesta

5.3.7. Fase Modulo Web de Análisis de Algoritmos

A. Análisis

Para la fase de desarrollo de un aplicativo que servirá para el análisis de los resultados se contempla los requerimientos de la entidad donde se realiza la investigación, estos



requerimientos se dividen básicamente en el propósito del desarrollo del modelo y por otro lado se busca una interfaz que permita visualizar los aspectos técnicos del modelo.

Usuarios de sistema

En los usuarios del sistema se considera:

Analista de RRHH: Es el rol que monitorea y opera el aplicativo para la extracción de reportes del mismo en función a la nueva alimentación de datos.

Ingeniero BI: Es el rol que se encarga del mantenimiento del modelo a partir de la revisión de la performance y modificación del algoritmo del modelo, así como los reportes de evaluación del modelo.

Administrador TI: Es el rol que cumple con la gestión de usuarios y seguridad del sistema.

Requisitos Funcionales

El aplicativo debe considerar la seguridad de los datos por lo tanto debe gestionar los ingresos al mismo.

El aplicativo debe mostrar una interfaz que permita visualizar las predicciones del modelo para un determinado escenario de validación a partir del ingreso de datos. Esta interfaz debe permitir la exportación de los resultados del modelo clasificador.

El aplicativo debe mostrar una interfaz que me permita visualizar el código nativo del modelo construido para su mantenimiento.

El aplicativo debe mostrar una interfaz para evaluar el rendimiento de los algoritmos utilizados a fin de determinar que script utilizar.

Requisitos No Funcionales

El aplicativo debe ser web para que sea accedido desde cualquier ordenador sin restricción de instalación de software.

El aplicativo debe soportar exportar reporte en PDF o Excel.

Del análisis de requisitos anterior se obtiene las historias de usuario:

BACKLOG DE HISTORIAS DE USUARIO



Tabla 18:
Historial de usuario

Historia de Usuario	Descripción
Acceso de sistema	El sistema mostrara una interfaz con un menú de acceso para que se ingrese y validen las credenciales del usuario
Módulo de Clasificación	El sistema muestra una interfaz para recoger los datos procesados como reporte
Módulo de Modelo	El sistema muestra una interfaz para analizar y editar el modelo diseñado
Módulo de sistema	El sistema muestra una interfaz para la gestión de los usuarios

Nota: Historial de usuario PEAM

B. Diseño

En la fase de diseño se procede a establecer la arquitectura de la aplicación a construir:

Arquitectura de la solución

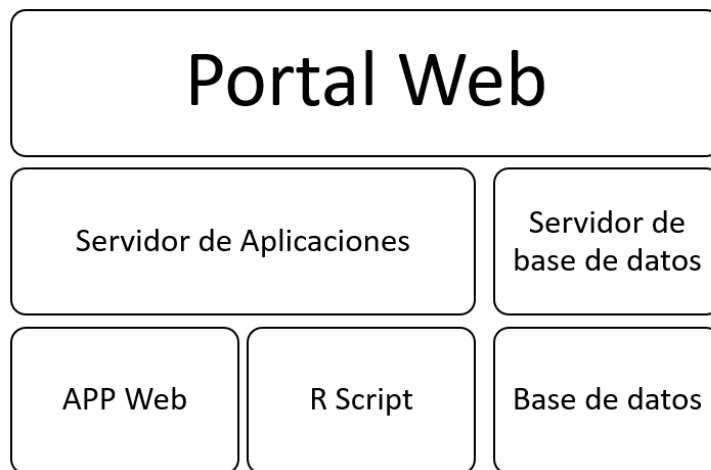


Figura 34: Arquitectura de la solución web
Fuente: Propia



Diseño de base de datos

Tecnología Web

En el aspecto de la codificación web se ha utilizado como lenguaje de programación PHP usando como base una estructura HTML5 con Javascript y CSS en la cual la relación es:

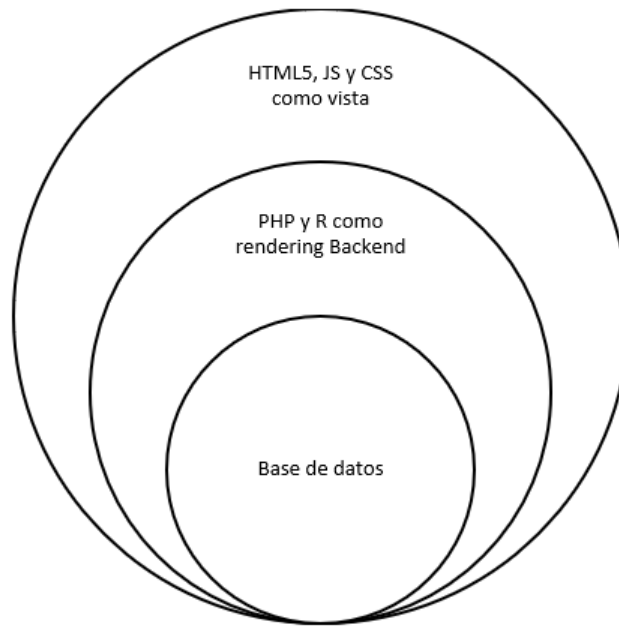


Figura 35: Capaz de estructura y tecnología web a construir
Fuente: Propia

C. Codificación

Para la codificación se ha utilizado una arquitectura monolítica, es decir que esta no obedece a patrones de diseño de software, y la razón se debe a que se trata en cuestión de un aplicativo de visualización de reportes con un mínimo trabajo transaccional, cual propósito es visualizar los resultados del modelo, es por ello que incluso no se necesita emplear FRAMEWORK de desarrollo alguno.

En la fase de codificación se utiliza la herramienta sublime text.

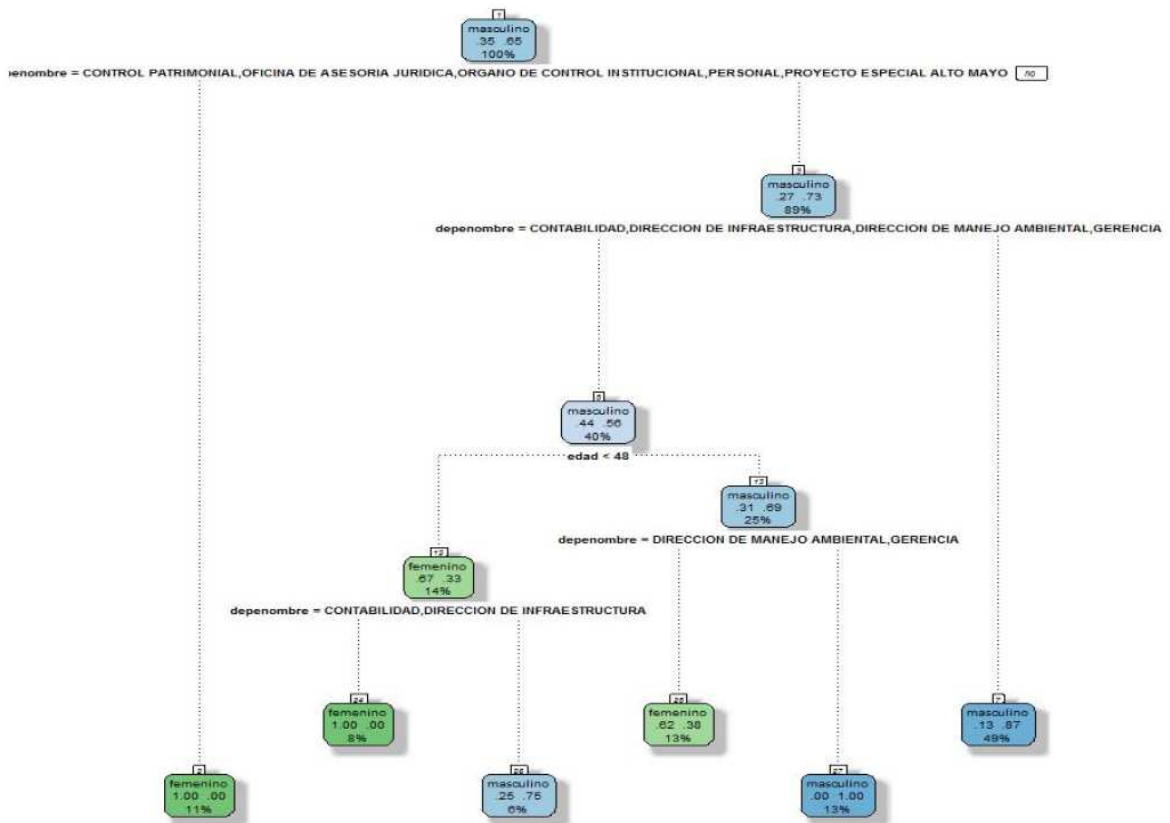


D. Pruebas

Sexo	Estado Civil	Nacionalidad	Nombre	Apellido Paterno	Apellido Materno	Edad	Tipo Edad	Dependencia	Tipo Permiso Real	Tipo Permiso Arbol	Tipo Permiso Red Bayesiana
masculino	soltero	nacional	ORLANDO	RAMÁZCUEZ	CUESTA	61	adulto mayor	DIRECCION DE DESARROLLO AGROPECUARIO	EXTERNO	EXTERNO	EXTERNO
femenino	casado	regional	ZOILA	PINEDO	PINEDO	62	adulto mayor	DIRECCION DE DESARROLLO AGROPECUARIO	EXTERNO	EXTERNO	EXTERNO
masculino	soltero	nacional	ULPIANO	VELA	MUÑOZ	52	adulto mayor	DIRECCION DE INFRAESTRUCTURA	EXTERNO	EXTERNO	EXTERNO
femenino	soltero	nacional	JOSEFA	MESAÑA	VÁSQUEZ	69	adulto mayor	DIRECCION DE MANEJO AMBIENTAL	EXTERNO	INTRANET	INTRANET
femenino	soltero	nacional	MILKA	ZAGACETA	GARCÍA MEZ	46	adulto	ORGANO DE CONTROL INSTITUCIONAL	EXTERNO	EXTERNO	EXTERNO
masculino	soltero	nacional	SEGUNDO MIGUEL	DIAZ	LA TORRE	56	adulto mayor	DIRECCION DE DESARROLLO AGROPECUARIO	EXTERNO	EXTERNO	EXTERNO
masculino	casado	regional	JULIO CESAR	PISCO	TUESTA	55	adulto mayor	OFICINA DE ADMINISTRACION	EXTERNO	EXTERNO	EXTERNO
femenino	casado	regional	SILVIA	CARRANZA	RAMOS	51	adulto mayor	CONTABILIDAD	EXTERNO	EXTERNO	EXTERNO
masculino	casado	regional	LUIS FERNANDO	JARAMILLO	MORALES	60	adulto mayor	OFICINA DE PRESUPUESTO, PLANIFICACION, ESTUDIOS Y ORDENAMIENTO TERRITORIAL	EXTERNO	EXTERNO	EXTERNO
femenino	casado	regional	ASIRIA	DIAZ	PAREDES	43	adulto	PERSONAL	INTRANET	INTRANET	INTRANET

Figura 36: Primeras pruebas de grafica web
Fuente: propia





Rattle 2016-sep.-09 17:50:19 Administrador

Figura 37: Grafica prueba arbol de decisiones

Fuente: PHP Console Propia



CAPITULO VI – CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

- a) Se procedió con la recopilación de datos a partir del análisis de datos de los archivos de tipo Backup entregados por la entidad, los cuales son 03 bases de datos correspondientes a los sistemas que se utilizan en la parte administrativa, y que contienen la información de las personas y los eventos de RRHH como es el caso de los permisos y las licencias, utilizando como variables los datos del personal, tanto como personales o como relacionados por el área donde se labora. Asimismo, se construyeron los algoritmos de procedimiento para extraer dichos datos.
- b) Para el set de datos recopilados luego de la fase de extracción, se determina que la calidad de datos y tipo de datos pertenecen a variables categóricas de tipo binomial y polinomial que permiten aplicar los algoritmos clasificadores, dada esta naturaleza se determina que los algoritmos de árbol de clasificación y redes bayesianas son los que se adecuan a los datos a examinar para generar el aprendizaje en la fase de entrenamiento. Para esto se desarrolló el modelo determinando en la fase de evaluación que el algoritmo de árbol de decisiones tiene un menor rendimiento sobre la red bayesiana en cuanto al propósito de predicción.
- c) Se implementó un portal web de tipo analítico para visualizar los datos procesados por el modelo en función a reportes del modelo, por otro lado, se diseñó un módulo que permite observar los resultados de diversos algoritmos en entrenamiento para determinar el script del modelo a utilizar.
 - a) En el caso de rendimiento se a considerado 3 y 6 meses sedetermino que la el árbol de decisión tiene mejor redimiento que red bayesiana teniendo en cuenta que este algoritmo no cuenta con muchos parámetros a configurar
 - b) En cuanto tiempo utilizado el árbol de decisiones consumo mucho tiempo de procesamiento a diferencia de la red bayesiana



Recomendaciones

- a) Se recomienda realizar los procesos ETL para integración y aseguramiento de la calidad de los datos a analizar, así como también el formato pre requisito en función al estado clase u objetivo del modelo.
- b) Se recomienda categorizar las variables en sus estados real, polinomial o binomial para el modelo.
- c) Las redes bayesianas son útiles cuando se tratan de estado clase de tipo binomial pero pierden su capacidad cuando se tratan de estados polinomiales.

REFERENCIAS

- Ignacio Perversi . (2007). *Aplicacion de mineria de datos para la exploracion y deteccion de patrones delictivos de argentina*. Buenos aires: Instituto tecnologico de buenos aires.
- Alina, D. C., & Marrero Delgado, F. (2013). *El modelo scor y el balanced scorecard, una poderosa combinacion intangible para la gestion empresarial*. Villa Clara.
- Andrade, J. F., Cadena, C. A., & Cuenca, M. J. (2013). *El balanced scorecard, una herramienta para la planeacion estrategica*. Quito.
- Aspiazu, G. C. (2010). *Proceso de la mineria de datos*.
- Bazan, E. A. (2013). *Implementacion del Balanced Scorecard para mejorar la toma de decisiones de la empresa NOVACUMB S.I.R.L*. Chiclayo.
- Beard, D. (2005). *Balanced Score Card*.
- Berrios, F. S. (2012). *Ingles Colaborativo*. Santiago de Chile: Universidad de Chile.
- Broggi, A. (2010). *Metodologia para la mejor administracion de los recursos humanos en la gestion de empresas de servicio en etapa de maduracion*. Buenos Aires.
- CLEMENTE ANTONIO MARTÍNEZ ÁLVAREZ. (2012). *APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS PARA MEJORAR*. Santiago de Chile: Universidad de Chile.
- Cowley, M., & Domb, E. (2012). *Beyond Strategic Vision*. Taylor & Francis.
- Crecenegocios. (1 de Febrero de 2014). *Pymex.pe*. Obtenido de <http://pymex.pe/pymes/estrategias-de-crecimiento/la-importancia-de-las-herramientas-de-gestion-empresarial>
- Datosperu. (Lunes de Febrero de 2014). *DatosPeru*. Obtenido de <http://www.datosperu.org/peru-cobra-peru-sa.php>
- Diaz, Jorge Alfredo Salas. (2013). *Implementacion de una aplicacion web utilizando mineria de datos para mejorar la gestion de facturacion en la empresa PEXPORT SAC*. Chiclayo - Peru: Universidad Señor de Sipan.
- Ducker, P. (2009). *Gerencia para el futuro*. Madrid: Prentice Educación.
- Gauchat, J. D. (2012). *El gran libro de HTML5, CSS3 y Javascript*. Barcelona: Marcombo.
- Herman Jose, Ramirez Jose, Ferri Cesar. (2004). *Introducción a la minería da datos*. Madrid: pearson prentice hall.
- Heurtel, O. (2011). *PHP 5.3: Desarrollar un sitio Web dinamico e interactivo*. Ediciones ENI.
- Hober Willy Siccha Vega. (2012). *Minería de Datos aplicados a las ventas con Tarjeta de Crédito realizados en las tiendas Saga Falabella en la ciudad de Lima*. Lima - Peru: universidad tecnologica del peru.
- Kaplan, R., & Norton, D. (1996). Using the Balanced Scorecard as a Strategic Management System. *Harvard Business Review*.
- Kaplan, R., & Norton, D. (2007). *Mapas Estratégicos* . Ediciones Gestion 2000.
- Logreira, C. (2011). *Mineria de datos y su incidencia en la toma de decisiones empresariales en el contexto CRM*. Barranquilla: Revista Ingenieria Solidaria.
- López, O. (2010). *Creando una empresa de éxito*. Barcelona, España: Biblioteca Administrativa.
- Magretta, H. (2002). *Modelo de negocio: México - México*. Distrito Federal, México: Negociar estudios.



- Marcano Aular, Y. Z., & Talavera Pereira, R. (2007). *Minería de datos como soporte a la toma de decisiones empresariales*. Maracaibo.
- Maria Jose, Y. N., & Carvajal Ramos, P. C. (2013). *Modelo de Administracion estrategica denominado Balanced Score Card (BSC) para las PYMES financieras de la ciudad de Babahoyo*. Babahoyo.
- Maza, M. A. (2012). *Javascript*. Innovacion Y Cualificacion.
- Mendoza Peña, D. A. (2015). *Cambio en el modelo administrativo del Complejo Cinematográfico "Mis Cines", aplicando el Modelo de Negocio Canvas*. Quito: Quito.
- Niven, P. R. (2003). *El cuadro de mando integral paso a paso*. Barcelona: Gestion 2000.
- Norton, D. P., & Kaplan, R. (2009). *El Cuadro de Mando Integral*. Gestion 2000.
- Norton, K. y. (1992). The Balanced Scorecard Measures that Drive Performance. *Harvard Business Review*.
- Osterwalder, A. (2009). *Desarrollo de modelos de negocio*. Madrid, España: Díaz de Saltos.
- Ramírez A., A. Y. (2007). Técnicas de Minería de Datos Aplicadas a la Construcción de Modelos de Score Crediticio. *Mathematical Problems in Engineering*.
- Ricart, J. E. (2009). *The missing link in strategic management*. IESE Uni. Navarra.
- Rivadeneira, R. M. (2014). *academia.edu*. Obtenido de http://www.academia.edu/4190459/EL_Balanced_Scorecard_desde_la_perspectiva_del_aprendizaje
- Roque, B. R. (2011). *Balanced Scorecard con apoyo de las TI como soporte para la administracion de proyectos en CONTRATISTAS GENERALES S.A.C.* . Lima.
- Sadath, L. (2013). *Data Mining: A Tool for Knowledge*. International Journal of Innovate Technology and Exploring Engineering.
- Sholom M. Weiss, Nitin Indurkha. (1998). *Predictivo de minería de datos: Una Guía Práctica*. Estados unidos: Morgan Kaufmann Publishers.
- Sumathi S. y Shivanandam S. (2006). *Introduction de minería de datos y sus aplicaciones.* . USA: Springer-Verlag.
- Treacy, M. (2007). *The discipline of market leaders: Choose your customers*. Basic books.
- Valdez, D. P. (2007). *Los diferentes lenguajes de programacion para web*.
- Witten y Frank. (2005). *Minería de datos: prácticos máquinas herramientas de aprendizaje y técnicas, segunda edición*. España: Elsevier.
- Zott, A. &. (2001). *Gestión estratégica y venta en línea*.



ANEXOS

Anexo 1 Territorio Mapa

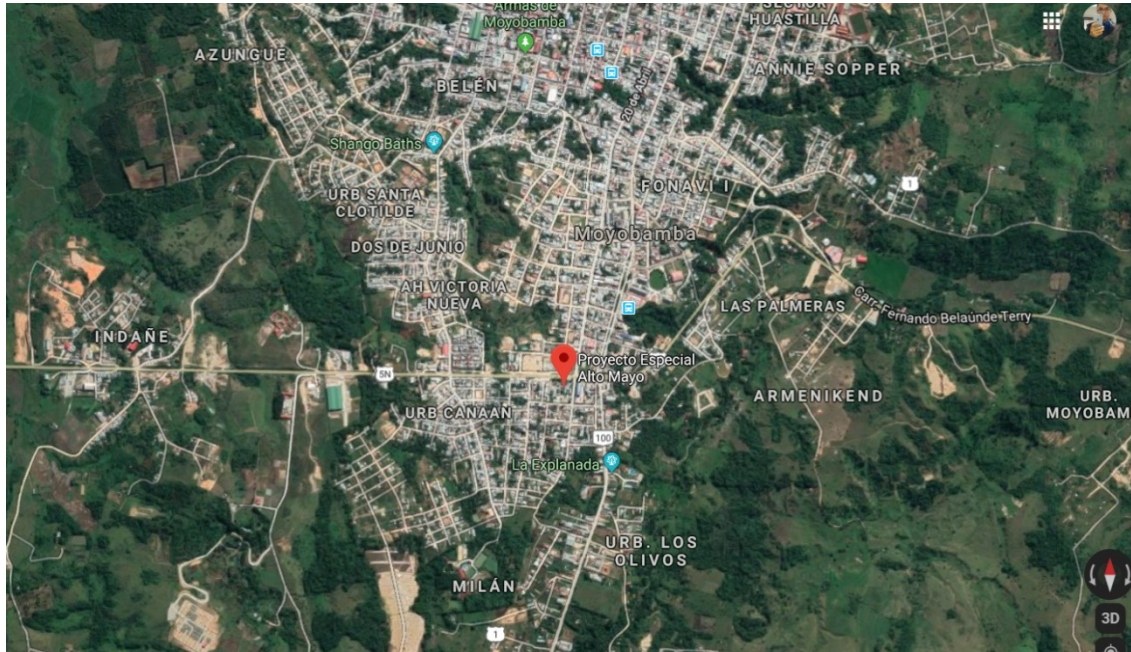


Figura 38: Territorio - ubicacion PEAM

Fuente: Elaboración Propia

Anexo 2 : Base del PROYECTO ESPECIAL ALTOMAYO

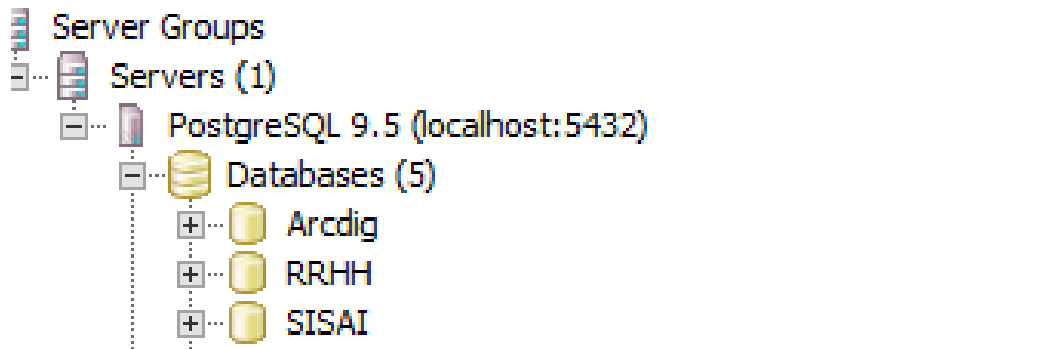


Figura 39: Base de datos PEAM

Fuente: Elaboración Propia



Anexo 3: Solicitud para inicio del trabajo de investigacion

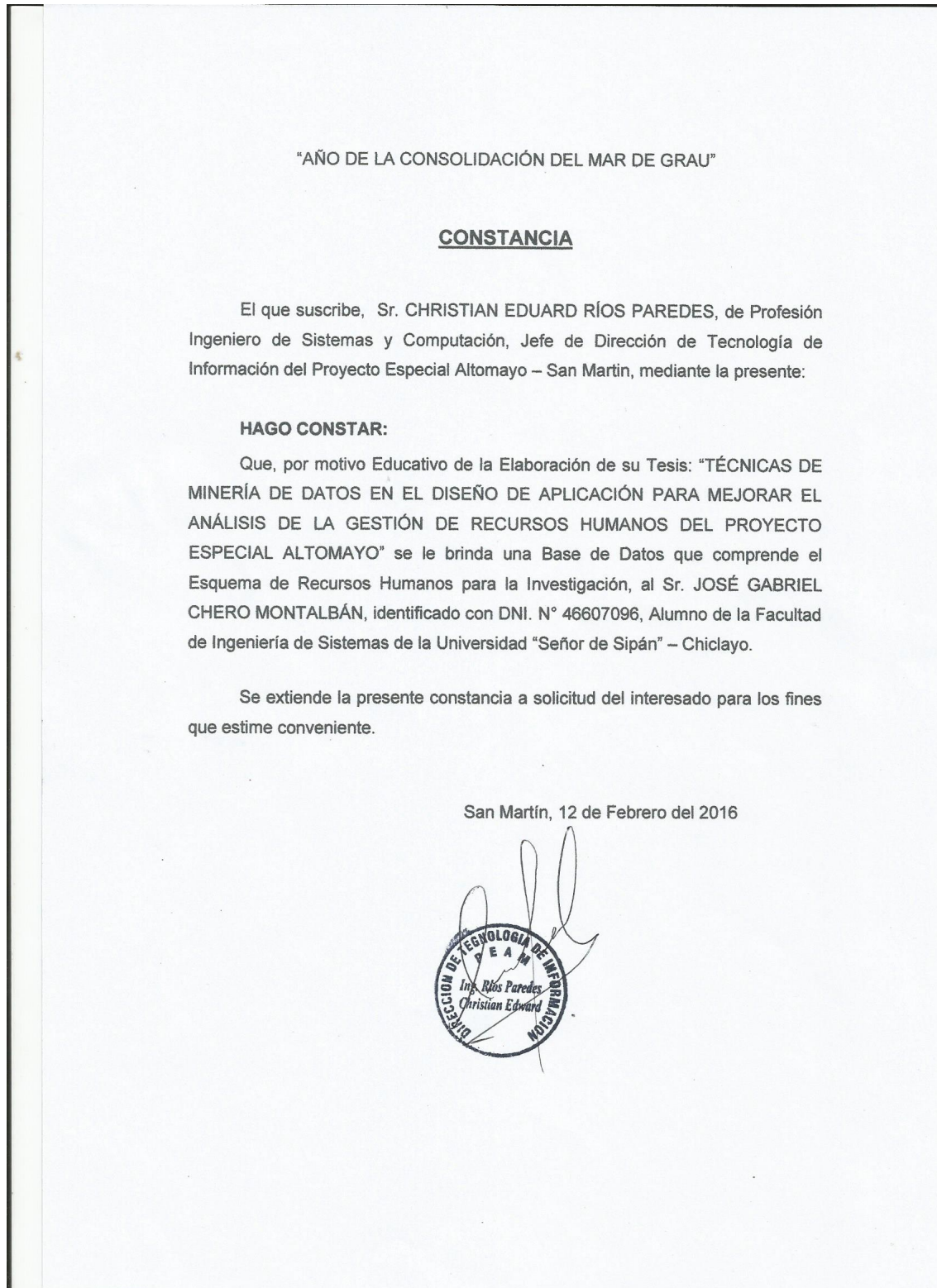


Figura 40: Solicitud para inicio del trabajo de investigacion
Fuente: Elaboración Propia



Anexo 4: Estructura Organica Proyecto Especial Altomayo

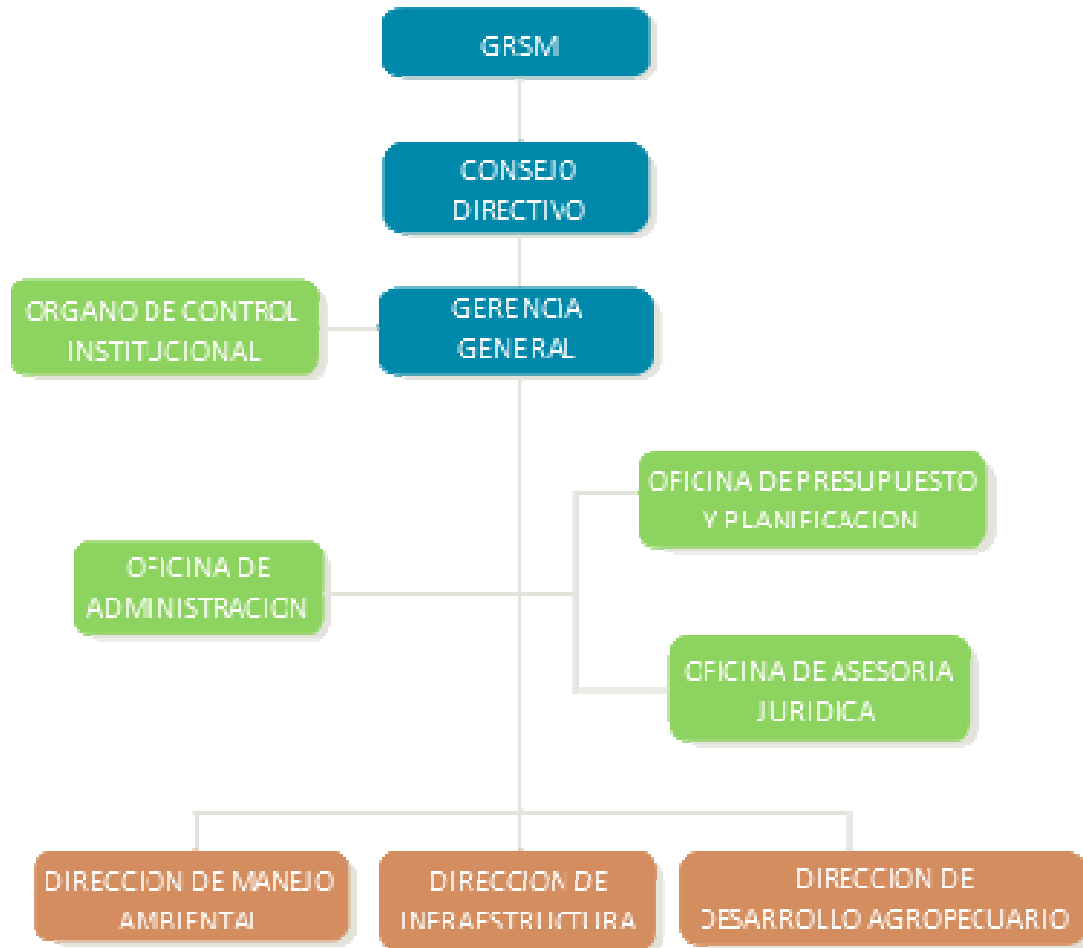


Figura 41: Estructura organica PEAM

Fuente: <http://www.peam.gob.pe/>

