



**FACULTAD DE INGENIERÍA, ARQUITECTURA Y
URBANISMO**

**ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS**

TESIS

**IMPLEMENTACIÓN DE UN SISTEMA BASADO EN
ALGORITMOS COMPUTACIONALES COMO AYUDA A
LOS PROCESOS PROFESIONALES EN EL
MINISTERIO DE TRANSPORTES Y
COMUNICACIONES - LIMA, 2018**

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

Autor:

Bach. Cubas Guevara Rubén Alberto

Asesor:

Dr. Ramos Moscol Mario Fernando

Línea de Investigación:

Infraestructura, Tecnología y Medio Ambiente

**Pimentel, Perú
2020**



DEDICATORIA

Gracias a mis padres por su apoyo incondicional para poder seguir con mi formación profesional y principalmente a Dios todopoderoso que nos guía por buen rumbo.

AGRADECIMIENTO

Mi especial agradecimiento a mis padres por ser los principales promotores de nuestros sueños, por confiar y creer en nuestras expectativas, por los consejos, valores y principios que nos han inculcado, A la Universidad Señor de Sipán, a sus docentes; por ser el lugar del conocimiento adquirido en estos años.



RESUMEN

El objetivo de la presente tesis es implementar un algoritmo computacional para ayudar a los procesos de selección profesional en el Ministerio de Transportes y Comunicaciones del Perú-MTC

La tarea clave fue conformar las relaciones que hay entre evaluador y postulante empleando modelos basados en contenido. Para ellos se representó las relaciones de manera gráfica que permitieron conjuntar esa representación mediante una red bayesiana, conformados por atributos y una variable clase.

Los procesos de selección de profesional en el Ministerio de Transporte y Comunicaciones-MTC fueron implementados en un sistema aplicando la metodología de desarrollo de basado en Prototipos

Se definió un algoritmo computacional, que permita evaluar a los postulantes de una convocatoria para selección de personal basado en resultados históricos de sus evaluaciones.

La implementación de un sistema utilizando el Algoritmo Bayesiano TAN en los procesos de selección tiene un promedio de precisión de 79.33% y exactitud de 70.33%. También se observó que el tiempo promedio en dichos procesos de selección disminuye con la implementación de un sistema utilizando el algoritmo bayesiana Tan de 124.17 minutos a 23.76 minutos.

Palabras clave. Red bayesiana, selección profesional, clasificador, proceso



ABSTRACT

The objective of this thesis is to implement a computational algorithm to help the professional selection process in the Ministry of Transportation and Communications of Peru-MTC

The key task was to shape the relationships between evaluator and applicant using content-based models. For them, the relationships were represented graphically that allowed to combine that representation through a Bayesian network, consisting of attributes and a class variable.

The professional selection processes in the Ministry of Transportation and Communications-MTC were implemented in a system applying the prototype-based development methodology

A computational algorithm was defined, which allows evaluating the applicants of a call for personnel selection based on historical results of their evaluations.

The implementation of a system using the Bayesian TAN Algorithm in the selection processes has an average of 79.33% accuracy and 70.33% accuracy. It was also observed that the average time in these selection processes decreases with the implementation of a system using the Bayesian algorithm Tan from 124.17 minutes to 23.76 minutes.

Keyword: Bayesian network, professional selection, classifier, process



INDICE

DEDICATORIA	2
AGRADECIMIENTO	3
RESUMEN	4
ABSTRACT.....	5
INDICE.....	6
INDICE DE TABLAS	9
INDICE DE ILUSTRACIONES	10
I. INTRODUCCION.....	13
1.1. Realidad problemática.....	13
1.2.1. La Realidad Problemática en el Ministerio de Transportes y Comunicaciones	15
1.2. Antecedentes de Estudio	16
1.3. Teorías relacionadas al tema	22
1.3.1. La selección de personal	22
1.3.3. Técnicas de Minería de datos.....	27
1.3.3.1. Árboles de decisión	27
1.3.3.2. Redes Neuronales	30
1.3.3.3. Técnicas estadísticas	32
1.3.3.4. Técnicas bayesianas	33
1.3.3.4.1. Teorema de Bayes	33
1.3.3.5. Algoritmos de Clasificación.....	39
1.3.3.5.1. Algoritmo K 2	40
1.3.3.5.2. Algoritmo C4.5.....	40
1.3.3.5.4. Algoritmo Naive Bayes.....	41
1.3.3.5.5. Algoritmo Naive Bayes TAN	42
1.3.4. Lenguajes de Programación	45
1.3.4.1. Php.....	46
1.3.4.2. JavaScript	46
1.3.4.3. Html.....	47



1.3.4.4.	Visual .NET.....	47
1.3.5.	Entorno de Programación	48
1.3.5.1.	Eclipse	48
1.3.5.2.	Netbeans	48
1.3.5.3.	Visual Studio.NET	49
1.3.6.	Metodologías De Desarrollo De Software	50
1.3.6.1.	RUP	50
1.3.6.2.	Metodología XP	52
1.3.6.3.	Metodología basada en prototipos	53
1.4.	Formulación del problema	54
1.5.	Justificación e Importancia de Estudio.....	54
1.6.	Hipótesis.....	54
1.7.	Objetivos	55
1.7.1.	Objetivos General	55
1.7.2.	Objetivos Específicos.....	55
II.	MATERIAL Y METODOS	57
2.1.	Tipo y diseño de investigación.....	57
2.1.1.	Tipo de investigación	57
2.1.2.	Diseño de investigación	57
2.1.3.	Población y muestra.....	57
2.2.	Variables, Operacionalizacion	58
2.2.1.	Variable Independiente	58
2.2.2.	Variable Dependiente.....	59
2.3.	Técnicas e instrumentos de recolección de datos, validez y confiabilidad.	60
2.4.	Procedimiento de análisis de datos.....	61
2.4.1.	Análisis estadístico e interpretación de los datos.....	61
2.5.	Aspectos Éticos	64
2.6.	Criterios de rigor científico	65
III.	RESULTADOS.....	68
3.1.	Resultados en tablas y figuras	68
3.2.	Discusión de resultados	78



3.3.	Aporte práctico.....	79
3.3.1.	Información General del Ministerio de Transportes y Comunicaciones.	79
3.3.2.	Visión del Ministerio de Transportes y Comunicaciones.	80
3.3.3.	Cuadro Comparativo de las Metodologías de Desarrollo de Software.....	81
3.3.4.	Metodología de Desarrollo de Software	83
3.3.4.1.	Metodología Basada en Prototipos.....	84
3.3.4.1.1.	Identificación de requerimientos básicos	84
3.3.4.1.2.	Diseño del prototipo inicial.....	87
3.3.4.1.3.	Implementación.....	116
3.3.4.1.4.	Pruebas y mejoras del Prototipo.....	130
IV.	CONCLUSIONES Y RECOMENDACIONES	147
4.1.	Conclusiones	147
4.2.	Recomendaciones.....	148
V.	REFERENCIAS	149
VI.	ANEXOS	153



INDICE DE TABLAS

Tabla 1 : <i>Variable Independiente</i>	58
Tabla 2: <i>Variable Dependiente</i>	59
Tabla 3: <i>Técnica e instrumentos de recolección de datos, validez y confiabilidad</i>	60
Tabla 4: <i>Matriz de Confusión</i>	62
Tabla 5: <i>Matriz de Confusión: Proceso de Selección 01</i>	70
Tabla 6 : <i>Matriz de Confusión: Proceso de Selección 02</i>	72
Tabla 7 : <i>Matriz de Confusión: Proceso de Selección 03</i>	74
Tabla 8: <i>Resultado Promedio de los Procesos de Selección</i>	76
Tabla 9: <i>Tiempo de las acciones a seguir en el Proceso de Selección</i>	77
Tabla 10: <i>Cuadro Comparativo de las Metodologías de Desarrollo de Software</i>	81
Tabla 11: <i>Requerimiento No Funcionales</i>	84
Tabla 12: <i>Requerimientos Funcionales</i>	85
Tabla 13: <i>Cuadro Comparativo de Algoritmo Computacionales</i>	92
Tabla 14: <i>Atributos y/o Variables empleado para el funcionamiento del Algoritmo TAN</i>	96
Tabla 15: <i>Variable Clase</i>	97
Tabla 16 : <i>Distribución de Probabilidad de la Variable Clase</i>	98
Tabla 17 : <i>Conteo de la Variable edad condicionada a la clase</i>	99
Tabla 18 : <i>Distribución de la Probabilidad de la Variable Edad condicionada a la Variable Clase</i>	99
Tabla 19 : <i>Conteo de la Variable Formación Académica condicionada a la Variable Clase</i>	99
Tabla 20 : <i>Distribución de la Variable Formación Académica condicionada a la variable clase.</i>	100
Tabla 21 : <i>Distribución de la probabilidad de la variable Edad y Formación Académica condicionada a la variable clase.</i>	101
Tabla 22 : <i>Distribución de la Variable Ultima Experiencia Laboral condicionada a la variable clase.</i>	104
Tabla 23 : <i>Cantidad de Información mutua de cada par de variables condicionada a la variable clase</i>	108
Tabla 24: <i>Ítems de la Evaluación Curricular</i>	123
Tabla 25: <i>Pruebas de Requerimientos</i>	131
Tabla 26: <i>Login del Encargado de Selección Profesional</i>	132
Tabla 27: <i>Creación y Visualización de Convocatorias</i>	133
Tabla 28: <i>Registro de Curriculum Vitae</i>	135
Tabla 29: <i>Registro de Evaluaciones Presenciales</i>	139
Tabla 30: <i>Generación de Red Bayesiana y Visualización de Resultados</i>	141



INDICE DE ILUSTRACIONES

Ilustración 1 : *Estructura de un árbol de decisión*29

Ilustración 2 : *Esquema de un Modelo Neuronal*.....32

Ilustración 3 : *Ejemplo de Teorema de Bayes*34

Ilustración 4 : *Red Bayesiana 1*36

Ilustración 5 : *Red Bayesiana 2*.....37

Ilustración 6 : *Algoritmo Naive Bayes*42

Ilustración 7: *Estructura Naive Bayes*42

Ilustración 8: *Estructura de Naive Bayes TAN*43

Ilustración 9: *Algoritmo Naive Bayes TAN*43

Ilustración 10: *Funcionalidades Implementadas*.69

Ilustración 11: *Resultado de Indicadores del Proceso de Selección 01*72

Ilustración 12: *Resultado de Indicadores del Proceso de Selección 02*74

Ilustración 13: *Resultado de los Indicadores del Proceso de Selección 03*76

Ilustración 14: *Tiempo de todas las acciones a seguir en el Proceso de Selección*77

Ilustración 15: *Flujograma del desarrollo del sistema basado en algoritmos computacionales* 83

Ilustración 16: *Prototipo Interfaz Inicial*87

Ilustración 17: *Prototipo Registro de Postulante*.....87

Ilustración 18: *Prototipo Login Administrador*88

Ilustración 19: *Prototipo de Visualización de Procesos o Convocatorias*.....89

Ilustración 20: *Prototipo de Visualización de los Datos Personales de Postulantes*89

Ilustración 21: *Prototipo de Visualización de Capacitaciones de Postulantes*90

Ilustración 22: *Prototipo de Visualización de Experiencias Laborales de Postulantes*90

Ilustración 23: *Prototipo de Visualización de Exámenes Psicotécnico de los Postulantes*91

Ilustración 24: *Prototipo de Visualización de Entrevistas Laborales de los Postulantes*91

Ilustración 25 : *Algoritmo Bayesiano Aumentado a Árbol*93

Ilustración 26 : *Ejemplo de Red Bayesiano aumentado a Árbol*.....94

Ilustración 27 : *Variables: Edad-Formación Académica*98

Ilustración 28 : *Grafo no dirigido completo cuyos nodos son las variables predictorias* 109

Ilustración 29 : *Árbol no dirigido de las variables predictorias*..... 110

Ilustración 30 : *Árbol dirigido con las variables predictorias*..... 110

Ilustración 31 : *Modelo Algoritmo TAN con las variables predictorias y con la variable clase* 111

Ilustración 32: *Arquitectura del Software* 116

Ilustración 33: *Código Fuente de la Conexion a la base de datos* 117

Ilustración 34: *Diseño de la Base de datos* 118

Ilustración 35: *Código Fuente del cálculo de las probabilidades* 119

Ilustración 36: *Interfaz de Pantalla Principal* 121

Ilustración 37: *Interfaz de Registro de Datos Personales del Postulante*..... 121

Ilustración 38: *Interfaz de Registro de Formación Académica y Cursos o Capacitaciones* 122

Ilustración 39: *Interfaz de Registro de Experiencia Laboral de los Postulantes*..... 123



Ilustración 40: <i>Interfaz de Ingreso al Sistema</i>	124
Ilustración 41: <i>Visualización de Convocatorias</i>	125
Ilustración 42: <i>Listado de Postulantes</i>	125
Ilustración 43: <i>Visualización de Datos Personales del Postulante</i>	126
Ilustración 44: <i>Visualización de Cursos y Capacitaciones del Postulante</i>	126
Ilustración 45: <i>Visualización de Experiencia Laboral del Postulante</i>	127
Ilustración 46: <i>Visualización de Examen Psicotécnico de los Postulantes</i>	127
Ilustración 47: <i>Visualización de Entrevista Personal de los Postulantes</i>	128
Ilustración 48: <i>Visualización de las Redes Bayesianas utilizando el Algoritmo TAN</i>	129
Ilustración 49: <i>Visualización de las Redes Bayesianas utilizando el Algoritmo TAN</i>	129
Ilustración 50: <i>Visualización de las Redes Bayesianas utilizando el Algoritmo TAN</i>	129
Ilustración 51: <i>Visualización de las Redes Bayesianas utilizando el Algoritmo TAN</i>	129



I INTRODUCCIÓN



I. INTRODUCCION

1.1. Realidad problemática

La supervivencia de una empresa se deberá a la forma de cómo se incorporara un nuevo personal para realizar actividades en decisiones muy importantes en esa. . Como vemos, el desempleo es uno de los factores importantes de nuestro país; pero en esta vez nos muestra una tasa de desempleo urbano con una evolución positiva en relación con otros países Latinoamericanos, lo cual fue señalado por la Organización Internacional del Trabajo (OIT); ya que en el 2005 la tasa de desempleo urbano del Perú alcanzó el 9.6 % y en el 2006 descendió a 8.5%; debido al crecimiento consecutivo del Producto Bruto Interno (El Comercio, 2008). Con eso deducimos que el problema que para disminuir las tasas de subempleo o su calidad deberá mejorar ciertas políticas en el mercado peruano. (Céspedes, 2006).

Las empresas peruanas, con respecto a la mano de obra juvenil, se puede decir que en un primer momento no necesariamente prefieren a los jóvenes, sino a trabajadores sin importar su edad; sin embargo, por existir ciertas ocupaciones de gran dinamismo, es ahí en donde los jóvenes son más adecuados; así vemos que el 57% de estas empresas no tienen ninguna dificultad en contratar jóvenes para alguna ocupación específica; sólo que se han encontrado inconvenientes como que entre los entrevistados, la gran mayoría no se desempeña bien, debido a deficiencias en la capacitación o formación previa. (Empleos para los jóvenes, 2006).

La OIT, reveló que a nivel mundial el desempleo aumento a 58 millones de personas. En el caso de Perú, 1 de cada 5 jóvenes de PET (Población en edad de trabajar) no realizan actividades algunas que no sean trabajar o estudiar; es así como de 1990 al 2007 la PEA ocupada por los jóvenes incrementó por demás de 200 mil trabajadores, entonces el trabajo se incrementó para ellos de 59.4% a 66%. La situación es que a los jóvenes les dan trabajos por temporadas es decir contratos entre 3 o 6 meses, que no quiere decir que no existan posibilidades de trabajos a los jóvenes.



En el caso del desempleo en la ciudad de Chiclayo, elaborado por la Encuesta de Hogares Especializada de Empleo 2007 del Ministerio de Trabajo y Promoción del Empleo (MINTRAE), la PEA es de 221 735 personas, de las cuales 205 666 personas están ocupada, teniendo como tasa de desempleo 7.2% , esto se debe también a la desocupación y subempleo (Banco Central de la Reserva, 2008), ya que se estimó que en el año 2000 la PEA regional sea de 336 902 personas (30.8% de la población total y el 45.2% de la población de 15 años a más).

La tasa del subempleo y los empleos informales sigue en aumento a pesar que la economía ha mostrado con tasas de crecimiento, esto se debe por los altos costos de despidos, mano de obra con baja productividad y la poca cobertura de los seguros de salud y pensiones; en el caso del departamento de Lambayeque los que se encuentran ocupados son el 4.1% , los productos por trabajador 1 404, el ingreso laboral 394, Empleos formales 13.0% y los Empleos sin contrato 83.9%, como vemos lo que prima en nuestro departamento son los empleos informales, esto se debe a que al haber excesivos costos en lo que se refiere al valor agregado que generan la creación de los empleos, no se puede satisfacer el salario del trabajador y el beneficio del empleador, ni tampoco los costos por contratación. (Moneda, 2008).

Las empresas peruanas con el propósito de darle importancia a los criterios de selección, piden que en la mayoría de sus evaluaciones se califiquen en escala de 1 a 10; los resultados arrojan que lo que más les importa a las empresas en un joven postulante las aptitudes generales de trabajo (8.4%), debido al hecho de que dichas empresas ven el conocimiento en el interior de ellas es específico, es decir que cualquier trabajador que ingrese a laborar pasará por un periodo de aprendizaje; es uno de los criterios más importantes, sin dejar de lado la comunicación oral (7.5%), el nivel educativo (7.0%) entre otros.

El reclutamiento y la selección del personal son los aspectos más críticos en la gestión de las empresas, esto se debe a la carencia del proceso formal de selección, lo cual ha conducido a muchas empresas el llevar esta tarea de un modo improvisado; según el



estudio sobre Benchmarking de gestión de recursos humanos 2008, elaborado por la consultora PriceWaterhouseCoopers (PWC), sobre una muestra de 150 empresas de nuestro país, se encontró que el 18% del personal de las empresas peruanas se renueva cada año, es decir una de cada cinco personas, esta proporción varía en función del sector económico donde compiten las empresas; y lo más contradictorio es que si bien el 76% de las actividades prioritarias de las empresas se dedican a los procesos de selección es decir al reclutamiento de personal, mientras a la retención de los talentos solo es 17% .

Los jóvenes también tienen una visión o expectativa sobre la empresa en donde les gustaría desarrollarse, es por ello que tienen otro modo de pensar, asumen el desafío y el compromiso de una manera diferente, mostrándose más seguros en sus decisiones, es por eso que sus elementos imprescindibles son el desarrollo profesional y aprendizaje (68.8%), existencia de un buen clima de trabajo (33.2%), y contar con un buen sueldo (20.8%); descontando la posibilidad de ascenso rápido (3.6%).

1.2.1. La Realidad Problemática en el Ministerio de Transportes y Comunicaciones

La selección profesional en la Institución obliga a tomar en cuenta aspectos como una delicada evaluación y cuidadosa apreciación del solicitante ello requiere seguir un procedimiento que permita determinar si el empleado está capacitado para ello.

El proceso de selección profesional en la Institución consta de las siguientes etapas relacionadas entre sí: Convocatoria, Selección, Suscripción y Registro del Contrato; los procesos proceso mismo de selección efectiva de personal impone tener en cuenta el rol que cumplen elementos denominados criterios.

Los criterios en la Institución carecen de un grado de confiabilidad, es decir no es bien segura. Lo que resulta que al tiempo la exactitud de los resultados de todos los procesos no se puedan confiar.

Los aspectos de los resultados de los procesos de selección son:



- a) El encargado de la oficina de evaluar y seleccionar al postulante, no tiene conocimiento acerca de los criterios sobre las bases de convocatoria, lo que ocasionaría una ineficiente selección.
- b) El tiempo en los procesos de selección es fundamental por motivo de transparencia, los cuales en la evaluación de CV se demora para elegir al postulante adecuado según su perfil.
- c) El bajo rendimiento de los servidores o personal es por la ineficiente selección lo que obligaría a la institución a cambiar personal de otras oficinas (rotacion) o realizar otros procesos de selección.

Mediante Decreto de Urgencia N° 005-2018, publicado en el Diario Oficial El Peruano, el gobierno peruano promulgo ciertas medidas económicas de reducir el gasto público , lo que trajo como consecuencia reducción de personal en la Oficina de Personal del MTC por lo siguiente en la demora de los resultados de las convocatorias de selección profesional.

1.2. Antecedentes de Estudio

(Beltran Pascual, Muñoz Martínez, & Muñoz Alamillos, 2014). En su Artículo “Redes bayesianas aplicadas a problemas de credit scoring. Una aplicación práctica”, estos autores en su artículo trataron como objetivo implementar un clasificador de préstamos bancarios por medio de un algoritmo bayesiano, a través de una base de datos de clientes donde se tenía registrado la decisión sobre si el cliente acepta o no el préstamo. El diseño de investigación fue descriptivo, donde se la población fue 1767 registros. Los cuales se valoraron 19 atributos de acuerdo a ciertos atributos los más importantes fueron el tipo de trabajo, su profesión, la calidad de vida, sus ingresos y gastos, etc. y concluyo que las redes bayesianas al utilizar de manera óptima sus variables dan buenos resultados en la precisión de un 78,6 % y en cuanto haya más registros ese porcentaje se va aumentando.



(Quepuy Perleche, 2016). En su tesis “Sistema de gestión y soporte de toma de decisiones basado en algoritmos de bayes y cluster para mejorar los procesos analíticos del área comercial de una empresa educativa”, este autor en su tesis tenía como objetivo mejorar el proceso de toma de decisión de una empresa con la implementación de algoritmos (Bayes y clúster) , fue de tipo aplicada ya que al realizar la implementación del algoritmo realizo una encuesta en al aérea usuaria conformada con una población de 6 personas donde concluyo que al implementar dichos algoritmos los procesos de la toma de decisiones por medio de reportes el tiempo tomado se disminuyó de 19 horas (1140 min) a media hora (60min) es decir la información se dio de manera rápida .

(Alarcon Jaimes, 2015). En su tesis “Optimización del clasificador Naive Bayes usando árbol de decisión C4.5”, el objetivo de su tesis fue optimizar el clasificador Naive Bayes (NB) usando el algoritmo del árbol de decisión C4.5 , fue de tipo experimental, su población fue un conjuntos de casos de paciente con problemas oftalmológicos (24 casos) donde el clasificador deberá indicarle y darle la probabilidad de que paciente tendrán que usar lentes o no usarlos , para esos utilizo ciertas características como por ejemplo su edad , su sufre de enrojecimiento o lagrimeo , etc. , al realizar la optimación del algoritmo concluyo que tuvo una mejora el clasificador NA del 66.85 % a un porcentaje de 83.7% de aciertos

(Lázaro Gonzales, 2015). En su tesis denominado “Propuesta de sistema experto para detección temprana de enfermedades neoplásicas en pacientes de la clínica san Bartolomé-huacho, 2014”, este autor propone crear un sistema experto con redes bayesianas para una temprana detección de enfermedades de los pacientes de mencionada clínica. Se realizó una investigación aplicada, teniendo en consideración la utilización del conocimiento pre existente como antecedente. Según la finalidad del estudio, es una investigación tecnológica porque permite descubrir conocimientos para resolver problemas usando la tecnología. El diseño es de tipo pre-experimental. La



población, objeto de investigación, estuvo constituida por Expertos 88 profesionales de la salud, de la clínica San Bartolomé se obtuvo como resultado un tamaño de muestra a 20 profesionales y por razones de precisión y estratégica del investigador, el estudio se realizara a la población entera de 22 médicos de la clínica, quienes serán los usuarios finales del Sistema Experto basado en Redes Bayesianas, médicos quienes realizarán la detección temprana de las enfermedades neoplásicas; al realizar este sistema el autor determinó el grado de relación entre la calidad del Sistema Experto basado en Redes Bayesianas y la calidad de la información requerida para la detección temprana de enfermedades neoplásicas en pacientes, obteniéndose un coeficiente $r=0.491$ de correlación positiva moderada y el $p\text{-valor}=0.02$ el cual indica que si hay correlación entre las variables, es decir no a entender que la aplicación o utilización de redes bayesianas aplicadas en la medicina es de gran importancia para el apoyo social ya que permite diagnosticar enfermedades a tiempo y poder combatirlas.

(Artica Chacon & Ocaña Sudario, 2015). Estos autores en su tesis “Aspecto Nutricional apoyado por un Modelo Bayesiano y su relación en el pronóstico de Morbilidad del Neonato de la gestante”, los autores propusieron determinar la manera en que el aspecto nutricional apoyado por un Modelo Bayesiano se relaciona con el Pronóstico de Morbilidad del Neonato de la gestante. El trabajo de investigación fue tipo descriptivo correlacional y predictivo, ya que buscaron la relación entre la morbilidad neonatal con el aspecto nutricional de la madre gestante. Se usó la Metodología de Sampieri, la población de esta tesis fue de 120 días de mediciones de las madres gestantes quedando como muestra en 27 días y concluyeron que cuando aplicaron la interfaz gráfica web a la madres apoyado con el modelo bayesiano pudieron medir la morbilidad del neonato comprobando su estado es medio con un 67% aprox.



(Tume Ayala, 2014). En su tesis “Desarrollo de inteligencia de negocios para analizar empresas financieras: cajas municipales”, tuvo como objetivo desarrollar inteligencia de negocios utilizando analysis service de sql server para analizar información estratégica de empresas financieras-cajas municipales. El diseño de investigación fue descriptiva y longitudinal. Se empleó el método científico como método general y como específicos al método analógico, el inferencial y prototipo de sistemas, utilizó el algoritmo árboles de decisión y Bayes Naive) La población estaba, trabajo con los datos de créditos de ahorros y financieros de las cajas municipales de ahorro y crédito del Perú, y concluyo que la minería de datos ha permitido realizar predicciones de las colocaciones en las entidades micro financieras cajas municipales. Así de los créditos otorgados se concluye que el 50.30% son utilizados para capital de trabajo, el 23.30% para consumo y el 6.4% para activo fijo, entre otros. Además se observó que el 96.7% d elos créditos colocados se han realizado con capital propio de la entidad micro financiera, el 1.6 % agro banco, el 0.9% banco de la nación y el 0.8% Cofide (En ambos casos se utilizó el algoritmo de bayes navie).

(Malca Bulnes, 2015). En su tesis “Modelo algorítmico para la clasificación de una hoja de planta en base a sus características de forma y textura”, el autor tuvo como objetivo mediante métodos de clasificación (arboles de decisión, redes neuronales y las redes bayesianas la obtención de un modelo algorítmico para la clasificación de una hoja de planta por medio de sus características de forma y textura. Las variables empleadas fueron 14 atributos (7 de forma y 6 de textura) además de la clase y concluyo que luego de la implementación de los diferentes modelos de clasificación, de realizo una comparación en base a los resultados: la precisión obteniéndose que el árbol de decisión obtuvo en 10 casos el 74.3% la red neuronal 80.95 % y la red bayesiana el 79.52%



(Botelbo, Soprani, Rodrigues, & Frizera, 2017). En su artículo “Nuevo Enfoque para la Clasificación de Señales EEG usando la Varianza de la Diferencia entre las Clases de un Clasificador Bayesiano”, tuvieron como objetivo analizar la intención de movimiento, la activación muscular e inicio de movimiento durante los movimientos de extensión de la rodilla. La población fueron Diez sujetos sanos (9 mujeres, 1 hombre; lado dominante derecho; $21,17 \pm 1,7$ años), participaron voluntariamente en este estudio. Este grupo fue escogido aleatoriamente de un grupo de voluntarios homogéneo en términos de género entre las edades de 18 a 25 años. Los criterios de exclusión fueron estar por fuera de este rango de edad y personas con problemas motores. Se empleó también un sistema de sensores inerciales TechMCSR (Technaid S. L., España). Este sistema proporciona una orientación espacial de los sensores en tiempo real. En los experimentos se emplearon dos sensores, uno colocado en el muslo y otro sobre la pantorrilla del sujeto. El uso de sensores inerciales o unidades de medición inercial (IMUs por sus siglas en inglés) permite validar el inicio del movimiento sin la necesidad de ambientes estructurados. Los resultados muestran que puede ser viable trabajar con señales SMR de EEG para predecir la intención de movimiento, incluso trabajando con los miembros inferiores. Por lo tanto, el análisis SMR puede ser una alternativa para la predicción de movimiento incluso en comparación con el uso de MRCP (Jiang et al., 2015) cuando es necesario un mejor tiempo de anticipación. Estos resultados pueden ser usados en el control de dispositivos robóticos para apoyar las tareas de rehabilitación con un adecuado nivel de detección sin sobrecargar el sistema de control. El método de clasificación de EEG se llevó a cabo basado en la frecuencia y el tiempo con un clasificador bayesiano y posteriormente con un clasificador de umbral con la varianza de la diferencia de estos discriminantes.

(Oviedo Balladares, 2017). En su tesis Clasificador Bayesiano basado en Distribución Normal y su aplicación en la predicción de tipo de Ictericia Neonatal, tuvo como



objetivo determinar los factores que influyen en la clasificación del tipo ictericia neonatal en el servicio de neonatología de EsSalud Cusco. La metodología empleada es cuantitativa, sustantiva básica y de diseño descriptivo correlacional, de tipo transversal porque se dió en un momento dado, la población es de 680 neonatos y como muestra a 246 recién nacidos con ictericia en el año 2015. Se aplicó el modelo de clasificación Bayesiana basado en la distribución normal, para predecir el tipo de ictericia neonatal, siendo este clasificador de tipo lineal. se obtuvo que el modelo de clasificación bayesiana basado en la distribución normal que mejor predice el tipo de ictericia neonatal en el servicio de neonatología de EsSalud Cusco, es de tipo lineal tanto para fisiología y patológica y los factores que influyen en la clasificación del tipo de ictericia neonatal en el servicio de neonatología de EsSalud Cusco, son antecedentes y patologías familiares. También se concluyó que la bondad del modelo de clasificación de ictericia neonatal basado en la distribución normal es de 65%

(Cordova Roque, 2017). En su tesis “Análisis Predictivo de Muerte y Sobrevida de Pacientes Hospitalizados mediante Clasificadores Supervisados”, el autor en su investigación tuvo como objetivo predecir la muerte o sobrevida de los pacientes hospitalizados a través de una clasificación de sus datos, la metodología que utilizo en KDD(Proceso de extracción de conocimiento) , en cual tuvo como población y muestra 24 500 registro de pacientes del Hospital Regional Honorio Delgado que cuentan con 26 atributos como datos de los pacientes y diagnósticos de diversas enfermedades ; y concluyo que el planteamiento de técnicas de clasificadores supervisados un 99 % de acierto indico que se puede pronosticar el fallecimiento y la vida de los pacientes .

(Ramon Campoverde, 2016). En su tesis “Desarrollo de un sistema web utilizando Redes Bayesianas para enseñanza del idioma inglés en la escuela Zoila Alvarado de Jaramillo”, la autora como propuso crear un sistema web para poder enseñar el idioma ingles a niños de 5 a 6 años, utilizando redes bayesianas. Para ello se empleó la técnica



de los Test Adaptativos Informatizados, seleccionando las Redes Bayesianas como estructura para representar el modelo del alumno donde los niños puedan visualizar temas, preguntas y conceptos y reconocer la voz. La metodología de desarrollo utilizada fue UWE (UML-Based Web Engineering), la autora empleó métodos de investigación: el inductivo y analítico sintético y concluyó que la aplicación web SEEII ayudó a que los estudiantes desarrollen las habilidades para escuchar y pronunciar las palabras evaluadas.

1.3. Teorías relacionadas al tema

1.3.1. La selección de personal

La selección de personal en un proceso donde de una u otras manera están involucradas, el cual es llevado por un personal especialista en el tema donde seleccionan trabajadores para el ejercicio de ciertas funciones o cargos, por ese motivo es uno de los principales procesos de una entidad sea pública o privada. (Canos, 2008).

Al vincular un personal a una entidad, empresa influye en el ámbito organizacional y las relaciones interpersonales por ser parte de un equipo de trabajo.

Las empresas o entidades para evitar pérdidas en el ámbito logístico y tiempo obligan al especialista a realizar una eficiente selección de su personal.

Todo proceso debe iniciar con una solicitud para poder ocupar un cargo en una vacante (Canos, 2008)

Objetivo:

- Las entidades deberán realizar un adecuado trabajo para elegir al postulante más adecuado a través de la obtención de su información.
- El proceso debe ser lo mayor eficiente posible para elegir a los candidatos más aptos.
- Para poder conocer al postulante con el mejor perfil de puesto se deberá minimizar el tiempo que se utiliza el proceso.



Fases del proceso de selección de personal.

La fase del proceso de selección debería estar en función de variables para cada caso, es decir que el proceso debe ser diferente si se trata de personal ejecutivo, de confianza o personal operativos.

Las fases de los procesos están en función a instrumentos que son utilizados a las necesidades de ellos, los cuales podrán añadirse u omitirse, eso dependerá de la creatividad o inteligencia de la responsable del encargo del proceso de selección (Atalaya Pisco, 2001).

Según (Atalaya Pisco, 2001), los indica los siguientes fases o etapas del proceso de selección de personal.

a) Planeamiento del proceso de Selección de Personal.

Según el autor nos indica que es etapa muy importante donde se elabora los objetivos, planes, elaboración de cronogramas de trabajo hasta el presupuesto es decir se determina como se realizara el proceso.

Es decir aquí se deberá elaborar el conocido “Perfil de puesto” que contiene la información relevante respecto al puesto, las funciones o tareas, responsabilidades que deberá conocer el postulante.

b) Reclutamiento

En esta etapa comienza con la publicación del aviso de las vacantes para el concurso para la recepción de los postulantes y verificación curricular, previos los postulantes deberán verificar y analizar los requisitos mínimos del perfil de puesto; es una etapa en buena ya que permite conocer la materia prima es decir los más aptos al puesto, puede ser interno o externo.

c) Evaluación

Al realizar el reclutamiento del proceso de selección se debe aplicar ciertos instrumentos de evaluación a los postulantes, que pueden consistir en una evaluación psicológica, de conocimientos y entrevista personal; es muy importante



que dichos instrumentos guarden relación con lo solicitado en los perfiles de puesto. Mediante la observación se ve su desenvolvimiento de las competencias requeridas lo que hace que la entrevista sea un instrumento muy importante en los procesos de selección.

d) Selección

Al término de evaluación de los postulantes se realiza la verificación de todo el referente a ellos por ejemplo sus datos, documentos, su entrevista realizada. Cuyo instrumento o técnica importante de todo proceso el cual nos permite encontrar vacíos en su currículum el cual nos permite conocer sus aspiraciones, plan de vida la adaptación al puesto de trabajo, es decir nos permite apreciar los intereses de la persona, sus valores y costumbres en cual es fundamental.

e) Decisión

De acuerdo al análisis de los informes de los postulantes se debe determinar a los finalistas de cada vacante o puesto de trabajo que pueden ser dos o más, pero dependerá la decisión, sobre a quién contratar, a la comisión que lleva a cabo el proceso donde le menciona que se presente al jefe en un día y hora determinado según las bases del concurso. Más adelante se deberá realizar los exámenes medios y firma de contrato.

f) Inducción

Al realizarse la incorporación a la entidad o empresa, el nuevo trabajador debe familiarizarse con el puesto, compañeros de áreas, al ambiente físico y obviamente con su jefe inmediato, y saber cuáles son las políticas y normas de la empresa o entidad que eso asegura el éxito del nuevo trabajador ya que puede potenciar sus conocimiento y plasmarla en sus actividades cotidianas.



g) Control y Seguimiento del Proceso de Selección

En esta etapa del procedimiento de selección de personal se verifica si realmente se realizó una eficiente selección, porque si el postulante con más alto puntaje al realizarle un control y seguimiento de sus funciones y/o actividades también obtiene un trabajo eficiente en su puesto, pues se concluye que ambos factores están correlacionados y el proceso va bien. Lo más recomendable sería que esto se debe realizar constantemente de por lo menos cada 3 meses al nuevo empleado.

Protagonistas de un Proceso de Selección de Personal

Según (Atalaya Pisco, 2001) , los protagonistas del proceso de selección de personal son básicamente:

a) El área usuaria que solicita uno o más trabajadores.

Este protagonista inicia el proceso con la determinación del puesto que desea cubrir, es decir en el área donde establece el perfil de puesto donde contiene las competencias adecuadas que debe tener cierta persona para que cubra la posición en base a los conocimientos , aptitudes que requiera .

Este actor o protagonista deberá elaborar junto con el área encargada de recursos humanos de la entidad.

b) Recursos Humanos

Este protagonista es la responsable de realizar el proceso de selección y dar como resultado al postulante más adecuado de acuerdo al perfil de puesto; es decir aprueba o desaprueba.

Previo a eso, tiene como función revisar los perfiles de puestos

c) Postulantes

Son las personas que desean ocupar el cargo o puesto de la empresa o entidad que convoca un proceso de selección, que al pasar el proceso de evaluación otorgara tu tiempo en las funciones dispuestas por esa.



Los 3 actores o protagonista se deberán interrelacionar, tomando decisiones, evaluando el proceso

1.3.2. Minería de Datos

Según (Hernandez Orallo, Ramirez Quintana, & Ferri Ramirez, 2004), nos dicen “Es el proceso de extraer el conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos”, es decir al aprovechar la gran cantidad de información almacenados en grandes bases de datos nos ayudaría a realizar tareas que sean aprovechadas por las empresas para que puedan tener las herramientas necesarias para lograr sus objetivos . Para Larrieta & Santillán (2007), “La minería de datos es el proceso que tiene como propósito descubrir, extraer y almacenar la información relevante de amplias bases de datos, a través de programas de búsqueda e identificación de patrones”, esto ayudaría a las grandes empresas a tomar mejores decisiones mediante una información clara y precisa para que puedan lograr un mejoramiento en su productividad. Además añade Kopanmakis & Theodoluidis (2003) “Es el proceso de descubrimiento de conocimiento sobre repositorio de datos complejos mediante la extracción oculta y potencialmente útil en forma de patrones globales y relaciones estructurales implícitas entre los datos”, que al extraer información de un conjuntos de datos almacenados pueda transformarla hacia una estructura que pueda ser usada utilizando métodos de inteligencia artificial, estadística y sistemas de base de datos. es así como vemos queda en claro que la Minería de datos tiene por objetivo fundamental, como lo dice (Hernandez Orallo, Ramirez Quintana, & Ferri Ramirez, 2004) “El encontrar modelos inteligibles a partir de los datos”, es decir utilizar a los datos como los elementos fundamentales de los cuales a través de las distintas técnicas que esta presenta, se pueda analizar y extraer el conocimiento novedoso y útil, es así como ese conocimiento deberá ser lo más



comprensible posible, ya que el usuario final, el cual tendrá acceso, no es una persona experta en las técnicas de minería de datos.

Ya teniendo conociendo lo que significa Minería de datos, nos toca revisar y analizar las tareas y técnicas que presenta y que la hacen ser importante.

1.3.3. Técnicas de Minería de datos

Como podemos ver con el pasar de los años se han creado varios tipos de investigaciones sobre Minerías de Datos, en el que se pensó en desarrollar diversas técnicas, las cuales pueden ser utilizadas para categorizar métodos de sistemas de minería de datos, ya sea a través de los tipos de base de datos o el tipo de conocimiento que se desea extraer, así como técnicas que se pueden aplicar en el proceso, estas técnicas de minería de datos se originan a partir de la inteligencia artificial y de la estadística las cual veremos una a una a continuación:

1.3.3.1. Árboles de decisión

Según (Hernandez Orallo, Ramirez Quintana, & Ferri Ramirez, 2004) nos dice que “Los árboles de decisión son una serie de decisiones o condiciones organizadas en forma jerárquica, a modo de árbol, las cuales son muy útiles para encontrar estructuras en espacios de alta dimensionalidad y en problemas que incluyen datos categóricos y numéricos”, al ser una herramienta de gran utilidad que permite a un individuo , a las empresas u organizaciones intercambiar ideas y acciones en el que puedan elegir la mejor alternativa según sus beneficios , es decir la más acertada para las soluciones de sus problemas. Además, añade Han & Kamber (2000) “Un árbol de decisión es un diagrama de flujo, con estructura de árbol, en donde los nodos internos representan validaciones sobre los atributos, las ramas representan las salidas de las validaciones y los nodos hojas representan las clases”. Esta técnica es usada para las tareas de regresión y de clasificación, ya que se puede tener que cada nodo es un atributo que llega a otro nodo ya que para clasificarla el flujo



sigue hacia debajo de acuerdo a la variable clase que se le asigna a dicho nodo

El árbol de decisión es una técnica enmarcada dentro del desarrollo de métodos y sistemas de razonamiento utilizados en investigaciones de inteligencia artificial y programación de aplicaciones, por su estructura son fáciles de comprender y analizar; su utilización cotidiana se puede dar en diagnósticos médicos, predicciones meteorológicas, controles de calidad, y otros problemas que necesiten de análisis de datos y toma de decisiones. (Calavancha Zuñiga).

Lo que no dice el autor, es que al ser una técnica aplicada a la inteligencia artificial o estadísticas puedes usarse para analizar casos de la vida cotidiana y analizar datos en diagnósticos para predecir una enfermedad o en casos meteorológicos, ante un desastre natural por lo que en estos casos se tiene en mayoría grandes cantidades de datos que es lo que requiere un árbol de decisión.

En la ciencia económica, se puede utilizar para la predicción de resultados de diversas políticas o estrategias a seguir, en las finanzas se puede utilizar para el análisis de riesgo crediticio, decisiones de inversión o decisiones de gestión financiera. (Calavancha Zuñiga)

Es decir en el campo financiero las empresas u organizaciones pueden pronosticar mediante probabilidades diversas alternativas de resultados económicos; es decir en cuanto a beneficios o pérdidas en sus estados económicos y consideren los riesgos.

El Árbol de Decisiones se basa en la aplicación de un conjunto de reglas SI-ENTONCES, utiliza funciones lógicas que nos llevarán a disyunciones de posibles resultados (SI compro acciones de una minera ENTONCES su precio puede subir O bajar) (Calavancha Zuñiga)



Los arboles de decisiones cuentas tiene la facilidad de ayudar a los individuos o empresas a exponer sus problemas de manera clara y sus posibles consecuencias de sus decisiones, es por eso que las probabilidades que se obtienes les ayuden a tomar adecuadamente sus decisiones a través de diversos esquemas.

Elaboración de un Árbol de Decisiones

Para construir un árbol de decisión, se debe tener en cuenta los siguientes elementos:

El Nodo de decisión, que se representa con un cuadrado, llamado punto de decisión o punto de arranque donde se fija la acción.

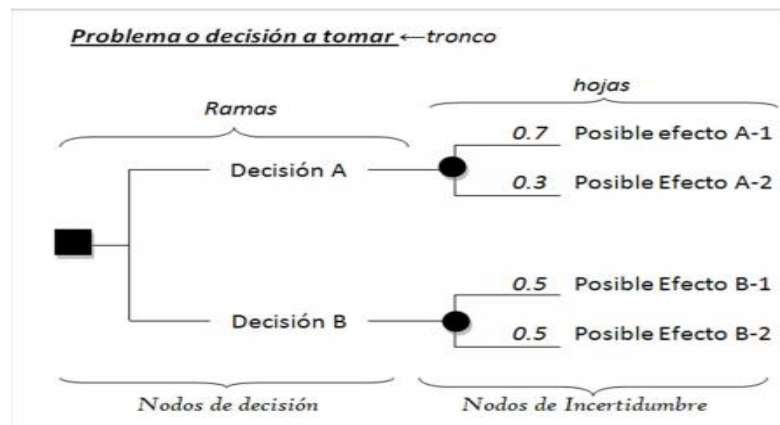
Línea o conectores, que se diseñan de izquierda a derecha que unen elementos mediante una secuencia conlleva al nodo de azar o probabilístico.

Nodo de Azar, se representan mediante círculos que está formado por los diferentes posibles efectos de cada decisión.

Nodo Terminal, se muestra en triángulos que son los desenlaces finales

En el siguiente gráfico comprenderemos la estructura del árbol de decisiones.

Ilustración 1 : Estructura de un árbol de decisión



Fuente: (Calavancha Zuñiga, pág. 5)

En el gráfico anterior, podemos observar dos partes en la elaboración de un árbol de decisión, donde que el nodo de decisión y los conectores o línea que parten de las llamadas ramas que conforman los nodos de decisión pueden ver más ramas o sub ramas y llaman hojas a los posibilidades de cada decisión, donde comienza con un figura geométrico, un cuadrado y deberá finalizar en un círculo. Los nodos de azar o incertidumbre donde están los posibles efectos de cada decisión, esta parte empieza con un círculo.

Para evaluar el árbol de decisión la empresa o entidad deberá elegir la mejor opción con el mejor valor de los posibles resultados los valor pueden ser en porcentajes que deberá llevar a 100% o 1.

Nos muestran una visión gráfica, además de evaluar las variables de las acciones a ser efectuada, es decir permite analizar las posibles consecuencias de tomar una decisión.

1.3.3.2. Redes Neuronales

Según (Hernandez Orallo, Ramirez Quintana, & Ferri Ramirez, 2004) “las redes neuronales, permiten a los problemas complejos modelizar en los que puede haber interacciones entre variables no lineales”. Como en el caso de los árboles de decisión, las redes neuronales pueden usarse en problemas de clasificación, de regresión y de agrupamiento, este tipo de técnica trabaja directamente con datos numéricos y requieren gran cantidad de datos para poder ponerlo en entrenamiento y también capacidad para la generalización de problemas no lineales, su desventaja es que es difícil de comprenderlo.



Existen varias definiciones de lo que es un red neuronal artificial, teniendo un como por ejemplo.

Una Red Neuronal es un modelo computacional, paralelo, compuesto de unidades procesadoras adaptivas con una alta interconexión entre ellas (Hassoun, 1995), al ser un modelo computacional con recursos que permite estudiar sistemas complejos, simulándolos por medio de una computadora.

Algunos ejemplos de modelados computacionales pueden ser: pronosticar el clima, construir aeronaves o aviones, estudiar fenómenos de la naturaleza (terremotos, maremotos).

Además también se pueden emplear en la medicina, los modelos computacionales ayudarían en crear instrumentos de decisión en los médicos, con un conjunto de datos de análisis de pacientes de diversas enfermedades o ayudar a modelar o diseñar medicinas y fármacos para obtener un medicamento eficaz.

Modelos Neuronales

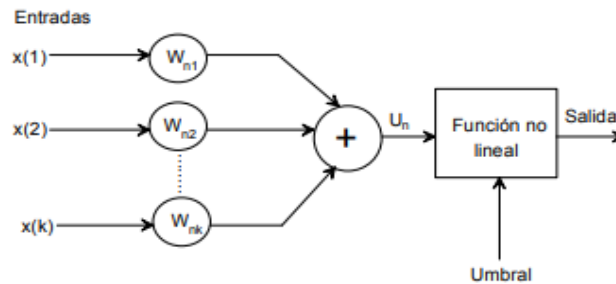
Según (Serrano, Antonio, Soria, Emilio, & D. Martin, 2010) , en una red neuronal artificial se tienen cuatro elementos básicos. , que son:

- a) Entradas; que son variables o conjuntos de conexiones o señales, que contiene un peso (determinan la intensidad de la señal de entrada) correspondiente a cada una de las entradas, es decir es aquí donde se almacena la información que vienen de otras neuronas.
- b) Un sumador o función de agregación , es decir se encarga de sumar ponderadamente las diversas entradas con los pesos sinápticos
- c) Una función de activación o de transferencia, donde el valor obtenido con la función de agregación, se desplaza por un función de activación y da la salida de la neurona como por ejemplo la función escalón, sigmoidea, etc.



d) Un umbral exterior, que determina la activación de la neurona.

Ilustración 2 : Esquema de un Modelo Neuronal



Fuente: (Serrano, Antonio, Soria, Emilio, & D. Martin, 2010)

Matemáticamente las operaciones a realizar serian

La función no lineal. Normalmente se asocia el umbral U_n mediante una entrada (que vale -1) y un peso adicional asociado. Es decir

$$\text{umbral} = -W_{n0} \Rightarrow \begin{cases} U_n = \sum_{j=0}^k W_{nj} \cdot x(j) \\ x(0) = 1 \end{cases} \Rightarrow \text{salida} = \rho(U_n)$$

$$\text{salida} = \rho(U_n - \text{umbral})$$

1.3.3.3. Técnicas estadísticas

Según Bonilla & Ojeda (2006), nos dicen que las técnicas estadísticas desarrollan modelos cuya función es mejorar la comprensión los escenarios viales en relación con los datos y de esa forma contar con información valiosa que te permita poder tomar una decisión, pueden expresarse en fórmulas matemáticas algebraicas como el nivel de confianza, población, el tamaño de la muestra, estimación de una media, muestreo, etc.



Las técnicas estadísticas son aplicables en diferentes ramas como la medicina, contabilidad, ingeniería, etc., como en actividades de educación, en las políticas y en todo donde se debe tomar decisiones.

1.3.3.4. Técnicas bayesianas

Según Bonilla & Ojeda (2006), nos dicen que este tipo de técnica nos permite calcular la probabilidad, donde pueden usarse en problemas descriptivos en relaciones de dependencia e independencia o predictiva para usarse como métodos de clasificación, unos de los clasificadores más usados es el Teorema de Bayes, algunos de los algoritmos más usados son el clasificador Bayesiano Naive, los métodos basados en máxima verisimilitud.

1.3.3.4.1. Teorema de Bayes

El Teorema de Bayes, “es la regla básica para realizar inferencias. Nos permite evaluar la creencia que tenemos de un suceso o conjunto de sucesos a la luz de nuevos datos u observaciones, es decir, nos permite pasar de una probabilidad a priori P (suceso) a la probabilidad a P (suceso | observaciones” (Bonilla Gordillo & Ojeda Schuldt, 2006, pág. 28).

$$P(h/O) = \frac{P(O/h) \cdot P(h)}{P(O)}$$

En la fórmula anteriormente mencionado. Podemos visualizar como calcular la probabilidad condicional a posteriori $P(h/O)$, donde $P(O/h)$ es la probabilidad de O en la hipótesis h , y $P(h)$, $P(O)$ es la probabilidad a priori, es decir el teorema de Bayes hace uso de las probabilidades a priori subjetivas.

Ejemplo.



La probabilidad de que suene una alarma si se ha ocurrido algún accidente es de 97% y de que suene la alarma si no ha ocurrido ningún accidente es de 2% y de que haya un accidente es de 1%. ¿Cuál es la probabilidad de no haya ocurrido ningún accidente si es que la alarma haya funcionado

I = Producir un incidente.

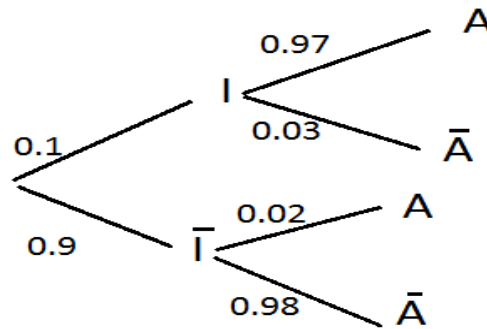
\bar{I} = No producir incidente

A = Sonar la alarma.

\bar{A} = No sonar la alarma

Aplicamos el teorema de Bayes.

Ilustración 3 : Ejemplo de Teorema de Bayes



Fuente: Elaboración Propia

$$P(\bar{I}/A) = \frac{P(A/\bar{I}) \cdot P(\bar{I})}{P(A)}$$

$$P(\bar{I}/A) = \frac{0.02 * 0.9}{0.1 * 0.97 + 0.9 * 0.02}$$

$$P(\bar{I}/A) = 0.157$$



Entonces la probabilidad que no haya habido ningún incidente dado que haya funcionado la alarma es 0.157.

Definición de las Redes Bayesianas

Una red bayesiana, “es un grafo dirigido y acíclico cuyos nodos representan variables y cuyos arcos representan relaciones de dependencia entre sus variables. Si existe un arco desde un nodo A hacia algún otro nodo B”. (Bonilla Gordillo & Ojeda Schuldt, 2006, pág. 36).

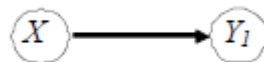
Las redes bayesianas, es un modelo probabilístico con un conjunto de variables aleatorias que se relacionan a través de un gráfico dirigido, es decir nos dan una visualización gráfica de cómo las variables se relacionan entre ellas

Las redes bayesianas se encuentran entre los modelos gráficos más populares. La principal diferencia, con respecto a otros modelos, está en que sus arcos son dirigidos y representan dependencia condicional entre las variables. El nombre proviene del hecho que gran parte de la teoría relevante con este tipo de redes se basa en la estadística bayesiana. (Santesteban Rojas, Utría Pérez, & Hernández Reyes, 2012, pág. 4)

Mediante el grafo podemos darnos cuenta acerca de que las variables pueden estar relacionadas mediante una independencia condicional de una variable o grupos de variables dada otras variables para un razonamiento probabilístico, donde los nodos (representan las variables aleatorias proposicionales) y arcos (la dependencia de probabilidad), el nodo o variable donde apunta el arco es dependiente (causa y efecto).



Cuando un arco une a un nodo X con un nodo Y, el nodo X se llama padre de Y, y el nodo Y se llama hijo de X.



Con la siguiente notación, según el grafico donde Y₁ condicionalmente dependiente de X = P (Y₁/X) y cuando las dos variables aleatorias X e Y₁ son independientes si:

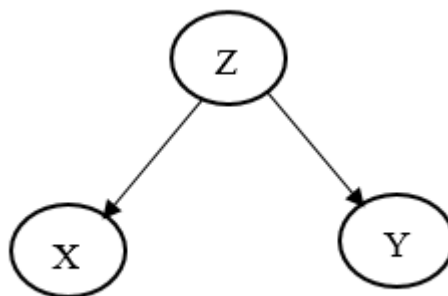
$$P(X, Y_1) = P(X) P(Y_1)$$

$$P(X/ Y_1)=P(X)$$

$$P (Y_1/X)=P(Y_1)$$

Veamos el siguiente gráfico:

Ilustración 4 : Red Bayesiana 1



Fuente: Elaboración Propia

Las condiciones equivalentes dadas las tres variables pueden son:

$$P(X, Y / Z) = P(X/Z) P(Y/Z)$$

Ejemplo

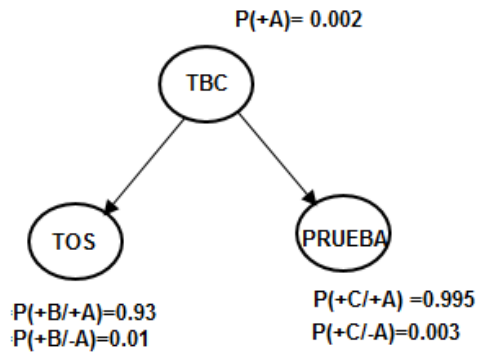


Dada una red bayesiana formado por tres variables denominado Enfermedad TBC, un síntoma que es la tos y la prueba.

Denotamos las variables TBC: nodo A, TOS nodo B y Prueba nodo C
 Cuando A es binaria, lo denotamos por un +A, es decir “el paciente tiene TBC y -A es paciente no tiene TBC; en el Nodo B cuando lo denotamos como +B en paciente presenta tos y cuando es -B no presenta tos, en el Nodo C cuando lo denotamos con +C significara el resultado de la prueba positivo y un -C un resultado negativo de la prueba.

Los datos que debemos saber son:

Ilustración 5 : Red Bayesiana 2



Fuente: Elaboración Propia

$P (+A) = 0.002$, indica a priori que un 0.02 % de los pobladores tienen TBC.

$P (+B/+A)=0.93$, indica que cuando hay TBC, los pacientes presentan tos en un 93% de los casos.

$P (+B/-A)=0.01$, indica que cuando no hay TBC, los paciente presentan tos en un 1 % de los casos. Y no presentan en un 99.99 %

$P(+C/+A) = 0.995$, indica que cuando hay TBC , la prueba a los pacientes es positivo en un 99.5 % de los casos.



$P (+C/-A) = 0.003$, indica que cuando no hay TBC, la prueba de los pacientes es positivo en un 0.03 % y negativo en un 99.97 %

Al tener estos datos, podemos calcular las siguientes probabilidades.

a) La probabilidad de C (Prueba) y B (Tos)

$$P (+C) = P (+C/+A) P (+A) + P (+C/-A) P (-A)$$

$$P (+C) = 0.995 * 0.002 + 0.003 * 0.998$$

$$P (+C) = 0.00199 + 0.002994$$

$$P (+C) = 0.004984$$

$$P (-C) = P (-C/+A) P (+A) + P (-C/-A) P (-A)$$

$$P (-C) = 0.005 * 0.002 + 0.997 * 0.998$$

$$P (-C) = 0.00001 + 0.995006$$

$$P (-C) = 0.995016$$

$$P (+B) = P (+B/+A) P (+A) + P (+B/-A) P (-A)$$

$$P (+B) = 0.93 * 0.002 + 0.01 * 0.998$$

$$P (+B) = 0.00186 + 0.00998$$

$$P (+B) = 0.01184$$

$$P (-B) = P (-B/+A) P (+A) + P (-B/-A) P (-A)$$

$$P (-B) = 0.07 * 0.002 + 0.99 * 0.998$$

$$P (-B) = 0.00014 + 0.98802$$

$$P (-B) = 0.98816$$

b) En el caso que la prueba haya dado positivo. ¿Qué probabilidad ahí que la persona sufra de TBC?



Ha veces las pruebas no tiene un fiabilidad absoluta, es decir existe la posibilidad de haber un falso positivo.

$$P(+A/+C) = \frac{P(+A)P(+C/+A)}{P(+C)} = \frac{0.002 \cdot 0.995}{0.004984} = \frac{0.00199}{0.004984} = 0.399$$

Es decir, de acuerdo al resultado de la prueba, hay un 39.9 % de probabilidad de que una persona sufra de TBC.

De igual manera podemos calcular:

$$P(-A/+C) = \frac{P(-A)P(+C/-A)}{P(+C)} = \frac{0.998 \cdot 0.003}{0.004984} = \frac{0.002994}{0.004984} = 0.6007$$

- c) En el con la siguiente variable que es Tos podemos calcular la probabilidad de que la persona tenga TBC, sabiendo que presenta tos.

$$P(+A/+B) = \frac{P(+A)P(+B/+A)}{P(+B)} = \frac{0.002 \cdot 0.93}{0.01184} = \frac{0.00186}{0.01184} = 0.1570$$

$$P(-A/+B) = \frac{P(-A)P(+B/-A)}{P(+B)} = \frac{0.998 \cdot 0.01}{0.01184} = \frac{0.00998}{0.01184} = 0.8429$$

Es decir observamos que solo hay un 0.157 de probabilidad de que tenga TBC.

1.3.3.5. Algoritmos de Clasificación

Los algoritmos de clasificación “encuentran patrones en los datos que le damos y los clasifica en grupo, luego comprar nuevos datos y los ubica en uno de ellos y se puede predecir de que se trata”. (Sandoval, 2018, pág. 37)

La clasificación como una técnica de la Minería de Datos permite saber en qué clase pertenece un objetivo o individuo, es decir a través de un conjunto de variables o atributos nos permite obtener a que clase pertenece un atributo o variable. Los algoritmos de clasificación es un tipo de aprendizaje supervisado ya que permite realizar analizando datos históricos.

En una clasificación hay dos tipos de variables, la variable clase o dependiente es la que vamos a hallar; y los atributos o variables independientes que va a permitir ser usados para poder predecir un resultado.



Las variables clases dependientes pueden ser discretos o categórico. Por ejemplo en un clasificación binaria, se puede predecir si un producto tendrá varias ventas, si un cliente cumplirá con pagar las cuotas de un prestamos solicitado podemos predecir con un SI o un NO.

A continuación detallaremos algunos algoritmos de clasificación:

1.3.3.5.1. Algoritmo K 2

El Algoritmo K 2 “se puede considerar el primer algoritmo basado en búsqueda u optimización de una métrica bayesiana y muchos trabajos posteriores se han basado en él”. (Bonilla Gordillo & Ojeda Schuldt, 2006, pág. 43) .

El algoritmo es optimiza, explora la búsqueda en la redes donde están las variables o atributos de una base de datos, donde ordena las variables, y se procede a calcula la métrica.

$$g(i, \pi_i) = \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)} \prod_{k=1}^{r_i} N_{ijk}!$$

1.3.3.5.2. Algoritmo C4.5

El método C4.5 “está basado en árboles de decisión, el cual utiliza la ganancia de información por medio del cálculo de la entropía para medir qué tan bien un atributo separa el conjunto de instancias de acuerdo a sus clases”. (CRUZ-GUERRERO, ALONSO-LA VERNIA, FRANCO-ARCEGA, & SIMON-MARMOLEJO, 27, pág. 12)

El algoritmo C4.5 tiene como antecedente al ID3 tiene como principal función la construcción de un árbol de decisión que son usados para la clasificación, su creación inicia con un nodo raíz vacío, luego se calcula la



entropía de las variables independientes, el atributo con la mayor entropía de información de ubicará en el nodo más alto del árbol, y así sucesivamente con los demás atributos de manera iterativa hasta las hojas.

1.3.3.5.3. Algoritmo HC

El Algoritmo HC, “es un algoritmo local de ascensión de colinas (Hill Climbing) por el máximo gradiente basado en la definición de una vecindad”. (Bonilla Gordillo & Ojeda Schuldt, 2006, pág. 44)

El Algoritmo Hill Climbing , cuando realiza su clasificación lo podía iniciar con el Naive Bayes , al realizar el cálculo de la métrica con los grafos vecinos

1.3.3.5.4. Algoritmo Naive Bayes.

Hipótesis H map: Máxima Posteriori

El clasificador Naive Bayes, es un clasificador bayesiano donde se utiliza el teorema de Bayes, a través de una red bayesiana se obtiene un probabilidad de una variable denominada clase y otras variables llamados atributos.

El clasificador supone que las variables (atributos) están condicionalmente independientes dado la variable clase, es decir los atributos son los nodos hojas y hay un único nodo raíz que es la clase.

Cuando se requiere clasificar un caso donde existan desde A_1, A_2, \dots, x_n variables o atributos y una variable clase donde tiene k categorías, se utiliza la clasificación Naive Bayes, donde:

$$P(C|A_1, \dots, A_n) = \frac{P(A_1, \dots, A_n|C)P(C)}{P(A_1, \dots, A_n)}$$

Y la hipótesis Map se define como:

$$c_{MAP} = \arg \max_{c \in \Omega_c} P(c|a_1, \dots, a_n) = \arg \max_{c \in \Omega_c} P(c) \prod_{i=1}^n P(a_i|c)$$



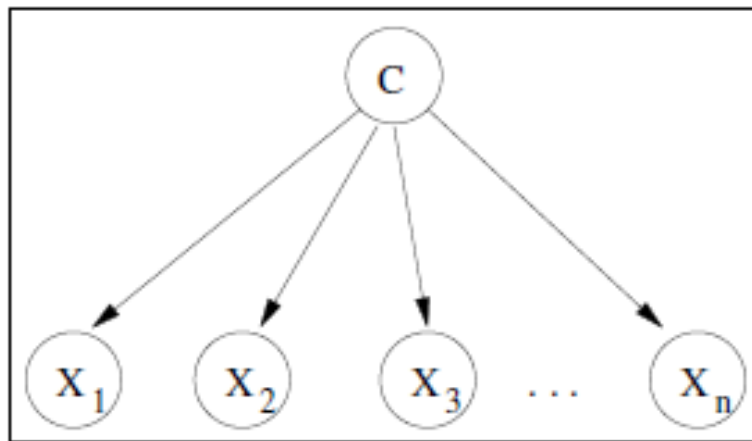
Ilustración 6 : Algoritmo Naive Bayes

Aprendizaje_Bayesiano_Naive(ejemplos)
 Para cada posible valor del resultado v_j
 Obtener estimación $p'(v_j)$ de la probabilidad $P(v_j)$
 Para cada valor a_i de cada atributo a
 Obtener estimación $p'(a_i/v_j)$ de la probabilidad $P(a_i/v_j)$

Clasificar ejemplo(x)
 devolver

$$v_{NB} = \arg \max_{v_j \in V} P(v_j) \prod_i P(a_i / v_j)$$

Ilustración 7: Estructura Naive Bayes



Fuente: (Alarcon Jaimes, 2015)

1.3.3.5.5. Algoritmo Naive Bayes TAN

El Algoritmo TAN, “Friedman y col. (Friedman, 1997) presentan un algoritmo denominado Tree Augmented Network (TAN) el cual consiste básicamente en una adaptación del algoritmo de Chow-Liu ”. (González & Rodríguez , 2006, pág. 7)

Este algoritmo propuesto por Freidnman en 1997, realiza una clasificación más combinada que el NB, es decir las variables o atributos ya no supone



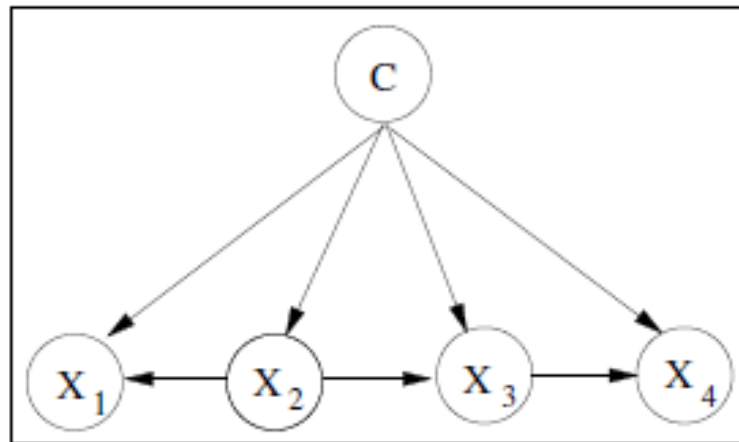
que están independientes, dada la clase, sino que hay admite que hay un grado de dependencia entre ellas.

Para obtener el grado de dependencia de las variables, utilizo o tuvo en cuenta la cantidad de información mutua entre ellas, el cual crea una estructura en forma de árbol.

La fórmula calcular para La cantidad de información mutua entre las variables discretas x e Y condicionada a la variable C se define como:

$$I(X, Y | C) = \sum_{i=1}^n \sum_{j=1}^m \sum_{r=1}^w p(x_i, y_j, c_r) \log \left(\frac{p(x_i, y_j | c_r)}{p(x_i | c_r)p(y_j | c_r)} \right)$$

Ilustración 8: Estructura de Naive Bayes TAN



Fuente: (González & Rodríguez , 2006)

Ilustración 9: Algoritmo Naive Bayes TAN

- Paso 1. Calcular $I(X_i, X_j | C)$ con $i < j, i, j = 1, \dots, n$
- Paso 2. Construir un grafo no dirigido completo cuyos nodos correspondan a las variables predictoras X_1, \dots, X_n . Asignar a cada arista conectando las variables X_i y X_j un peso dado por $I(X_i, X_j | C)$
- Paso 3. A partir del grafo completo anterior y siguiendo el algoritmo de Kruskal construir un árbol expandido de máximo peso
- Paso 4. transformar el árbol no dirigido resultante en uno dirigido, escogiendo una variable como raíz, para a continuación direccionar el resto de las aristas
- Paso 5. Construir un modelo TAN añadiendo un nodo etiquetado como C y posteriormente un arco desde C a cada variable predictora X_i



Fuente: (Artica Chacon & Ocaña Sudario, 2015)

Información Mutua

La información mutua, “es una medida general de asociación o dependencia entre variables procedente del ámbito de la teoría de la información que es capaz no sólo de captar asociaciones lineales, como lo hacen los coeficientes de correlación de Pearson o Spearman,”. (Cortes de la Fuente, 2011, pág. 20)

En el caso discreto, la información mutua entre dos variables X e Y se define de la siguiente forma:

$$I(X, Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)},$$

$p(x)$ es la probabilidad de $X=x$, $p(y)$ es la probabilidad de $Y=y$ y $p(x,y)$ es la probabilidad conjunta de que $X=x$ e $Y=y$.

Por otro lado, en el caso continuo la información mutua entre dos variables X e Y se define como sigue:

$$I(X, Y) = \int_Y \int_X f_{XY}(x, y) \log \frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} dx dy,$$

Siendo $f_{XY}(x,y)$ la función de densidad conjunta de X e Y , y $f_X(x)$ y $f_Y(y)$ las funciones de densidad marginales de X e Y respectivamente.

Entropía de Shannon

La entropía de una variable x , “se interpreta como la cantidad de información que posee dicha variable, o también como la incertidumbre asociada a esa variable”. (Cortes de la Fuente, 2011, pág. 22) , es decir la entropía está íntimamente relacionada con la entropía.



Donde una variable aleatoria discreta X con M posibles estados y sean x_1, \dots, x_M los posibles valores que puede tomar esa variable. Entonces, la entropía de X , $H(X)$, se define de la siguiente forma:

$$H(X) = -\sum_{i=1}^M p(x_i) \log p(x_i).$$

Donde $p(x_i)$ es la probabilidad de

$X=x_i$. Se adoptará la convención de que $0 \cdot \log 0 = 0$ por consideraciones de continuidad.

La entropía de X se interpreta como la cantidad de información que tiene dicha variable, o también como la incertidumbre asociada a esa variable.. En cambio, para casos deterministas en los que $p=1$ para un estado de la variable y 0 para los demás, la entropía será 0 ya que no existe incertidumbre con respecto al valor que tomará X . (Cortes de la Fuente, 2011, pág. 22)

La entropía conjunta de dos variables X e Y está definida como

$$H(X, Y) = -\sum_{i=1}^{M_X} \sum_{j=1}^{M_Y} p(x_i, y_j) \log p(x_i, y_j)$$

1.3.4. Lenguajes de Programación

Un lenguaje de programación, es un “Conjunto de reglas o normas que permiten asociar a cada programa correcto un cálculo que será llevado a cabo por un ordenador (sin ambigüedades), por tanto, un lenguaje de programación es un convenio o acuerdo acerca de cómo se debe de interpretar el significado de los programas de dicho lenguaje” (Ureña Almagro, 2010, pág. 4).



Algunos lenguajes de programación tenemos a los siguientes:

1.3.4.1. Php

El lenguaje de programación PHP , se utiliza mayormente para el desarrollar aplicaciones web y sus siglas en inglés traducidas al español significa Procesador de Hipertexto , tienes una similar combinación con los lenguajes de script, C y Perl es de código abierto, este lenguaje permite funcionar e introducirse con el lenguaje HTML, con PHP se puedan generar páginas con contenidos dinámicos entre más funcionalidades, muchas aplicaciones web están creada con este lenguaje entre las cuales citamos Joomla y Grupal (gestores de contenidos de páginas web) es de adquisición libre lo que quiere decir no tiene ningún costo, para programar con PHP necesitaremos un servidor para ello pero como este lenguaje es multiplataforma, podríamos utilizar cualquiera de los principales servidores web, para lo cual utilizaremos el servidor Apache. Al trabajar con PHP, el programador puede construir páginas web para cooperaciones, multinacionales, grande y pequeñas empresas, para que estas puedan procesar , visualizar su información por medios de formularios con contenido dinámicos. (Coggeshall, 2005).

1.3.4.2. JavaScript

El lenguaje JavaScript es un lenguaje totalmente orientado a objeto, es dinámico, débilmente tipado, basado en prototipos, su sintaxis es parecida al lenguaje C, una de sus funcionalidades en como navegador web pueden utilizar paginas dinámicas y rediseñar sus interfaces, mejorarlas para la satisfacción de los usuarios (client- side)

JavaScript, “es un lenguaje de programación interpretado, dialecto del estándar ECMAScript. Se define como orientado a objetos, basado en



prototipos, imperativo, débilmente tipado y dinámico”. (AlfonsoGU, 2012, pág. 1).

Es una de los lenguajes con mayor demanda laboral , ya que está en todas partes , por ejemplo en los sistemas operativos , base de dato , juegos , en movil. (Flanagan, 2002)

1.3.4.3. Html

El lenguaje HTML sus siglas estas en inglés pero al traducirlas al español significa lenguaje de marcas de hipertexto, este lenguaje utiliza etiquetas o marcas que no son más que breves instrucciones del comienzo al final, esto es la forma de cómo el texto aparecerá en su navegador, los elementos, imágenes que parecerán en la pantalla de su ordenador, este lenguaje es uno de los populares, ya que los navegadores Mozilla, Firefox, Netscape o Explorer lo utilizan para construcción de documentos, este lenguaje puede ser editado en cualquier plataforma .

Unos de las facilidades de usar Html , es que los documentos de textos que ese encuentran en diferentes ordenadores pueden estar conectados a través de una red o vías de telecomunicaciones (Lujan Mora, 2002)

1.3.4.4. Visual .NET

Visual Basic. “es un lenguaje de programación orientado a objetos creado por la Microsoft. Este lenguaje incorpora todas las herramientas necesarias para la creación de cualquier aplicación para Windows”. (Xiloj, 2015, pág. 1).

La programación en VB, ahora con el Visual .Net 2019, posee una interfaz graficas renovadas, donde las soluciones se cargan sin la necesidad de abrir el proyecto completo, que te permitirá diseñar formularios de Windows, al ser un lenguaje muy conocido y utilizado por



muchos programadores, el fácil encontrar documentos fuente, información que permita crear o desarrollar aplicaciones desde lo más pequeñas hasta más complejas. (Rodríguez Bucarelly, Rodríguez Bucarelly, Piñeyro Torres, Morel Pichardo, & Hilario Sanchez, 2008)

1.3.5. Entorno de Programación

1.3.5.1. Eclipse

Eclipse, “es una plataforma de desarrollo open source basada en Java. Eclipse fue desarrollado originalmente por IBM como el sucesor de su familia de herramientas para VisualAge ”. (Garrido Cobo, Lorente Puchades, & Almirall, 2013, pág. 8)

Eclipse al ser un IDE, donde posee variedad de herramientas de código abierto para desarrollar proyectos integrados, con la ayuda que contiene un resaltador o coloreado de código de sintaxis en su editor de texto que ayuda a completar automáticamente un código, por medio de su potente depurado ayuda al programador a mejorar su código fuente, al igual que contiene un extensa colección de plug-ins, de distinta manera que obtenerlas en la modalidad de pagos, gratuitos.

1.3.5.2. Netbeans

NetBeans, “es un entorno modular para el desarrollo de aplicaciones informáticas, escrito en Java. Este IDE (Interfaces Development Environment) está desarrollado para la construcción de sistemas informáticos de diversa índole: aplicaciones de escritorio, para la web o para dispositivos móviles”. (Gómez Jiménez, 2012, pág. 22).

Netbeans al ser fundada en el año 2000 por Sun MicroSystem , al utilizar la plataforma Java donde utiliza sus librerías y demás componentes,



permite desarrollar aplicaciones web dinámicas, de escritorio y móviles , se considera como un IDE muy conocido para los programadores de desarrollo de software, que está en un crecimiento notable. Considerada al igual que Eclipse es una multiplataforma de código abierto ya que también soporta Android, Php , C , C++ , etc

Su acceso a las base de datos, es decir por medio de esta plataforma podemos conectarnos a diferentes base de datos como Oracle , MySql , y ejecutar o realizar consultas , editar , visualizar sus tablas.

Es gratuito, por lo que encontramos variedad de documentación en internet sobre la utilización de este IDE.

1.3.5.3. Visual Studio.NET

Visual Studio .Net, “Es la nueva versión de la familia de herramientas de desarrollo de software de Microsoft, naturalmente orientadas hacia su nuevo entorno de programación: .NET Framework. Es posible la escritura de programas empleando sólo el SDK de .NET Framework”. (Blanco, 2002, pág. 67)

Al ser una plataforma que utiliza el lenguaje de programación .net, también puede utilizar otros lenguajes como C#, C++, F# .

Es una herramienta muy productiva, en la mejora y corrección de código fuente para un desarrollo de mayor velocidad y seguridad, además en las últimas versiones de esta plataforma, esta plataforma viene incorporado la herramienta Azure (servicios en la nube de desarrollar aplicaciones cloud first).

Asimismo Visual Studio tiene a Xamarin para conectar y compilar aplicaciones móviles para Windows Android e iOS

1.3.6. Metodologías De Desarrollo De Software

Una metodología “es un conjunto integrado de técnicas y métodos que permite abordar de forma homogénea y abierta cada una de las actividades del ciclo de vida de un proyecto de desarrollo”. (INTECO, 2009, pág. 37).

Cuando hablamos de metodologías de desarrollo de software, al formar parte indispensables del proceso del desarrollo del mismo desde el requerimiento por parte de los usuarios hasta la el cumplimiento del objetivo, realizan varias actividades de planificación, seguimiento y evaluación.

Las metodologías tienen como objetivo dar a conocer técnicas para el modelado de sistemas que sirven para la creación de software de calidad.

Actualmente, existen varias, a continuación presentamos algunos nombres de las metodologías de desarrollo de software, como Rational Unified Process (RUP), la metodología SCRUM, la metodología basada en prototipo, la metodología XP.

1.3.6.1. RUP

El RUP (Rational Unified Process), o Proceso Unificado Racional “es una metodología que tiene como objetivo ordenar y estructurar el desarrollo de software, en la cual se tienen un conjunto de actividades necesarias para transformar los requisitos del usuario en un sistema Software”. (Guzman Ortiz, 2018, pág. 6).

Esta metodología utiliza el Lenguaje Unificado de Modelado UML, es la más utilizada por su adaptabilidad a los requerimientos de las organizaciones a nivel mundial para el análisis, diseño, implementación y documentación de sistemas orientados a objetos.

La metodología se basa en estos principios:



1. Todo proceso tiene que adaptarse a los requerimientos y/o características propias de una empresa o entidad.
2. Estos requerimientos ya sean funcionales o no tienen prioridades por parte del usuario, en lo que el desarrollador deberá realizar un balance de las mismas para satisfacer a los usuarios.
3. Para poder llevar a cabo todas las fases o proceso de la metodología se debe tener la colaboración de todo el equipo responsable de llevar a cabo cumplir con los objetivos.
4. En cada fase del proceso iterativamente se debe dar un avance del proyecto a realizar, dando a conocer a los usuarios y solicitando las opiniones correspondientes sobre la calidad, estabilidad, rendimiento del producto.
5. La metodología permite a los desarrolladores motivarse al uso de conceptos reutilizables, donde pueden utilizar el UML.
6. Cuando el desarrollador realiza la iteración de cada proceso deberá enfocarse en brindar una mejor calidad de software, para asegurarse en desarrollar un mejor producto.

La metodología RUP se divide en 4 iteraciones o ciclos de vida:

Inicio

Es la primera fase o ciclo de la metodología, donde se va a definir el modelado del negocio, es decir la identificación de requisitos, de los casos de uso y sus actores, donde se determina que recursos (estimar un coste de tiempo – recursos) van a ser designados al proyecto mediante un plan de negocio.

Elaboración

En esta fase, ya se debe tener concretado los casos de usos identificados en la primera fase para su análisis, diseño de la arquitectura de sistema, o sea un preliminar diseño.

Construcción



En esta fase se debe desarrollar el producto (minimizando los costes para la optimización de recursos), es proporcionarlo documentalmente, a través de sucesivas interacciones de análisis, diseños e implementación, también desarrollar manuales para facilitar al usuario.

Transición

Es la etapa donde ya se le entrega el software o producto al usuario, con sus respectivas capacitaciones sobre la configuración, funcionalidades y facilidades del uso del software (podría utilizar una versión Beta para realizar las validaciones del nuevo sistema), y posteriormente desarrollar las versión nuevas del producto.

1.3.6.2. Metodología XP

La metodología XP, “es una metodología ligera de desarrollo de aplicaciones que se basa en la simplicidad, la comunicación y la realimentación del código desarrollado”. (Melendez Valladares, Gaitan, & Perez Reyes, 2015, pág. 26)

XP o Programación Extrema – Extreme Programming, es una de las metodologías ágiles para desarrollar software muy usada y exitosa, de manera eficiente y efectiva. La metodología fomenta el trabajo en equipo como la clave del éxito mediante las relaciones interpersonales y la retroalimentación del equipo con el cliente.

La metodología XP comprende 4 fases que son:

Planificación del Proyecto

En esta primera fase se juntan todos los requerimientos, el corto ya que puede tomar pocos días, donde se puede convocar a reuniones para poder realizar un cronograma de las interacciones a realizar, con el fin de crear la arquitectura del proyecto.

Diseño

El diseño del sistema o programa se deberá crear de manera sencilla y simple con las funcionalidades principales del mismo, es decir se desarrolla la interfaz donde el usuario interactuará con el sistema.



Codificación

En esta fase de la codificación los programadores del proyecto y el cliente deben estar comunicados para se pueda codificar todo lo que el programa necesita, aquí se debe tener en cuenta que el código utilizado funcione correctamente

Pruebas

En esta fase se hacen las pruebas de acuerdo a los requerimientos del cliente, con el fin de comprobar y verificar que estos estén correctamente funcionando con las iteraciones de los pasos anteriores.

1.3.6.3. Metodología basada en prototipos

Esta metodología, “Es un enfoque basado en un proceso iterativo en que a partir de un análisis de requisitos inicial se realiza prototipos cada vez más complejos. El cliente da su opinión y se incorporan sus sugerencias para la siguiente iteración”. (Ganzabal Garcia, 2014, pág. 36).

La ventaja de esta metodología es que el programador del software se da cuenta el avance del proyecto y lo que el cliente requiere.

Comprende las siguientes etapas:

Identificar requerimientos básicos:

En esta etapa inicial se identifica el problema, las necesidades o requerimientos de los usuarios para determinar la funcionalidad el sistema o software para su solución.

Diseño del prototipo inicial

Se diseña el prototipo inicial, como modelo preliminar, donde los usuarios puedan tener la idea de cómo se van a implementar sus requerimientos en el sistema.

Implementación del prototipo

Se realiza la codificación, mediante una arquitectura definida.



Pruebas y mejora del prototipo

En esta última etapa se realizan las correspondientes pruebas con el fin de verificar las funcionalidades del sistema y la corroboración con los requerimientos solicitados, con las retroalimentaciones correspondientes.

1.4. Formulación del problema

¿En qué medida la implementación de un sistema basado en algoritmos computacionales ayudará a los procesos de selección profesional en el Ministerio de Transporte y Comunicaciones?

1.5. Justificación e Importancia de Estudio

En la presente tesis, es importante, ya que mediante la implementación de un sistema basado en algoritmos computacionales ayudara a los procesos de profesionales en lo que respecta al tiempo, costo y también en elegir el mejor perfil de la persona que se encuentra postulando.

Además, no sólo beneficiará a la Oficina de Recursos Humanos, que es el área encargada de la selección y evaluación de los perfiles profesionales, sino que se piensa beneficiar y relacionar a las demás áreas de la institución, ya que dependiendo de las necesidades que éstas requieran, el área de Personal realizará la previa selección para tratar de cubrirlas y satisfacerlas con los nuevos aspirantes a dichos puestos de trabajo; así veremos que la institución se unirá y se complementará en equipo.

1.6. Hipótesis

Mediante la implementación de un sistema basado en algoritmos computacionales ayudará a mejorar los procesos de selección profesional.



1.7. Objetivos

1.7.1. Objetivos General

Implementar un sistema basado en algoritmos computacionales como ayuda a los procesos de selección profesional en el Ministerio de Transportes y Comunicaciones - MTC.

1.7.2. Objetivos Específicos

- a) Efectuar el análisis de requerimientos.
- b) Seleccionar el algoritmo computacional para el desarrollo del software.
- c) Implementar el sistema basado en algoritmo computacional
- d) Realizar las pruebas - testing del software.



II

MATERIAL Y METODOS

II. MATERIAL Y METODOS

2.1. Tipo y diseño de investigación

2.1.1. Tipo de investigación

En la presente tesis el tipo de investigación es Aplicada esto se debe porque se emplearon las teorías ya adquiridas y lo aplica a un problema determinado.

Por el enfoque: Cuantitativa.

Porque por medio del planteamiento de objetivos e hipótesis realiza la comprensión y solución del problema

Por el nivel de Tecnológico: Tecnológica:

Porque busca descubrir y conocer las técnicas que son más eficaces o apropiadas para producir cambios en el tema de investigación.

2.1.2. Diseño de investigación

El diseño de investigación en la tesis es descriptiva porque se describieron diferentes elementos de la situación problemática de la selección profesional en el Ministerio de Transportes y Comunicaciones - MTC.

2.1.3. Población y muestra

La población N en este estudio estuvo conformada por una muestra de 3 Procesos de Selección en la Oficina de Personal del Ministerio de Transportes y Comunicaciones – MTC.

2.2. Variables, Operacionalizacion

2.2.1. Variable Independiente

Sistema basado en algoritmos Computacionales

Tabla 1 : Variable Independiente

Variable Independiente	Indicadores	Fórmulas	Técnicas e instrumentos de recolección de datos
sistema basado en algoritmos computacionales	1.- Número de funcionalidades implementadas que funcionan de acuerdo a la especificación de requerimientos	$\frac{NRF}{TR} * 100$ <p>NRF= Numero de requerimientos que si funcionan TR=Total de Requerimiento</p>	
	2.- Precisión	$P = \frac{VP}{VP + FP}$ <p>Donde P= Precisión VP= Verdadero Positivo FP = Falso Positivo</p>	Informe de Pruebas funcionales del sistema
	3.- Accuracy	$A = \frac{VP + VN}{VP + VN + FP + FN}$ <p>Donde A : Accuracy VP : Verdadero Positivo VN :Verdadero Negativo FP : Falso Positivo FN : Falso Negativo</p>	

Fuente: Elaboración Propia



2.2.2. Variable Dependiente.

Ayuda a los procesos de selección profesional.

Tabla 2: *Variable Dependiente*

Variable Dependiente	Indicadores	Técnicas e instrumentos de recolección de datos
Procesos de selección profesional	-Tiempo de todas las acciones a seguir en el proceso de selección profesional.	Revisión documental Encuestas Análisis estadístico

Fuente: Elaboración Propia



2.3. Técnicas e instrumentos de recolección de datos, validez y confiabilidad.

En esta investigación de tesis las técnicas e instrumentos son:

Tabla 3: Técnica e instrumentos de recolección de datos, validez y confiabilidad

Técnica	Uso	Instrumento
Encuesta.	Se procedió a realizar la entrevista a los empleados de la Oficina de Personal que seleccionan el personal para conocer cómo está la evaluación de la selección profesional en el Ministerio de Transportes y Comunicaciones – MTC del Perú.	Cuestionario
Análisis documentario	Se revisaron los registros de selección profesional de la Oficina de Personal en el Ministerio de Transportes y Comunicaciones – MTC del Perú.	Análisis de documentos.
Observación.	Se observaron los procesos de selección profesional de la Oficina de Personal en el Ministerio de Transportes y Comunicaciones – MTC del Perú.	Registro de observación.
Test Funcionales	Las pruebas del sistema	Informe de test funcionales

Fuente: Elaboración Propia



2.4. Procedimiento de análisis de datos

La recolección de la información sirvió para poder realizar las interpretaciones y toma de decisiones con el propósito de dar alternativas de solución del problema. Se contó con la colaboración del personal administrativo que ayudaron sobre la selección profesional en el Ministerio de Transporte y Comunicaciones.

Participaron en la investigación: empleados, personal administrativo y el investigador, a través de las técnicas como: observación, encuestas, análisis de documentos, informes e internet.

2.4.1. Análisis estadístico e interpretación de los datos

En esta tesis se utiliza el cálculo de probabilidades como la condicionada y el teorema de redes bayesianas con el Clasificador bayesiano aumentada a árbol (TAN)

También se utilizó el cálculo de la Entropía y la Cantidad de Información Mutua entre cada variable o atributo.

Además, se analizó la información y se utilizó la estadística descriptiva con la utilización de porcentajes; el cual se representaron en forma de gráficos y cuadros.



Indicadores

a. Numero de funcionalidades implementadas que funcionan de acuerdo a la especificación de requerimientos:

Se realizó una prueba de funcionalidad del sistema al encargado de la Oficina de Personal de acuerdo a los requerimientos.

$$\frac{NRF}{TR} * 100$$

NRF= Numero de requerimientos que si funcionan

TR=Total de Requerimiento

Para calcular los indicadores: exactitud, precisión se debe realizar una matriz de confusión, que nos dará una mejor idea de cómo está clasificando nuestro modelo.

Tabla 4: Matriz de Confusión

Matriz de Confusión		Predicción	
		Seleccionado	No Seleccionado
Real	Seleccionado	Verdadero Positivo (VP)	Falso Negativo (FN)
	No seleccionado	Falso Positivo (FP)	Verdadero Negativo (VN)

Fuente: Elaboración Propia



Precisión: También llamado valor de precisión positivo. Es la razón de los casos positivos que fueron identificados correctamente, es decir nos da la probabilidad de que, dada una predicción positiva, la realidad sea positiva también.

$$P = \frac{VP}{VP + FP}$$

Donde

P= Precisión

VP= Verdadero Positivo

FP = Falso Positivo

La Accuracy (Exactitud): Es la razón del total de predicciones que son correctas (tanto positivas como negativas) respecto al total. Se determina utilizando la ecuación:

$$A = \frac{VP + VN}{VP + VN + FP + FN}$$

Donde

A= Accuracy

VP = Verdadero Positivo

FP = Falso Positivo

VN= Verdadero Negativo

FN =Falso Negativo



Indicadores

Tiempo:

Se refiere al tiempo que se toma los procesos de selección (en minutos)

Las siguientes actividades son:

- a) Verificación de Datos.
 - En pre test (sin sistema); se refiere a la verificación del CV documentado de los postulantes sin aplicación.
 - En post test (con sistema); se refiere a la verificación del CV de los postulantes en la aplicación
- b) Procesamiento de Datos:
 - En pre test (sin sistema), se refiere a la evaluación del CV (indicar un puntaje por cada requisito del perfil de puesto) de los postulantes.
 - En post test (con sistema), se refiere al proceso de cálculo del algoritmo bayesiano en la aplicación.
- c) Lectura de Resultado:
 - En pre test (sin sistema), se refiere a visualización con un informe el resultado final de acuerdo a todos los requisitos del puesto.
 - En post test (con sistema), se refiere a la visualización, de la red bayesiana del resultado final (probabilidades) de los postulantes.

2.5. Aspectos Éticos

En esta tesis, como investigador, se aseguró los derechos de las personas a estar protegidos (en cuanto a la información recolectada.)

Principio de beneficencia. - Que consiste en la obligación ético y moral profesional de haber realizado esta investigación en bienestar de los otros, promover la mejora, en este caso mejorar el proceso de selección profesional.



Está orientado a la acción, que debe realizarse de una forma responsable y de respeto

Principio de no maleficencia.- Principio que consiste en la obligación de no causar perjuicios o daños, de alguna manera por la obtención de una adicional información, el investigador pudo caer en acciones antiéticas.

Principio de justicia.- Este principio declara fundamentalmente en la igualdad de los seres humanos, donde está el respeto a sus deberes, al trato equivalente y misma consideración, es decir no pueden tratarse ni verse con un menor valor. En la investigación se consideró el respeto y el trato de igual entre los entrevistados.

Respeto a las personas o a su principio de autonomía. - La persona tratada participo voluntariamente, sin ninguna obligación personal o institucional, realizando su permiso para ser observado y entrevistado. Se respetó la autonomía quiere decir, otorgar valor a las elecciones y opiniones de personas autónomas, no obstruyendo su actos.

2.6. Criterios de rigor científico

Verdad o validez. Este criterio, su objetivo es el aseguramiento y demostración que la investigación de esta tesis, en la que se propone un sistema basado en algoritmos computacionales, se ha realizado de manera ajustada a los hechos, respetando los resultados, es decir, garantizando que el trabajo de investigación fue descrito e identificado con exactitud.

Aplicabilidad.- La aplicabilidad da la posibilidad de emplear los descubrimientos a otras realidades. Se garantizó que, al obtener unos buenos resultados, en base a una minuciosa investigación y apegados a la verdad, los resultados, de la implementación del Algoritmo Computacional, puedan ser usados o propuestos para otros investigadores.

Consistencia. Quiere decir que el grado de verosimilitud del presente trabajo tiene un reflejo real a los hechos, lo que al repetir comprobaría los mismos hechos, lo que demostraría que hay una seguridad y confianza en el trabajo realizado.

Neutralidad.- Principio para ser imparciales y objetivos, donde los resultados obtenidos en el trabajo no están direccionados a uno u otro, de manera parcializada, sino más bien demuestra independencia de los intereses o preferencias, que son garantía del investigador de los resultados.

III

RESULTADOS



III. RESULTADOS

3.1. Resultados en tablas y figuras

Indicador 1: Número de funcionalidades implementadas que funcionan de acuerdo a la especificación de requerimientos.

Este indicador se evaluó de acuerdo a una prueba del sistema a la persona encargada de la Oficina de Personal sobre las funcionalidades implementadas de acuerdo a los requerimientos (Ver Anexo B), los cuales se obtuvieron los siguientes resultados.

De las 14 funcionalidades, las 14 funcionan correctamente

Por lo tanto:

$$\frac{NRF}{TR} * 100$$

NRF= Numero de requerimientos que si funcionan

TR=Total de Requerimiento

Donde,

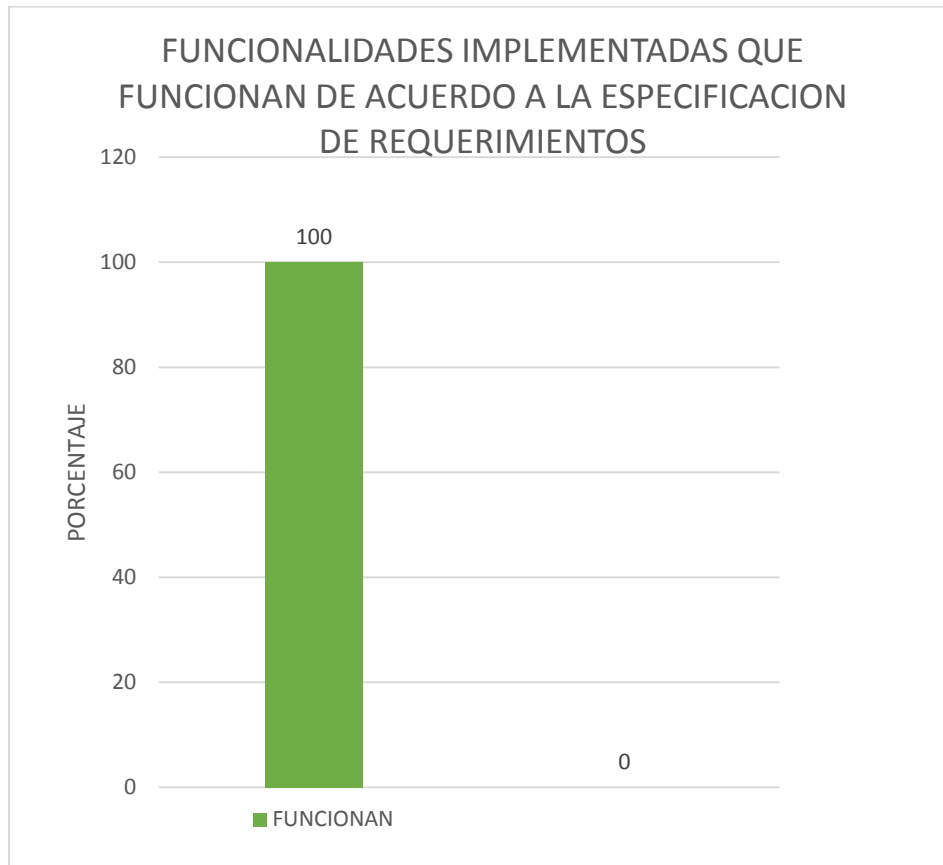
NRF= 14, TR= 14

$$= \frac{14}{14} * 100$$

$$= 100\%$$



Ilustración 10: Funcionalidades Implementadas.



Fuente: Elaboración propia

Indicador 2 y 3:

Precisión; nos da la probabilidad de que, dada una predicción positiva, la realidad sea positiva también.



Exactitud, nos da la probabilidad dadas todas las predicciones sean correctas (positiva y negativas) sobre el total de predicciones.

Para evaluar este indicador se realizó una matriz de confusión por cada proceso de selección.

Proceso 01: Asesor Legal.

Se registraron 11 postulantes los cuales se obtuvieron resultados descritos en el Anexo C.

De acuerdo a los resultados obtenidos, se elaboró una nuestra matriz de confusión de acuerdo a las categorías obtenidas.

Donde:

VP: Verdadero Positivo

FP: Falso Positivo

FN: Falso Negativo

VN: Verdadero Negativo

Tabla 5: *Matriz de Confusión: Proceso de Selección 01*

Matriz de Confusión	Predicción		Total
	Seleccionado	No Seleccionado	
			7
	Seleccionado	VP =4	FN=3
Real	No seleccionado	FP=1	VN=4
	Total	5	6

Fuente: Elaboración propia



Se halló la precisión con la siguiente fórmula

$$P = \frac{VP}{VP + FP}$$

$$P = \frac{4}{4 + 1}$$

$$P = \frac{4}{5}$$

$$P = 0.8 * 100$$

$$P = 80\%$$

Se halló la exactitud con la siguiente fórmula:

$$A = \frac{VP + VN}{VP + VN + FP + FN}$$

$$A = \frac{4 + 4}{4 + 3 + 1 + 3}$$

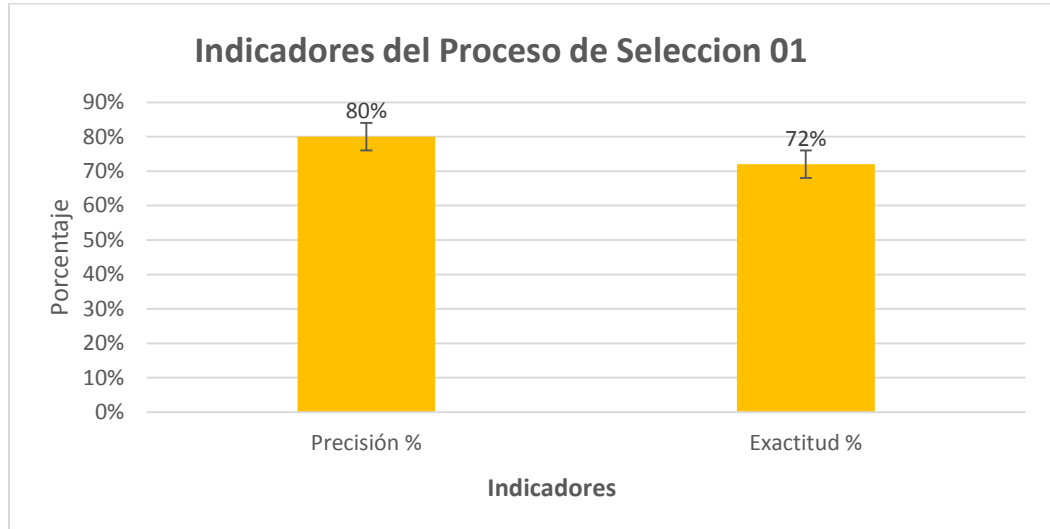
$$A = \frac{8}{11}$$

$$A = 0.72 * 100$$

$$A = 72 \%$$



Ilustración 11: Resultado de Indicadores del Proceso de Selección 01



Fuente. Elaboración Propia

Proceso N° 02: Asesor de Planeamiento

Se registraron 10 postulantes los cuales los cuales se obtuvieron resultados descritos en el Anexo D.

De acuerdo a los resultados, se elaboró una matriz de confusión de acuerdo a las categorías obtenidas

Tabla 6 : Matriz de Confusión: Proceso de Selección 02

Matriz de Confusión	Predicción		Total
	Seleccionado	No Seleccionado	
Real	Seleccionado	VP =3 FN=2	5
	No seleccionado	FP=1 VN=4	5
Total	4	6	

Fuente: Elaboración Propia



Se halló la precisión con la siguiente fórmula:

$$P = \frac{VP}{VP + FP}$$

$$P = \frac{3}{3 + 1}$$

$$P = \frac{3}{4}$$

$$P = 0.75 * 100$$

$$P = 75\%$$

Se halló la exactitud con la siguiente fórmula

$$A = \frac{VP + VN}{VP + VN + FP + FN}$$

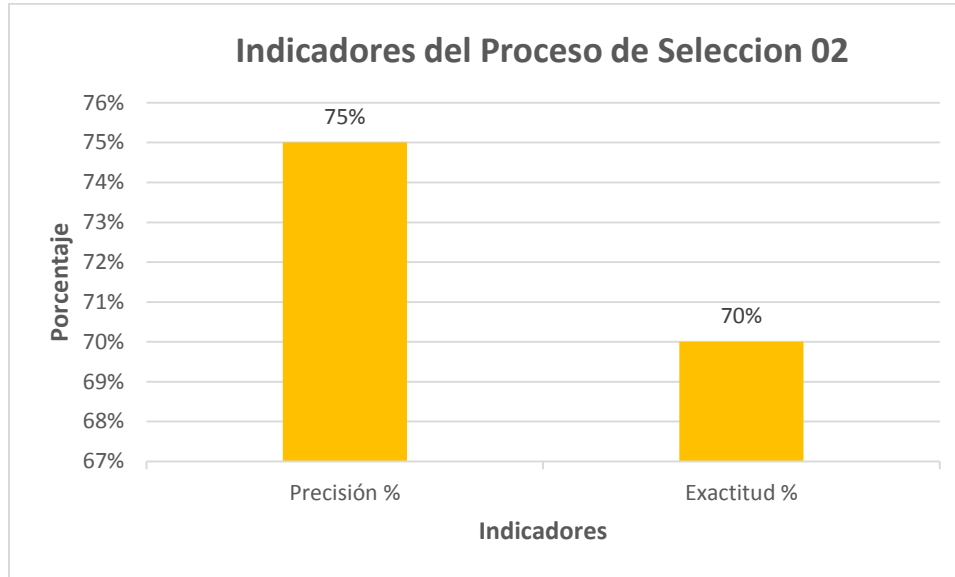
$$A = \frac{3 + 4}{3 + 4 + 1 + 2}$$

$$A = \frac{7}{10}$$

$$A = 0.7 * 100$$

$$A = 70 \%$$

Ilustración 12: Resultado de Indicadores del Proceso de Selección 02



Fuente: Elaboración Propia

Convocatoria N° 03: Analista de Sistemas

Se registraron 13 postulantes los cuales se obtuvieron resultados descritos en el Anexo E

Se elaboró una matriz de confusión de acuerdo a las categorías obtenidas.

Tabla 7 : Matriz de Confusión: Proceso de Selección 03

Matriz de Confusión	Predicción		Total
	Seleccionado	No Seleccionado	
Real	Seleccionado	VP =5 FN=3	8
	No seleccionado	FP=1 VN=4	5
Total	6	7	

Fuente: Elaboración Propia



Se halló la precisión con la siguiente fórmula:

$$P = \frac{VP}{VP + FP}$$

$$P = \frac{5}{5 + 1}$$

$$P = \frac{5}{6}$$

$$P = 0.83 * 100$$

$$P = 83\%$$

Se halló la exactitud con la siguiente fórmula:

$$A = \frac{VP + VN}{VP + VN + FP + FN}$$

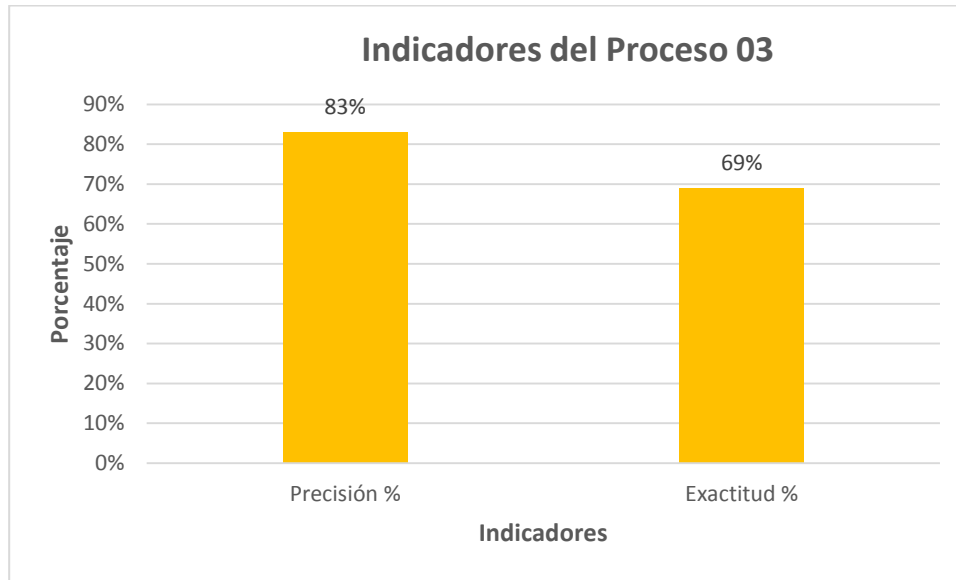
$$A = \frac{5 + 4}{5 + 4 + 1 + 3}$$

$$A = \frac{9}{13}$$

$$A = 0.69 * 100$$

$$A = 69 \%$$

Ilustración 13: Resultado de los Indicadores del Proceso de Selección 03



Fuente: Elaboración de Propio

La probabilidad promedio de los indicadores: Precisión y exactitud de los 3 procesos de selección queda de la siguiente manera:

Tabla 8: Resultado Promedio de los Procesos de Selección

Proceso	Precisión %	Exactitud %
Convocatoria N° 01: Asesor Legal	80%	72%
Convocatoria N° 02: Asesor de Planeamiento	75%	70%
Convocatoria N° 03: Analista de Sistemas	83%	69%
Promedio	79.33%	70.33%

Fuente: Elaboración Propia



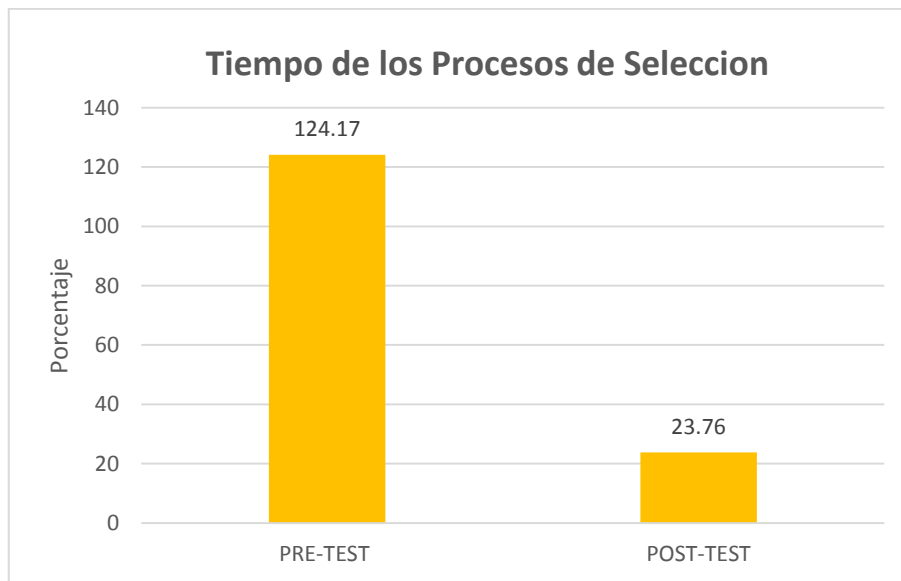
Indicador 4: Tiempo de todas las acciones en el proceso de selección profesional (minutos)
 Para la evaluación del indicador se consideró el promedio en minutos de cada uno de los tres tipos de actividades consideradas durante el proceso de selección.

Tabla 9: *Tiempo de las acciones a seguir en el Proceso de Selección*

TIPOS	PRE-TEST				POST-TEST			
	PS1	PS2	PS3	PROM	PS1	PS2	PS3	PROM
Registro y Verificación de datos	60.00	72.00	66.50	66.17	20.00	25.00	22.00	22.33
Procesamiento	29.00	41.00	37.00	35.67	0.40	0.25	0.15	0.27
Lectura de resultados	24.00	16.00	27.00	22.33	0.35	0.45	0.36	1.16
TOTAL				124.17				23.76

Fuente: Elaboración Propia

Ilustración 14: *Tiempo de todas las acciones a seguir en el Proceso de Selección*



Fuente . Elaboracion Propia



3.2. Discusión de resultados

Podemos observar que las 14 funcionalidades implementadas funcionan correctamente de acuerdo a la especificación de los requerimientos siendo un total del 100%.

Se observa que en el Proceso de Selección 01 nos da un porcentaje de Precisión de 80 % y de Exactitud de 72 %

Se observa que en el Proceso de Selección 02 nos da un porcentaje de Precisión de 75 % y de Exactitud de 70 %

Se observa que en el Proceso de Selección 03 nos da un porcentaje de Precisión de 83 % y de Exactitud de 69 %

En tal sentido se observa que la implementación de un sistema utilizando el algoritmo bayesiano Tan en los procesos de selección tiene un promedio de precisión de 79.33% y exactitud de 70.33%

También se observa que el tiempo promedio de los procesos de selección disminuye con la implementación de un sistema utilizando el algoritmo bayesiano Tan de 124.17 minutos a 23.76 minutos

3.3. Aporte práctico

3.3.1. Información General del Ministerio de Transportes y Comunicaciones.

El Ministerio de Transportes y Comunicaciones (MTC) , como parte del Poder Ejecutivo, es el responsable de la infraestructura y desarrollo de los sistemas de transporte , comunicaciones y telecomunicaciones del Perú, que es importante para integración a sea a nivel regional , nacional e internacional, facilitando el comercio para el bienestar de todos los peruanos.

De igual manera, este Ministerio es el que promueve de manera eficiente los sistemas ferroviarios, aéreos, terrestres, marítimos y fluviales, las redes de telecomunicaciones, también programas de concesiones en dichos rubros.

Lo que corresponde al transporte terrestre, es el promotor del mejoramiento y las construcciones de las nuevas carreteras y sistemas de transporte público; en lo aéreo en la verificación y regulación de las líneas aéreas comerciales en los aeropuertos, para que garantice la seguridad de los usuarios peruanos y extranjeros.

Los que corresponde al transporte marítimo, por medio de sus órganos de la competencia del rubro, son los encargados de estimular el mejoramiento de los puertos a nivel nacional para un adecuado traslado de personas y productos.



Los que corresponde a las comunicaciones por medio de sus órganos de la competencia del rubro, tienen la función de supervisar y controlar las actividades de prestación de los servicios de comunicaciones, además de realizar los trámites de las operaciones de estación de radio y tv en señal abierta.

3.3.2. Visión del Ministerio de Transportes y Comunicaciones.

Su visión es, “Ministerio distinguido por su eficiencia en la gestión de los sectores transportes y comunicaciones, garantizando servicios integrales, seguros y competitivos” (Ministerio de Transportes y Comunicaciones, 2017).

3.3.3. Cuadro Comparativo de las Metodologías de Desarrollo de Software

Tabla 10: Cuadro Comparativo de las Metodologías de Desarrollo de Software

FUNCION	SCRUM	XP	RUP	BASADO EN PROTOTIPO
Estabilidad	Emplea actuales y más productivas herramientas	Metodología eficiente, sencilla y ligera.	De acuerdo a la arquitectura que se utiliza, emplea casos de usos.	Es fácil de usar y modificar. Exige disponer de las herramientas adecuadas
Flexibilidad	En el trabajo en el equipo emplea técnicas y herramientas.	Para disponibilidad del usuario utiliza modelos de implementación.	Se adapta al espacio tiempo, necesidad y recursos disponibles	Alto grado de participación del usuario
Rendimiento	Las prioridades de los equipos de trabajo son muy productivas.	Para la elaboración final del sistema se toma en cuenta la calidad y productividad.	Equipos de trabajos enfocados en procesos definidos	Al crear software disminuye el riesgo de que estos no satisfagan a los usuarios.
Diseño	Es adaptable con diversos sistemas de desarrollo de software.	Su diseño se basa en las funcionalidades que contiene.	Sistemas orientados a objetos, iterativo e incremental	Permite un diseño rápido que estable una buena aceptación del software.
Implementación	Proyecto muy complejo.	Su utilización es en proyectos de poco realce, por ser ligero	Se adapta a diversos proyectos así sean complejos.	La implementación de sistema resulta menos compleja, porque los usuarios ya conocen o han realizado las pruebas con el prototipo.



Pruebas	Es adaptable y aplica la verificación y demostración.	Realiza pruebas en la implantación de aceptación, tiene una protección para evitar fallos.	Adaptabilidad y Verificación, Aplicando cada fase según sus funcionalidades y aislando las fallas.	Cuando el usuario no está conforme con una parte del diseño del prototipo, quiere decir que una parte de la prueba fallo , lo que debe corregirse hasta que el usuario de la conformidad
Tipo de Proyectos de Software	Es recomendable para organizaciones que quieran rápidas mejoras y no haya una fecha límite.	Aplicaciones Móviles, Proyectos pequeños	Grandes empresas con proyectos de un nivel medio y grande de complejidad	Software de Investigación
Etapas	Desarrollo simple no requiere de trabajo duro	<ul style="list-style-type: none"> ✓ Planificar ✓ Diseñar ✓ Codificar ✓ Pruebas 	<ul style="list-style-type: none"> ✓ Iniciar ✓ Elaborar ✓ Construir ✓ Transitar 	<ul style="list-style-type: none"> ✓ Investigación Preliminar ✓ Diseño Técnico ✓ Programación y Prueba ✓ Operación y Mantenición

Fuente. Elaboración Propia

En esta tesis se utilizará la metodología de desarrollo de software basado en prototipos, cuyas ventajas implican la reducción del riesgo, ya que nos permite tener una muestra del sistema que se está desarrollado en muy poco tiempo, con pocos recursos y que el usuario final va a poder visualizar antes algo que vaya a querer modificar.



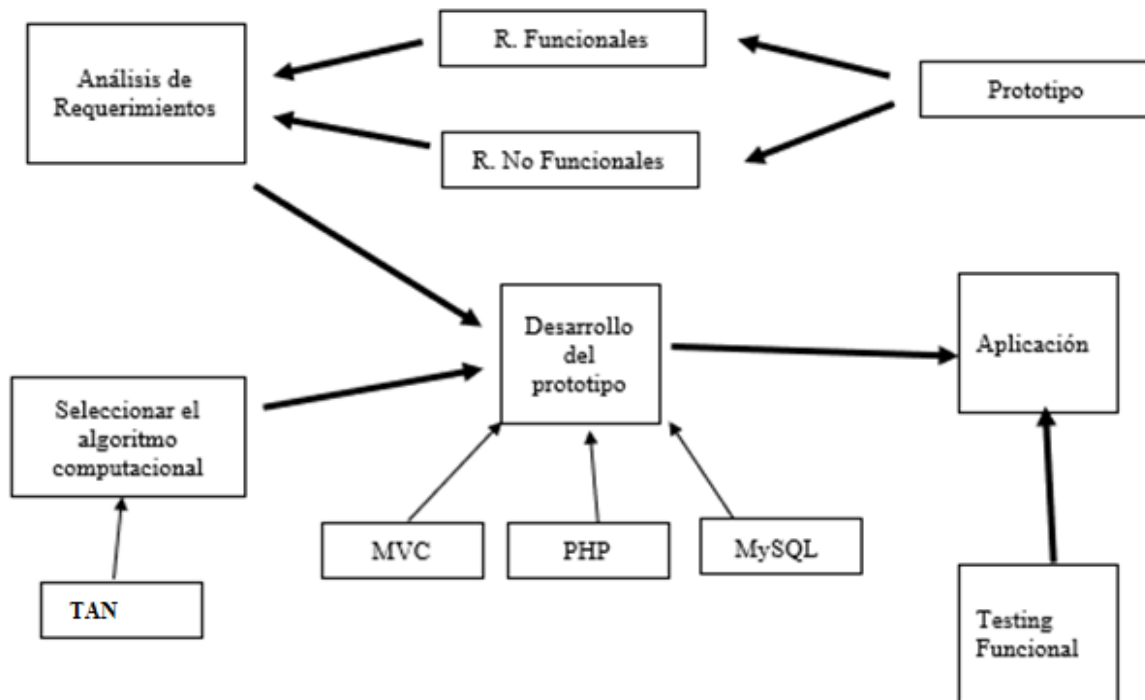
3.3.4. Metodología de Desarrollo de Software

En este trabajo de tesis, con el fin de cumplir con los objetivos trazados se utilizó la metodología de desarrollo de software basada en prototipos. Los motivos por lo que se eligió son:

- Solamente se desarrollaron los requerimientos que se tuvo conocimiento.
- Desarrollo del sistema por etapas.
- Pocos recursos, reducción de riesgos
- Hubo una muestra del sistema en poco tiempo

FLUJOGRAMA DE DESARROLLO DEL SISTEMA BASADO EN ALGORITMOS COMPUTACIONALES

Ilustración 15: *Flujograma del desarrollo del sistema basado en algoritmos computacionales*



Fuente. Elaboración Propia



3.3.4.1. Metodología Basada en Prototipos

Esta metodología permitió desarrollar actividades en las siguientes etapas:

3.3.4.1.1. Identificación de requerimientos básicos

La identificación de los requerimientos eran necesarios para la construcción del prototipo funcional; por lo tanto, permitió que se desarrollen funcionalidades para la solución de distintos problemas y tener una solución inmediata.

Los siguientes requerimientos no funcionales fueron:

Tabla 11: *Requerimiento No Funcionales*

IDENTIFICACION	NOMBRE	CONCEPTO
RNF01	Interfaz de Usuario	El sistema mostrará una interfaz de usuario amigable y sencilla
RNF02	Alertas o Mensajes de Error	El sistema debe contener y mostrar mensajes de error o alertas para el usuario

Fuente. Elaboración Propia



Los siguientes requerimientos funcionales fueron:

Tabla 12: Requerimientos Funcionales

IDENTIFICACION	NOMBRE	CONCEPTO
RF01	Creación de Convocatorias	El encargado del proceso de selección deberá crear convocatorias, así como cerrarlas
RF02	Visualización de convocatorias	El postulante deberá visualizar las convocatorias abiertas para que pueda ingresar a registrarse
RF03	Visualización del Perfil del Puesto	El postulante visualizara los requisitos mínimos para el puesto.
RF04	Registro de datos personales	El postulante deberá registrar: sus apellidos, nombres, fecha de nacimiento, documento nacional de identidad, email, dirección, teléfono, celular.
RF05	Registro de Formación Académica y Cursos	El postulante deberá registrar su grado académico y los cursos, talleres que tiene.
RF06	Registro de Experiencia Laboral	El postulante deberá registrar sus experiencias laborales: cargo que desempeño, año inicio y fin, empresa y/o institución
RF07	Login del encargado	El encargado de los procesos de selección deberá ingresar al sistema mediante un usuario y contraseña
RF08	Registro de Resultado Evaluación Psicotecnia	El encargado de los procesos de selección deberá registrar los resultados de la evaluación psicotécnica
RF09	Registro de resultado de	El encargado de los procesos de selección deberá registrar los resultados de la entrevista personal



	evaluación de entrevista	
RF10	Listado de Postulantes	El encargado del proceso de selección visualizará a los postulantes (sus datos personales, su experiencia laboral, estudios, evaluaciones)
RF11	Generación de Red	El encargado del proceso de selección visualizará la red generada
RF12	Visualización de Resultados	El encargado del proceso de selección visualizará los resultados (porcentajes) de los cálculos realizados por el sistema

Fuente. Elaboración Propia

3.3.4.1.2. Diseño del prototipo inicial.

El diseño es la segunda fase de la metodología de desarrollo basada en prototipos.

En esta sección se presentan de datos interfaz del sistema

Ilustración 16: Prototipo Interfaz Inicial



Fuente: Elaboración Propia

Ilustración 17: Prototipo Registro de Postulante



Fuente: Elaboración Propia



The image shows a web interface with two main sections: 'Formación Académica' and 'Experiencia Laboral'. The 'Formación Académica' section includes a dropdown for 'Grado Instrucción', a 'Capacitaciones' section with fields for 'Capacitación' and 'Institución', and date fields for 'Inicio' and 'Fin' with a 'Horas' field. The 'Experiencia Laboral' section includes fields for 'Cargo' and 'Empresa', and date fields for 'Inicio' and 'Fin'. Both sections have '+ Agregar' buttons. At the bottom, there are 'CANCELAR' and 'GUARDAR' buttons.

Fuente: Elaboración Propia

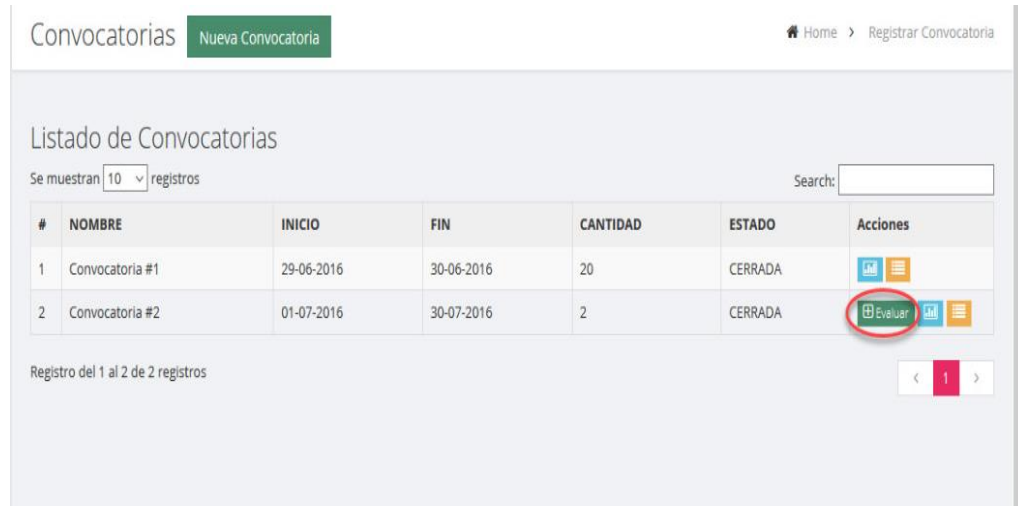
Ilustración 18: Prototipo Login Administrador

The image shows a login form titled 'Ingreso Seguro'. It features a yellow header bar with the title. Below the header, there are two input fields: 'Usuario' and 'Contraseña'. At the bottom, there are two buttons: 'INGRESAR' (green) and 'CANCELAR' (yellow).

Fuente: Elaboración Propia

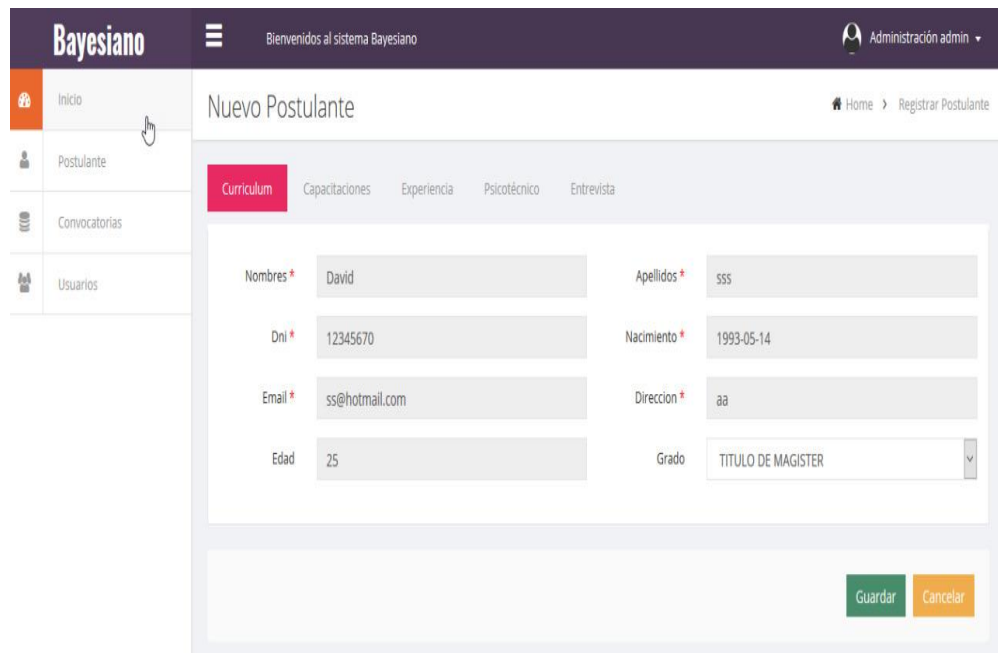


Ilustración 19: Prototipo de Visualización de Procesos o Convocatorias



Fuente: Elaboración Propia

Ilustración 20: Prototipo de Visualización de los Datos Personales de Postulantes



Fuente: Elaboración Propia



Ilustración 21: Prototipo de Visualización de Capacitaciones de Postulantes

Bayesiano Le ir Administración admin

Nuevo Postulante Home > Registrar Postulante

Curriculum **Capacitaciones** Experiencia Psicotécnico Entrevista

Nombres * David Apellidos * sss
 Dni * 12345670 Direccion * aa

Capacitaciones

Nombre	Institución	Inicio	Fin	Horas
ss	ss	2018-08-06	2018-08-06	20.00
xx	xx	2018-08-07	2018-08-07	10.00

Guardar Cancelar

Fuente: Elaboración Propia

Ilustración 22: Prototipo de Visualización de Experiencias Laborales de Postulantes

Bayesiano Le invitamos a explorar las funci Administración admin

Nuevo Postulante Home > Registrar Postulante

Curriculum Capacitaciones **Experiencia** Psicotécnico Entrevista

Nombres * David Apellidos * sss
 Dni * 12345670 Direccion * aa

Experiencias

Cargo	Empresa	Inicio	Fin	Tiempo
aa	aa	2010-08-01	2012-08-03	2.00
bb	bb	2012-05-01	2015-07-01	3.17

Guardar Cancelar

Fuente: Elaboración Propia



Ilustración 23: Prototipo de Visualización de Exámenes Psicotécnico de los Postulantes

The screenshot shows a web interface for 'Bayesiano' with a navigation menu on the left containing 'Inicio', 'Postulante', 'Convocatorias', and 'Usuarios'. The main content area is titled 'Nuevo Postulante' and has a breadcrumb 'Home > Registrar Postulante'. The 'Psicotécnico' tab is active, showing a form with the following fields:

- Nombres *: David
- Apellidos *: sss
- Dni *: 12345670
- Direccion *: aa
- Aptitud Académica *
- Preguntas Relacionados al Puesto *

 At the bottom right of the form are 'Guardar' and 'Cancelar' buttons.

Fuente: Elaboración Propia

Ilustración 24: Prototipo de Visualización de Entrevistas Laborales de los Postulantes

The screenshot shows the same 'Nuevo Postulante' form, but with the 'Entrevista' tab selected. The form fields are:

- Nombres *: David
- Apellidos *: sss
- Dni *: 12345670
- Direccion *: aa
- Orientación a la Calidad *
- Tolerancia al trabajo bajo presión *
- Orientación a resultados *
- Sentido de urgencia *
- Precisión en manejo de información *

 The 'Guardar' and 'Cancelar' buttons are visible at the bottom right.

Fuente: Elaboración Propia



Tabla 13: Cuadro Comparativo de Algoritmo Computacionales

CARACTERISTICAS	NAIVE BAYES	NAIVE BAYES AUMENTADA A ARBOL (TAN)	ALGORITMO HC (HILL-CLIMBER)	C 4.5
Precisión	63%	84%	80%	73%
Sensibilidad	40%	65%	61%	46,2%
Especificidad	76%	89%	80%	72.6%
Margen de Error	37%	16%	20%	27 %

Fuentes: (Pedro Gabriel, Rodriguez, & Cruz Barboza, 2015)

(Beltran Pascual, Muñoz Martinez, & Muñoz Alamillos, 2014)

El algoritmo a utilizar en esta tesis es Naive Bayes aumentando a árbol (TAN) ya que permite inducir modelos de probabilidad realizando inferencias con los datos, nuevos valores observados y los atributos tienen relación de dependencia dado la variable clase; lo cual pretende mejorar la tasa de precisión, exactitud durante la clasificación manteniendo la simplicidad del Naive Bayes.

3.4.4.2.2. Algoritmo Bayesiano (Tan).

En esta tesis después de recopilar información acerca de los algoritmos computacionales, se eligió para cumplir con los objetivos trazados, la implementación del algoritmo TAN (que quiere del inglés Tree Augmented Naive Bayes).

En este método se realizó la construcción de una red bayesiana un poco más compleja, es decir en vez de suponer que las variables o atributos sean independientes dada la clase, en el TAN se admiten que hay cierta dependencia entre las variables atributos, lo que permite crear una red de forma de árbol.

Cuando Friedman y col. (Friedman, 1997), da a conocer al TAN, nos dice que es una adaptación al Chow-Liu. (Friedman, 1997), es decir que el asume la cantidad de información mutua condicionada a la clase, en vez de la cantidad de información mutua que se basa el algoritmo de Chow-Liu

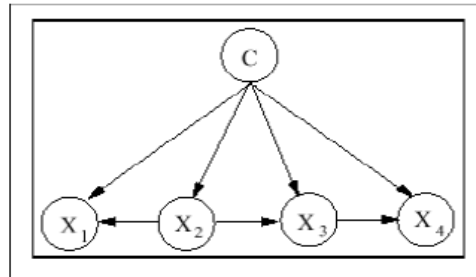
Ilustración 25 : Algoritmo Bayesiano Aumentado a Árbol

- Paso 1. Calcular $I(X_i, X_j | C)$ con $i < j, i, j = 1, \dots, n$
- Paso 2. Construir un grafo no dirigido completo cuyos nodos correspondan a las variables predictoras X_1, \dots, X_n . Asignar a cada arista conectando las variables X_i y X_j un peso dado por $I(X_i, X_j | C)$
- Paso 3. A partir del grafo completo anterior y siguiendo el algoritmo de Kruskal construir un árbol expandido de máximo peso
- Paso 4. transformar el árbol no dirigido resultante en uno dirigido, escogiendo una variable como raíz, para a continuación direccionar el resto de las aristas
- Paso 5. Construir un modelo TAN añadiendo un nodo etiquetado como C y posteriormente un arco desde C a cada variable predictora X_i

Fuente: (Artica Chacon & Ocaña Sudario, 2015)



Ilustración 26 : Ejemplo de Red Bayesiano aumentado a Árbol



Fuente : (Artica Chacon & Ocaña Sudario, 2015)

La cantidad de información mutua de obtiene mediante esta fórmula:

$$I(X, Y | C) = \sum_{i=1}^n \sum_{j=1}^m \sum_{r=1}^w p(x_i, y_j, c_r) \log \left(\frac{p(x_i, y_j | c_r)}{p(x_i | c_r) p(y_j | c_r)} \right)$$

Una vez calcula la $I(X, Y | C)$, se aplica el algoritmo de Kruskall, su fórmula es : $n(n - 1) / 2$, donde n es la cantidad de variables atributos a utilizar.

Luego se debe asignar las dos aristas de mayor peso para construir el árbol, de manera que no firma ciclos y así la siguiente.

En este algoritmo se empleará variables discretas y continuas, las cuales son:

Edad: Variable Discreta

Los valores que se fijaran para la clasificación son A: 18 - 25 años, B=25 - 35 años, C=35 - 45 años, D=45 - 55 años, E=55 - 65 años, D=65 a más. En la base de datos ya se encuentran agrupados.

Formación Académica: Variable Discreta

Los valores que se fijaran para la clasificación son: Título de Doctor, Egresado de Doctorado, Título de Magister, Egresado de Maestría, Título Universitario, Bachiller



Universitario, Título Técnico, Egresado de Universidad y/o Instituto, Estudiante de Universidad y/o Instituto, Secundaria Completa, Primaria Completa.

Capacitaciones: Variable Discreta

Los valores que se fijaran para la clasificación son las horas lectivas de capacitaciones, seminarios, cursos, talleres de los postulantes que están comprendidos por: De 12 hasta 50 horas lectivas, Entre 50 y 90 horas lectivas, Más de 90 horas lectivas.

Experiencia Laboral General: Variable continúa

Este atributo se refiere al tiempo laborado en toda su carrera profesional, para la clasificación o evaluación será en “años”, el cual se discretara en intervalos por misma amplitud.

Examen Psicotécnico: Variable Discreta

Los valores que se fijaran para la clasificación es con respecto a las notas que los postulantes obtuvieron los cuales se agrupo en categorías que son: Suspenso (de 0 a 10), Aprobado (entre 10 a 14), Notable (entre 14 a 17), Sobresaliente (entre 17 a 20).

Última Experiencia Laboral: Variable continúa

Este atributo se refiere al tiempo laborado en su último trabajo realizado, para la clasificación o evaluación será en “años”, el cual se discretara en intervalos por misma amplitud

Entrevista Laboral: Variable Discreta

Los valores que se fijaran para la clasificación es con respecto a las notas que los postulantes obtuvieron los cuales se agrupo en categorías que son: Suspenso (de 0 a 10), Aprobado (entre 10 a 14), Notable (entre 14 a 17), Sobresaliente (entre 17 a 20).



Tabla 14: Atributos y/o Variables empleado para el funcionamiento del Algoritmo TAN

Atributos o Variables	Tipo	Valores
Edad	Discreta	A=18 a 25
		B=25 a 35
		C=35 a 45
		D=45 a 55
		E=55 a 65
		D=65 a mas
Formación Académica	Discreta	Título de Doctor
		Egresado de Doctorado
		Título de Magister
		Egresado de Maestría
		Título Universitario
		Bachiller Universitario
		Título Técnico
		Egresado de Universidad y/o Instituto
		Estudiante de Universidad y/o Instituto
		Secundaria Completa
Primaria Completa		
Capacitación	Discreta	De 12 hasta 50 horas lectivas
		Entre 50 y 90 horas lectivas
		Más de 90 horas lectivas
Ultima Experiencia Laboral	Continua	años
Entrevista Laboral	Discreta	Suspense (de 0 a 10)
		Aprobado (entre 10 a 14)
		Notable (entre 14 a 17)
		Sobresaliente (entre 17 a 20)
Experiencia Laboral General	Continua	años
Examen Psicotécnico	Discreta	Suspense
		Aprobado
		Notable
		Sobresaliente

Fuente: Elaboración Propia



Y la variable clase están definidos por los siguientes valores.

Tabla 15: Variable Clase

Atributos o Variables	Tipo	Valores
Resultado	Discreta	Seleccionado No Seleccionado

Fuente: Elaboración Propia

Pasos del Algoritmo Bayesiano Aumentado a Árbol (Tan)

Paso 1.

Tenemos una base de datos con 251 postulantes de anteriores procesos de selección en el cual en base a esos se procederá a calcular la probabilidad de clasificación de nuevos postulantes.

Donde

$n = 7$ Variables (Atributos)

Peso = $7(7-1)/2 = 21$ pares

Calculamos la cantidad de información mutua de todas las variables de todos los pares de atributos condicionadas a la clase (Seleccionado), mediante la siguiente fórmula:

$$I(X, Y / C) = H(X / C) + H(Y / C) - H(X, Y / C)$$

Dónde: X = variable o atributo 1, Y = variable atributo 2 y C = variable clase

$H(X / C)$ = entropía condicional de la variable o atributo 1

$H(Y / C)$ = entropía condicional de la variable o atributo 2

$H(X, Y / C)$ = entropía conjunta de ambos atributos dada la clase

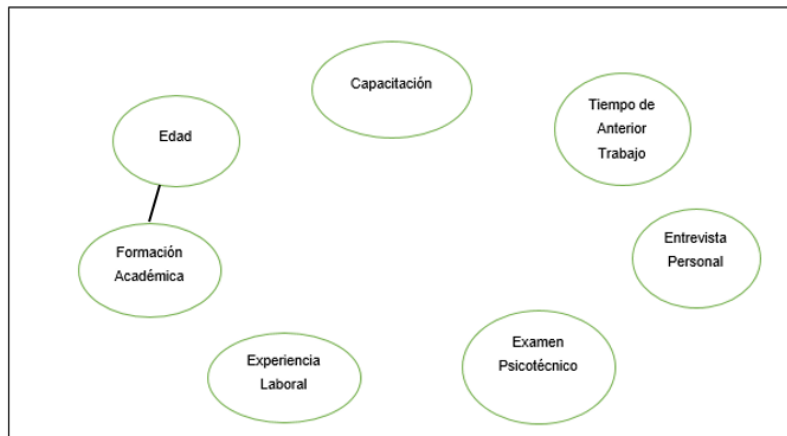
Veamos cómo se calcula la cantidad de información mutua del primer par de atributos.



Para calcular la cantidad de información mutua de las variables Edad y Formación Académica condicionadas a la clase Seleccionado / No Seleccionado. Primero tenemos que calcular las probabilidades de cada variable.

Dónde: X = edad; Y = Formación académica y C = Variable Clase

Ilustración 27 : Variables: Edad-Formación Académica



Fuente: Elaboración Propia

a) Cálculo de las distribuciones de probabilidad cuando la variables en discreta

- Distribución de la probabilidad de la Variable Clase C,

De un total de 251 postulantes según la base de datos, calculamos las probabilidades de la Clase C

$$P(C) = \frac{Cantidad}{Total\ de\ Postulantes}$$

Tabla 16 : Distribución de Probabilidad de la Variable Clase

Clase(C)	Valores	Notación(C)	Cantidad	Probabilidad
Resultado	Seleccionado	C_1	120	0.478
	No Seleccionado	C_2	131	0.522

Fuente: Elaboración Propia



- Distribución de tabla de conteo y probabilidad de la variable : Edad

Tabla 17 : Conteo de la Variable edad condicionada a la clase

Valores	Notación	C ₁	C ₂	Cantidad
A	A ₁ = a ₁₁	15	22	37
B	A ₁ = a ₁₂	17	28	45
C	A ₁ = a ₁₃	30	13	43
D	A ₁ = a ₁₄	19	23	42
E	A ₁ = a ₁₅	19	21	40
F	A ₁ = a ₁₆	20	24	44
Total		120	131	251

Fuente: Elaboración Propia

Aplicamos la probabilidad de cada valor del atributo condicionada a la variable Clase.

$$P(A_1/ C) = \frac{P(A_1 \cap C)}{P(C)}$$

Tabla 18 : Distribución de la Probabilidad de la Variable Edad condicionada a la Variable Clase

Clase	P(a ₁₁ /C)	P(a ₁₂ /C)	P(a ₁₃ /C)	P(a ₁₄ /C)	P(a ₁₅ /C)	P(a ₁₆ /C)
C ₁	0.125	0.141666667	0.25	0.158333333	0.158333333	0.166666667
C ₂	0.167938931	0.213740458	0.099236641	0.175572519	0.160305344	0.183206107

Fuente: Elaboración Propia

- Distribución de tabla de conteo y probabilidad de la variable: Formación Académica

Tabla 19 : Conteo de la Variable Formación Académica condicionada a la Variable Clase

Valores	Notación	C ₁	C ₂	Cantidad
Título de Doctor	A ₂ = a ₂₁	12	11	23
Egresado de Doctorado	A ₂ = a ₂₂	7	11	18



Título de Magister	$A_2 = a_{23}$	13	10	23
Egresado de Maestría	$A_2 = a_{24}$	5	8	13
Título Universitario	$A_2 = a_{25}$	21	18	39
Bachiller Universitario	$A_2 = a_{26}$	11	15	26
Título Técnico	$A_2 = a_{27}$	20	13	33
Egresado de Universidad y/o Instituto	$A_2 = a_{28}$	6	11	17
Estudiante de Universidad y/o Instituto	$A_2 = a_{29}$	2	4	6
Secundaria completa	$A_2 = a_{2A}$	10	13	23
Primaria completa	$A_2 = a_{2B}$	13	17	30
Total		120	131	251

Fuente: Elaboración Propia

Aplicamos la probabilidad de cada valor del atributo condicionada a la variable Clase

$$P(A_2/C) = \frac{P(A_2 \cap C)}{P(C)}$$

Tabla 20 : Distribución de la Variable Formación Académica condicionada a la variable clase.

Probabilidades	C_1	C_2
$P(a_{21}/C)$	0.1	0.08396946
$P(a_{22}/C)$	0.058333333	0.083969466
$P(a_{23}/C)$	0.108333333	0.076335878
$P(a_{24}/C)$	0.041666667	0.061068702
$P(a_{25}/C)$	0.175	0.13740458
$P(a_{26}/C)$	0.091666667	0.114503817
$P(a_{27}/C)$	0.166666667	0.099236641
$P(a_{28}/C)$	0.05	0.083969466



$P(a_{29}/C)$	0.016666667	0.030534351
$P(a_{2A}/C)$	0.083333333	0.099236641
$P(a_{2B}/C)$	0.108333333	0.129770992

Fuente: Elaboración Propia

- **Distribución de tabla de conteo y probabilidad de la variable: Edad- Formación Académica condicionada a las variables clase, donde**

$$P(A_1 \cap A_2 / C) = \frac{P(A_1 \cap A_2 \cap C)}{P(C)}$$

Tabla 21 : Distribución de la probabilidad de la variable Edad y Formación Académica condicionada a la variable clase.

$P(A_1 \cap A_2)$	Cantidad			Probabilidades	
	C_1	C_2	Total	$P(A_1 \cap A_2 / c_1)$	$P(A_1 \cap A_2 / c_2)$
$P(a_{11} \cap a_{21})$	0	0	0	0	0
$P(a_{11} \cap a_{22})$	0	0	0	0	0
$P(a_{11} \cap a_{23})$	0	0	0	0	0
$P(a_{11} \cap a_{24})$	0	0	0	0	0
$P(a_{11} \cap a_{25})$	5	2	7	0.042016807	0.015267176
$P(a_{11} \cap a_{26})$	1	4	5	0.008403361	0.030534351
$P(a_{11} \cap a_{27})$	2	4	6	0.016806723	0.030534351
$P(a_{11} \cap a_{28})$	3	4	7	0.025210084	0.030534351
$P(a_{11} \cap a_{29})$	1	1	2	0.008403361	0.007633588
$P(a_{11} \cap a_{2A})$	0	4	4	0	0.030534351
$P(a_{11} \cap a_{2B})$	3	3	6	0.025210084	0.022900763
$P(a_{12} \cap a_{21})$	1	2	3	0.008403361	0.015267176
$P(a_{12} \cap a_{22})$	0	2	2	0	0.015267176
$P(a_{12} \cap a_{23})$	2	4	6	0.016806723	0.030534351
$P(a_{12} \cap a_{24})$	0	3	3	0	0.022900763
$P(a_{12} \cap a_{25})$	2	5	7	0.016806723	0.038167939



$P(a_{12} \cap a_{26})$	0	1	1	0	0.007633588
$P(a_{12} \cap a_{27})$	6	0	6	0.050420168	0
$P(a_{12} \cap a_{28})$	1	6	7	0.008403361	0.045801527
$P(a_{12} \cap a_{29})$	1	2	3	0.008403361	0.015267176
$P(a_{12} \cap a_{2A})$	3	2	5	0.025210084	0.015267176
$P(a_{12} \cap a_{2B})$	1	1	2	0.008403361	0.007633588
$P(a_{13} \cap a_{21})$	5	1	6	0.042016807	0.007633588
$P(a_{13} \cap a_{22})$	0	1	1	0	0.007633588
$P(a_{13} \cap a_{23})$	6	2	8	0.050420168	0.015267176
$P(a_{13} \cap a_{24})$	1	1	2	0.008403361	0.007633588
$P(a_{13} \cap a_{25})$	4	1	5	0.033613445	0.007633588
$P(a_{13} \cap a_{26})$	2	2	4	0.016806723	0.015267176
$P(a_{13} \cap a_{27})$	4	0	4	0.033613445	0
$P(a_{13} \cap a_{28})$	2	1	3	0.016806723	0.007633588
$P(a_{13} \cap a_{29})$	0	1	1	0	0.007633588
$P(a_{13} \cap a_{2A})$	4	0	4	0.033613445	0
$P(a_{13} \cap a_{2B})$	2	3	5	0.016806723	0.022900763
$P(a_{14} \cap a_{21})$	3	1	4	0.025210084	0.007633588
$P(a_{14} \cap a_{22})$	3	3	6	0.025210084	0.022900763
$P(a_{14} \cap a_{23})$	3	2	5	0.025210084	0.015267176
$P(a_{14} \cap a_{24})$	0	1	1	0	0.007633588
$P(a_{14} \cap a_{25})$	2	4	6	0.016806723	0.030534351
$P(a_{14} \cap a_{26})$	0	4	4	0	0.030534351
$P(a_{14} \cap a_{27})$	2	4	6	0.016806723	0.030534351
$P(a_{14} \cap a_{28})$	0	0	0	0	0
$P(a_{14} \cap a_{29})$	0	0	0	0	0
$P(a_{14} \cap a_{2A})$	1	2	3	0.008403361	0.015267176
$P(a_{14} \cap a_{2B})$	5	2	7	0.042016807	0.015267176
$P(a_{15} \cap a_{21})$	0	4	4	0	0.030534351
$P(a_{15} \cap a_{22})$	4	1	5	0.033613445	0.007633588
$P(a_{15} \cap a_{23})$	1	2	3	0.008403361	0.015267176
$P(a_{15} \cap a_{24})$	1	3	4	0.008403361	0.022900763
$P(a_{15} \cap a_{25})$	6	0	6	0.050420168	0



$P(a_{15} \cap a_{26})$	2	1	3	0.016806723	0.007633588
$P(a_{15} \cap a_{27})$	3	2	5	0.025210084	0.015267176
$P(a_{15} \cap a_{28})$	0	0	0	0	0
$P(a_{15} \cap a_{29})$	0	0	0	0	0
$P(a_{15} \cap a_{2A})$	1	3	4	0.008403361	0.022900763
$P(a_{15} \cap a_{2B})$	1	5	6	0.008403361	0.038167939
$P(a_{16} \cap a_{21})$	3	3	6	0.025210084	0.022900763
$P(a_{16} \cap a_{22})$	0	4	4	0	0.030534351
$P(a_{16} \cap a_{23})$	1	0	1	0.008403361	0
$P(a_{16} \cap a_{24})$	3	0	3	0.025210084	0
$P(a_{16} \cap a_{25})$	4	5	9	0.033613445	0.038167939
$P(a_{16} \cap a_{26})$	4	4	8	0.033613445	0.030534351
$P(a_{16} \cap a_{27})$	3	3	6	0.025210084	0.022900763
$P(a_{16} \cap a_{28})$	0	0	0	0	0
$P(a_{16} \cap a_{29})$	0	0	0	0	0
$P(a_{16} \cap a_{2A})$	1	2	3	0.008403361	0.015267176
$P(a_{16} \cap a_{2B})$	1	3	4	0	0.022900763
TOTAL	120	131	251		

Fuente: Elaboración Propia

b) Cálculo de las distribuciones de probabilidad cuando las variables en continua.

Por ejemplo, la variable **Última Experiencia Laboral** es de tipo continua, entonces la discretizamos usando el método de igual anchura.

$$y = \text{Ultima Experiencia Laboral}$$

Pasos para la agrupación en intervalos de clase de igual amplitud:

- ✓ Se calcula el recorrido de las observaciones

$$R = Y_{max} - Y_{min} = 12 - 0.15 = 11.85$$

- ✓ El número de intervalos de clase que podemos tomar para agrupar los datos es: $k = 1 + 3,322 \log n$; donde n es la cantidad de datos

Método Surges: $k = 1 + 3,332 \log n$ $K = 1 + 3.322 \log 251 = 8.965956709$



Que aproximamos por el número natural siguiente: $k = 9$.

Por tanto, la amplitud de cada intervalo es:

$$d = \frac{R}{K} = \frac{11.85}{9} = 1.3166666$$

Al no ser exacta, aproximamos la cantidad anterior a un número ligeramente superior; por ejemplo, $d = 1.32$.

Como la amplitud de los intervalos la hemos tomado un poco mayor de lo que se obtiene en un principio, entonces el nuevo recorrido es:

$$R_0 = \text{Numero de Intervalos} * \text{amplitud} = 9 * 1.32 = 11.88$$

Como el recorrido original es 11.85, entonces sobra 0.03

$$0.15 - 0.03 = 0.12$$

- $0.12 + 1.32 = 1.44$
- $1.44 + 1.32 = 2.76$
- $2.76 + 1.32 = 4.08$
- $4.08 + 1.32 = 5.4$
- $5.4 + 1.32 = 6.72$
- $6.72 + 1.32 = 8.04$
- $8.04 + 1.32 = 9.36$
- $9.36 + 1.32 = 10.68$
- $10.68 + 1.32 = 12 + 0.03 = 12.03$

Finalmente, de obtiene un cuadro discretizamos la variable, luego se calcula las probabilidades

Tabla 22 : *Distribución de la Variable Ultima Experiencia Laboral condicionada a la variable clase.*

Intervalos	Notación (A)	Cantidad	C ₁	C ₂
0.12-1.44	$A_4 = a_{41}$	35	15	20
1.44-2.76	$A_4 = a_{42}$	36	21	15



2.76-4.08	$A_4 = a_{43}$	29	18	11
4.08-5.4	$A_4 = a_{44}$	41	13	29
5.4-6.72	$A_4 = a_{45}$	30	14	16
6.72-8.04	$A_4 = a_{46}$	32	13	19
8.04-9.36	$A_4 = a_{47}$	17	9	8
9.36-10.68	$A_4 = a_{48}$	22	14	8
10.68-12.03	$A_4 = a_{49}$	9	3	6
TOTAL		251		

Fuente: Elaboración Propia

Una vez calculadas las probabilidades de la variable clase y de los atributos: Edad y Formación académica, se calcula la Entropía condicional de cada variable:

Se calculó de la Entropía condicional de la Variable Edad $A_1 = x$, dada la clase = c:

$$\text{➤ } H(x/c) = -\sum_c p(c) \sum_x p(x/c) \cdot \log_2 p(x/c)$$

$$H(x/c) = -p(c = C_1)\{p(x = a_{11}/c = C_1) \cdot \log_2 p(x = a_{11}/c = C_1) + p(x = a_{12}/c = C_1) \cdot \log_2 p(x = a_{12}/c = C_1) + p(x = a_{13}/c = C_1) \cdot \log_2 p(x = a_{13}/c = C_1) + p(x = a_{14}/c = C_1) \cdot \log_2 p(x = a_{14}/c = C_1) + p(x = a_{15}/c = C_1) \cdot \log_2 p(x = a_{15}/c = C_1) + p(x = a_{16}/c = C_1) \cdot \log_2 p(x = a_{16}/c = C_1)\} - p(c = C_2)\{p(x = a_{11}/c = C_2) \cdot \log_2 p(x = a_{11}/c = C_2) + p(x = a_{12}/c = C_2) \cdot \log_2 p(x = a_{12}/c = C_2) + p(x = a_{13}/c = C_2) \cdot \log_2 p(x = a_{13}/c = C_2) + p(x = a_{14}/c = C_2) \cdot \log_2 p(x = a_{14}/c = C_2) + p(x = a_{15}/c = C_2) \cdot \log_2 p(x = a_{15}/c = C_2) + p(x = a_{16}/c = C_2) \cdot \log_2 p(x = a_{16}/c = C_2)\}$$

$$H(x/c) = -0.478 \{0.125 \cdot \log_2 0.125 + 0.141666667 \cdot \log_2 0.141666667 + 0.25 \cdot \log_2 0.25 + 0.158333333 \cdot \log_2 0.158333333 + 0.158333333 \cdot \log_2 0.158333333 + 0.166666667 \cdot \log_2 0.166666667\} - 0.522 \{0.167938931 \cdot \log_2 0.167938931 + 0.213740458 \cdot \log_2 0.213740458 + 0.099236641 \cdot \log_2 0.099236641 + 0.175572519 \cdot \log_2 0.175572519 + 0.160305344 \cdot \log_2 0.160305344 + 0.183206107 \cdot \log_2 0.183206107\}$$

$$H(x/c) = -0.476 \{-0.375 + -0.399418932 + -0.5 + -0.421002488 + -0.421002488 + -0.430827083\}$$



$$-0.524 \{-0.432273362 + -0.475800811 + -0.330754066 + -0.440662626 + -0.423383337 + -0.448572916\}$$

$$H(x/c) = -0.478 \{-2.547250991\} - 0.522\{-2.551447119\}$$

$$H(x/c) = 1.217809239 + 1.331631763$$

$$H(x/c) = 2.549441002 \text{ bits}$$

Se calculó de la Entropía condicional de la Variable Formación Académica $A_2 = y$, dada la clase = c.

➤ $H(y/c) = -\sum_c p(c) \sum_y p(y/c) \cdot \log_2 p(y/c)$

$$H(y/c) = -p(c = C_1) \cdot \{p(y = a_{21}/c = C_1) \cdot \log_2 p(y = a_{21}/c = C_1) + p(y = a_{22}/c = C_1) \cdot \log_2 p(y = a_{22}/c = C_1) + p(y = a_{23}/c = C_1) \cdot \log_2 p(y = a_{23}/c = C_1) + p(y = a_{24}/c = C_1) \cdot \log_2 p(y = a_{24}/c = C_1) + p(y = a_{25}/c = C_1) \cdot \log_2 p(y = a_{25}/c = C_1) + p(y = a_{26}/c = C_1) \cdot \log_2 p(y = a_{26}/c = C_1) + p(y = a_{27}/c = C_1) \cdot \log_2 p(y = a_{27}/c = C_1) + p(y = a_{28}/c = C_1) \cdot \log_2 p(y = a_{28}/c = C_1) + p(y = a_{29}/c = C_1) \cdot \log_2 p(y = a_{29}/c = C_1) + p(y = a_{2A}/c = C_1) \cdot \log_2 p(y = a_{2A}/c = C_1) + p(y = a_{2B}/c = C_1) \cdot \log_2 p(y = a_{2B}/c = C_1)\}$$

$$-p(c = C_2) \cdot \{p(y = a_{21}/c = C_2) \cdot \log_2 p(y = a_{21}/c = C_2) + p(y = a_{22}/c = C_2) \cdot \log_2 p(y = a_{22}/c = C_2) + p(y = a_{23}/c = C_2) \cdot \log_2 p(y = a_{23}/c = C_2) + p(y = a_{24}/c = C_2) \cdot \log_2 p(y = a_{24}/c = C_2) + p(y = a_{25}/c = C_2) \cdot \log_2 p(y = a_{25}/c = C_2) + p(y = a_{26}/c = C_2) \cdot \log_2 p(y = a_{26}/c = C_2) + p(y = a_{27}/c = C_2) \cdot \log_2 p(y = a_{27}/c = C_2) + p(y = a_{28}/c = C_2) \cdot \log_2 p(y = a_{28}/c = C_2) + p(y = a_{29}/c = C_2) \cdot \log_2 p(y = a_{29}/c = C_2) + p(y = a_{2A}/c = C_2) \cdot \log_2 p(y = a_{2A}/c = C_2) + p(y = a_{2B}/c = C_2) \cdot \log_2 p(y = a_{2B}/c = C_2)\}$$

$$H(y/c) = -0.478 \{0.1 \cdot \log_2 0.1 + 0.058333333 \cdot \log_2 0.058333333 + 0.108333333 \cdot \log_2 0.108333333 + 0.041666667 \cdot \log_2 0.041666667 + 0.175 \cdot \log_2 0.175 + 0.091666667 \cdot \log_2 0.091666667 + 0.166666667 \cdot \log_2 0.166666667 + 0.05 \cdot \log_2 0.05 + 0.016666667 \cdot \log_2 0.016666667 + 0.083333333 \cdot \log_2 0.083333333 + 0.108333333 \cdot \log_2 0.108333333\}$$



$$\begin{aligned}
 & -0.522\{0.083969466 \cdot \log_2 0.083969466 + 0.083969466 \log_2 0.083969466 + \\
 & 0.076335878 \cdot \log_2 0.076335878 + 0.061068702 \cdot \log_2 0.061068702 + \\
 & 0.13740458 \cdot \log_2 0.13740458 + 0.114503817 \cdot \log_2 0.114503817 + \\
 & 0.099236641 \cdot \log_2 0.099236641 + 0.083969466 \cdot \log_2 0.083969466 + \\
 & 0.030534351 \cdot \log_2 0.030534351 + 0.099236641 \cdot \log_2 0.099236641 + \\
 & 0.129770992 \cdot \log_2 0.129770992 \} \\
 H(y/c) = & -0.478 \{-0.332192809 + -0.239139581 + -0.347365512 + -0.191040104 \\
 & + -0.440050305 + -0.316017073 + -0.430827083 + -0.216096405 \\
 & + -0.098448177 + -0.298746875 + -0.347365512\} \\
 & -0.522\{-0.30010615 + -0.30010615 + -0.28332022 + -0.24631591 + \\
 & -0.39345774 + -0.358799989 + -0.33075407 + -0.30010615 + \\
 & -0.153692305 + -0.330754066 + -0.382300173\}
 \end{aligned}$$

$$H(y/c) = -0.478 \{-3.257289436\} - 0.522\{-3.37891282\}$$

$$H(y/c) = 1.55726985 + 1.76349633$$

$$H(y/c) = 3.32076618 \text{ bits}$$

Cálculo de la Entropía conjunta de la Variable Edad $A_1 = x$, Formación Académica $A_2 = y$, dada la clase = c.

$$\text{➤ } H(x, y/c) = -\sum_c p(c) \sum_y p(x, y/c) \cdot \log_2 p(x, y/c)$$

$$\begin{aligned}
 H(x, y/c) = & -0.478\{0 \cdot \log_2 0 \dots \dots \dots + 0.008333333 \cdot \log_2 0.008333333 \\
 & -0.522\{0 \cdot \log_2 0 + \dots \dots \dots + 0.022900763 \cdot \log_2 0.022900763\}
 \end{aligned}$$

$$\begin{aligned}
 H(x, y/c) = & -0.478\{0 + \dots \dots \dots + -0.057557422\} \\
 & -0.522\{0 + \dots \dots \dots + -0.124773905\}
 \end{aligned}$$

$$H(x, y/c) = -0.478\{-5.23454444\} - 0.522\{-5.4699\}$$

$$H(x, y/c) = 5.357447024 \text{ bits}$$



Una vez calculado la entropía se aplica la siguiente fórmula para hallar la cantidad de información mutua de la variable Edad: x y Formación académica = y , condicionada a la variable Clase (Seleccionado / No Seleccionado) = c

$$I(X, Y / C) = H(X/C) + H(Y/C) - H(X, Y / C)$$

$$I(X, Y / C) = 2.549441002 + 3.32076618 - 5.357447024$$

$$I(X, Y / C) = 5.870207182 - 5.357447024$$

$$I(X, Y / C) = 0.5127 \text{ bits}$$

Posteriormente se calculó la cantidad de información de las demás pares de variables, obteniéndolos en el siguiente cuadro:

Tabla 23 : Cantidad de Información mutua de cada par de variables condicionada a la variable clase

Nro.	Pares de Variables	Cantidad de Información Mutua $I(X, Y / C)$ - bits
1	EDAD - EXP. GENERAL	1.077504
2	EXP. GENERAL - FORMACIONACADEMICA	0.575971
3	ULT. EXPERIENCIA - FORMACIONACADEMICA	0.523709
4	ULT. EXPERIENCIA - EXP. GENERAL	0.517336
5	EDAD - FORMACIONACADEMICA	0.512773
6	EDAD - ULT. EXPERIENCIA	0.434536
7	CAPACITACION - FORMACIONACADEMICA	0.266406
8	ENTREVISTA - FORMACIONACADEMICA	0.238654
9	PSICOTECNICO - FORMACIONACADEMICA	0.166579
10	ENTREVISTA - EXP. GENERAL	0.163975
11	EDAD - CAPACITACION	0.152223
12	CAPACITACION - ULT. EXPERIENCIA	0.142027
13	CAPACITACION - EXP. GENERAL	0.134885
14	PSICOTECNICO - EXP. GENERAL	0.135639



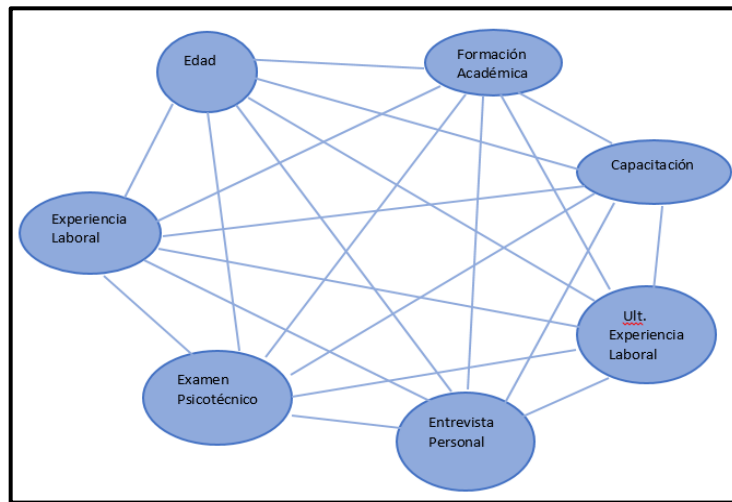
15	EDAD - ENTREVISTA	0.110261
16	ULT. EXPERIENCIA - PSICOTECNICO	0.098699
17	ULT. EXPERIENCIA - ENTREVISTA	0.095079
18	ENTREVISTA - PSICOTECNICO	0.065654
19	EDAD - PSICOTECNICO	0.061591
20	CAPACITACION - PSICOTECNICO	0.044742
21	CAPACITACION - ENTREVISTA	0.024506

Fuente: Elaboración Propia

PASO 2:

Es este paso se construye un grafo no dirigido completo donde las variables o atributos son los nodos, luego se asigna a cada arista conectado las variables cuyo peso está dado por $I(X, Y/C)$.

Ilustración 28 : Grafo no dirigido completo cuyos nodos son las variables predictorias



Fuente: Elaboración Propia

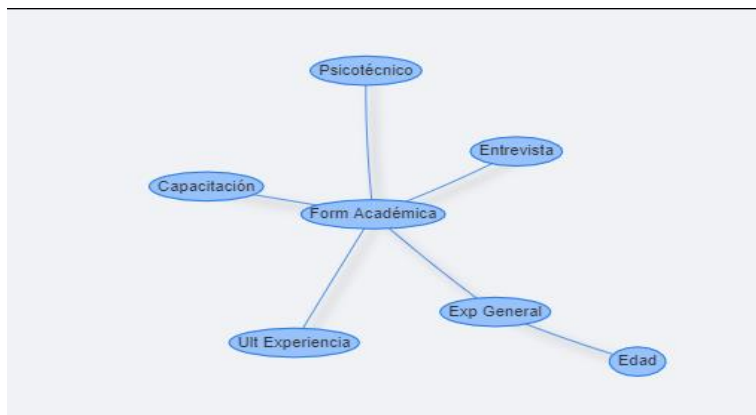
PASO 3:

Una vez teniendo el grafo del paso anterior, se aplica el algoritmo de Kruskal, donde se construye un árbol de máximo peso, luego se asigna las dos aristas de mayor peso al árbol a



crear, y añade al árbol de manera que no forme un ciclo, de ser así se descarta y se pasa a la siguiente hasta que se haya seleccionado n-1 arista.,

Ilustración 29 : Árbol no dirigido de las variables predictorias



Fuente: Elaboración Propia

PASO 4:

Se transforma el árbol no dirigido del construido en el paso anterior en uno dirigido, eligiendo una variable como la raíz, luego direccionar a las demás aristas.

Ilustración 30 : Árbol dirigido con las variables predictorias



Fuente: Elaboración Propia



PASO 5:

Al final se construye el modelo TAN, incorporando un nodo como C = variable clase y luego un arco desde C a cada variable predictorias o atributos.

EL POSTULANTE OBTIENE LOS SIGUIENTES DATOS:

- 1.- **EDAD:** Categoría D
- 2.- **CAPACITACION:** Entre 12 hasta 50 Horas Lectivas
- 3.- **EXAMEN PSICOTECNICO:** Aprobado
- 4.- **ULTIMA EXPERIENCIA LAB:** 1,75 Años
- 5.- **EXPERIENCIA LABORAL:** 5,3 Años
- 6.- **FORMACION ACADEMICA:** Título Universitario
- 7.- **ENTREVSTA PERSONAL:** Suspenso

Ilustración 31 : Modelo Algoritmo TAN con las variables predictorias y con la variable clase



Fuente: Elaboración Propia

Donde le denotamos las variables predictorias y la Variable clase de la siguiente manera:



Edad = A_1

Formación Académica = A_2

Capacitación = A_3

Ultima Experiencia Laboral = A_4

Entrevista Personal = A_5

Experiencia Laboral General = A_6

Examen Psicotécnico = A_7

Variable clase (Seleccionado / No Seleccionado) = C

El modelo clasificatorio obtenido es lo siguiente:

$$P(C / A_1, A_2, A_3, A_4, A_5, A_6, A_7) = p(c) \cdot p(A_1/c) \cdot p(A_6/A_1, c) \cdot p(A_3/A_2, c) \cdot p(A_2/A_6, c) \cdot p(A_7/A_2, c) \cdot p(A_4/A_2, c) \cdot p(A_5/A_2, c)$$

Esta estimación aparenta ser una buena pero puede suceder que durante las combinaciones los valores de las variables puedan ser ceros, entonces poder que disminuya ese problema, se puede aplicar los estimadores como por ejemplo la sucesión de Laplace, el cual se obtiene así:

$$\frac{N + 1}{M + r}$$

Para calcular la probabilidad se de tener en cuenta el número de casos favorables + 1, entre los casos totales más el número de alternativas del otros atributo (r) .

Donde:

N = número de casos de la variable X_i



M= número de casos de la variable X_j

r = número de posibles valores de X_i

Aplicando la fórmula de probabilidad general, condicional y la sucesión de Laplace calculamos:

a) Para la Variable Clase $C_1 =$ Seleccionado

$$p(C_1) = \frac{\text{Total de Seleccionados}}{\text{Total General}} = \frac{120}{251} = \frac{120 + 1}{251 + 2} = \frac{121}{253} = 0.4783$$

$$p(A_1/C_1) = \frac{A_1 \cap C_1}{C_1} = \frac{19}{120} = \frac{19 + 1}{120 + 6} = \frac{20}{126} = 0.158730$$

$$p(A_6/A_1, C_1) = \frac{A_6 \cap A_1 \cap C_1}{A_1 \cap C_1} = \frac{0}{19} = \frac{0 + 1}{19 + 9} = \frac{1}{28} = 0.03571$$

$$p(A_3/A_2, C_1) = \frac{A_3 \cap A_2 \cap C_1}{A_2 \cap C_1} = \frac{2}{22} = \frac{2 + 1}{22 + 3} = \frac{3}{25} = 0.12$$

$$p(A_2/A_6, C_1) = \frac{A_2 \cap A_6 \cap C_1}{A_6 \cap C_1} = \frac{6}{18} = \frac{5 + 1}{18 + 11} = \frac{7}{29} = 0.241379$$

$$p(A_7/A_2, C_1) = \frac{A_7 \cap A_2 \cap C_1}{A_2 \cap C_1} = \frac{4}{22} = \frac{4 + 1}{22 + 4} = \frac{5}{26} = 0.1923$$

$$p(A_4/A_2, C_1) = \frac{A_4 \cap A_2 \cap C_1}{A_2 \cap C_1} = \frac{4}{22} = \frac{4 + 1}{22 + 9} = \frac{5}{31} = 0.161290$$



$$p(A_5/A_2, C_1) = \frac{A_5 \cap A_2 \cap C_1}{A_2 \cap C_1} = \frac{1}{22} = \frac{1+1}{22+4} = \frac{2}{26} = 0.076923$$

b) Para la Variable Clase $C_2 =$ No Seleccionado

$$p(C_2) = \frac{\text{Total de No Seleccionados}}{\text{Total General}} = \frac{131}{251} = \frac{131+1}{251+2} = \frac{132}{253} = 0.5217$$

$$p(A_1/C_2) = \frac{A_1 \cap C_2}{C_2} = \frac{23}{131} = \frac{23+1}{131+6} = \frac{24}{137} = 0.1752$$

$$p(A_6/A_1, C_2) = \frac{A_6 \cap A_1 \cap C_2}{A_1 \cap C_2} = \frac{0}{23} = \frac{0+1}{23+9} = \frac{1}{32} = 0.03125$$

$$p(A_3/A_2, C_2) = \frac{A_3 \cap A_2 \cap C_2}{A_2 \cap C_2} = \frac{2}{18} = \frac{2+1}{18+3} = \frac{3}{21} = 0.14285$$

$$p(A_2/A_6, C_2) = \frac{A_2 \cap A_6 \cap C_2}{A_6 \cap C_2} = \frac{1}{17} = \frac{1+1}{17+11} = \frac{2}{28} = 0.071428$$

$$p(A_7/A_2, C_2) = \frac{A_7 \cap A_2 \cap C_2}{A_2 \cap C_2} = \frac{7}{18} = \frac{7+1}{18+4} = \frac{8}{22} = 0.3636$$

$$p(A_4/A_2, C_2) = \frac{A_4 \cap A_2 \cap C_2}{A_2 \cap C_2} = \frac{2}{18} = \frac{2+1}{18+9} = \frac{3}{27} = 0.11111$$

$$p(A_5/A_2, C_2) = \frac{A_5 \cap A_2 \cap C_2}{A_2 \cap C_2} = \frac{7}{18} = \frac{7+1}{18+4} = \frac{8}{22} = 0.363636$$



Reemplazamos los valores obtenidos en:

$$P(C_1 / A_1, A_2, A_3, A_4, A_5, A_6, A_7) = p(C_1) \cdot p(A_1/C_1) \cdot p(A_6/A_1, C_1) \cdot p(A_3/A_2, C_1) \cdot p(A_2/A_6, C_1) \cdot p(A_7/A_2, C_1) \cdot p(A_4/A_2, C_1) \cdot p(A_5/A_2, C_1)$$

$$0.4783 * 0.158730 * 0.03571 * 0.12 * 0.241379 * 0.1923 * 0.161290 * 0.076923$$

$$P(C_1 / A_1, A_2, A_3, A_4, A_5, A_6, A_7) = 0.000000187262587105245$$

$$P(C_2 / A_1, A_2, A_3, A_4, A_5, A_6, A_7) = p(C_2) \cdot p(A_1/C_2) \cdot p(A_6/A_1, C_2) \cdot p(A_3/A_2, C_2) \cdot p(A_2/A_6, C_2) \cdot p(A_7/A_2, C_2) \cdot p(A_4/A_2, C_2) \cdot p(A_5/A_2, C_2)$$

$$0.5217 * 0.1752 * 0.03125 * 0.14285 * 0.071428 * 0.3636 * 0.11111 * 0.363636$$

$$P(C_2 / A_1, A_2, A_3, A_4, A_5, A_6, A_7) = 0.00000042706871653682$$

Normalizamos:

$$S = P(C_1 / A_1, A_2, A_3, A_4, A_5, A_6, A_7) + P(C_2 / A_1, A_2, A_3, A_4, A_5, A_6, A_7)$$

$$S = 0.000000614331303642060$$

Obtenemos las probabilidades:

$C_1 =$ Que sea Seleccionado

$$P(C_1 / A_1, A_2, A_3, A_4, A_5, A_6, A_7) = \frac{0.000000187262587105245}{0.000000614331303642060} = 0.304 * 100 = 30.4 \%$$

$C_2 =$ Que NO sea Seleccionado

$$P(C_2 / A_1, A_2, A_3, A_4, A_5, A_6, A_7) = \frac{0.00000042706871653682}{0.000000614331303642060} = 0.695 * 100 = 69.6\%$$



3.3.4.1.3. Implementación

La arquitectura del software que se utilizó es el MODELO VISTA CONTROLADOR (MVC), porque tiene muchos parámetros como la reutilización de código, estructurar programas importante en todo desarrollo de software para facilitar el mantenimiento y desarrollo del mismo

Se programó en lenguaje de Programación PHP y base de datos MySQL.

Ilustración 32: Arquitectura del Software

Nombre	Tamaño	Modificado
conexion		13/09/2018 9:49:01 p. m.
controller		6/08/2018 9:28:06 p. m.
model		8/08/2018 10:55:50 p. m.
view		8/08/2018 11:01:13 p. m.
admin.php	1 KB	15/09/2018 4:45:09 p. m.
index.php	1 KB	25/06/2016 10:57:28 p. m.
		22/07/2018 12:01:26 p. m.

Fuente: Elaboración Propia

Carpetas de la Arquitectura del Software

Las carpetas descritas contienen los siguientes.

- **conexion:** contiene el archivo de conexión a la BD
- **controller:** controladores de acciones que resuelven las peticiones del Front-end
- **model:** contiene los archivos que usa cada controlador, y que permiten crear transacciones y consultas a BD.
- **view:** contiene los archivos Front-end, así como los archivos JavaScript.



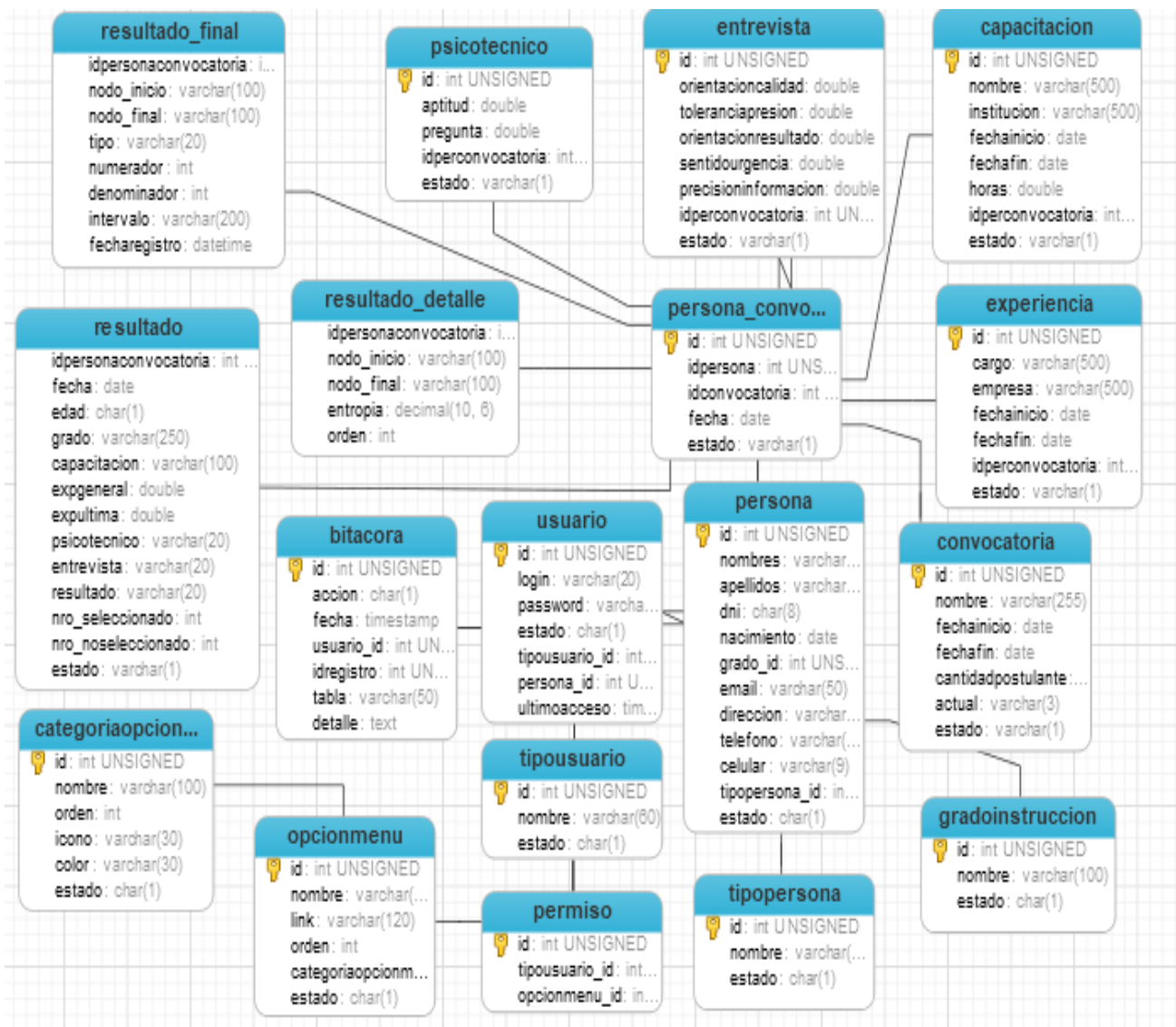
Ilustración 33: Código Fuente de la Conexión a la base de datos

```
// Constructor de la clase
function __construct(){
    global $gTipoBD;
    $this->gTipoBD=$gTipoBD;
    if($this->gTipoBD==1){
        $this->gServidor = "localhost";
    }
    if($this->gTipoBD==2){
        $this->gServidor = "localhost";
        $user='root';
        $pass='';
    }
    if($this->gTipoBD==3){
        $this->gServidor = "localhost";
        // $user='mampuerto';//sica
        // $pass='tBZPD*W#eAE';//sicaWeb2013
        $user='postgres';
        $pass='postgres';
    }
}
// Crea una conexión SQLSERVER 2000
//echo "Antes de conectar";
try {
    //print_r(PDO::getAvailableDrivers());
    if(isset($cnx) && $cnx){
        $this->gCnx = $cnx;
    }else{
        if($this->gTipoBD==1){
            $this->gCnx = new PDO("mysql:host=".$this->gServidor.";dbname=".$this->gBaseDatos,$user,$pass);
        }
        if($this->gTipoBD==2){
            $this->gCnx = new PDO("mysql:host=".$this->gServidor.";port=3306;dbname=".$this->gBaseDatos,$user,$pass);
            //$this->gCnx->setAttribute(PDO::MYSQL_ATTR_USE_BUFFERED_QUERY, true);
        }
        if($this->gTipoBD==3){
            if($this->gTipoConex==1){
                $this->gCnx = new PDO("pgsql:host=".$this->gServidor.";port=5432;dbname=".$this->gBaseDatos,$user,$pass);
            }else{
                echo "Antes de Post";
                $this->gCnx = new PDO("pgsql:host=".$this->gServidor.";port=5432;dbname=".$this->gBaseDatos,$user,$pass);
                echo $this->gCnx;
                echo "Inicia:.....".$this->gCnx->errorInfo();
            }
        }
    }
}
// $this->gCnx->setAttribute(PDO::ATTR_ERRMODE, PDO::ERRMODE_WARNING);
// $PDO->setAttribute(PDO::ATTR_PERSISTENT, true);
} catch (PDOException $e) {
```

Fuente: Elaboración Propia



Ilustración 34: Diseño de la Base de datos



Fuente: Elaboración Propia



Ilustración 35: Código Fuente del cálculo de las probabilidades

```

150     for ($i=0; $i < count($c_ordenado); $i++) {
151         $archivo_xml = $archivo_xml."<item><inicio>".$c_ordenado[$i]["INICIO"]."</inicio><final>".$c_ordenado[$i]["FINAL"]."</final></item>";
152     }
153
154     //var_dump($archivo_xml." /// ".$cantidad_xml);
155
156     $resultado=$objEvaluacion->insertarResultadoFinal($idperconvocatoria, $archivo_xml, $cantidad_xml);
157
158     $resultadofinal=$objEvaluacion->listarResultadoFinal($idperconvocatoria);
159
160     $seleccionado_inicial=1;
161     $noseleccionado_inicial=1;
162     $seleccionado_final=0;
163     $noseleccionado_final=0;
164
165     foreach ($resultadofinal as $dato) {
166         if($dato["tipo"]=="SELECCIONADO"){
167             $seleccionado_inicial = $seleccionado_inicial*$dato["resultado"];
168         }
169         if($dato["tipo"]=="NO SELECCIONADO"){
170             $noseleccionado_inicial = $noseleccionado_inicial*$dato["resultado"];
171         }
172     }
173
174     $seleccionado_final = ( $seleccionado_inicial/($seleccionado_inicial+$noseleccionado_inicial) )*100;
175     $noseleccionado_final = ( $noseleccionado_inicial/($seleccionado_inicial+$noseleccionado_inicial) )*100;
176
177     $decision = "";
178     if($seleccionado_final>=$noseleccionado_final){
179         $decision="SELECCIONADO";
180     }else{
181         $decision="NO SELECCIONADO";
182     }
183
184     $actualizarresultado=$objEvaluacion->actualizarResultado($idperconvocatoria, $seleccionado_final, $noseleccionado_final, $decision);
185
186     $response['inserted'] = true;
187
188 }
189 $objEvaluacion->finalizarTransaccion();
190
191 $response['RPTA'] = $resultado;
192 // $response['XML'] = $archivo_xml;
193 // $response['ordenado'] = $c_ordenado;
194
195

```

Fuente: Elaboración Propia

Todos los procesos para establecer la solución del problema planteado que permitirán saber acerca del comportamiento de los postulantes de durante la convocatoria de empleo en la institución.

El trabajo de investigación se basará en la información de los postulantes; que está constituida por los siguientes:

Lista de postulantes

Lista de entrevistas.

Lista de evaluaciones psicotécnicas

Lista de evaluaciones de CV.

Cuando el número de datos aumenta, la tasa de certeza será más alta, esto porque el método de Bayes utiliza la teoría de probabilidad.

El algoritmo a utilizar en esta tesis es Naive Bayes Aumentado a Árbol (TAN) donde permite realizar modelos probabilísticos, el cual los atributos tienen relación de dependencia dado la clase.

Es una técnica de predicción y clasificación supervisada, que crea modelos para predecir las probabilidades de resultados posibles. Al ser una técnica supervisada necesita de ejemplos para que funcionen, es decir datos históricos.

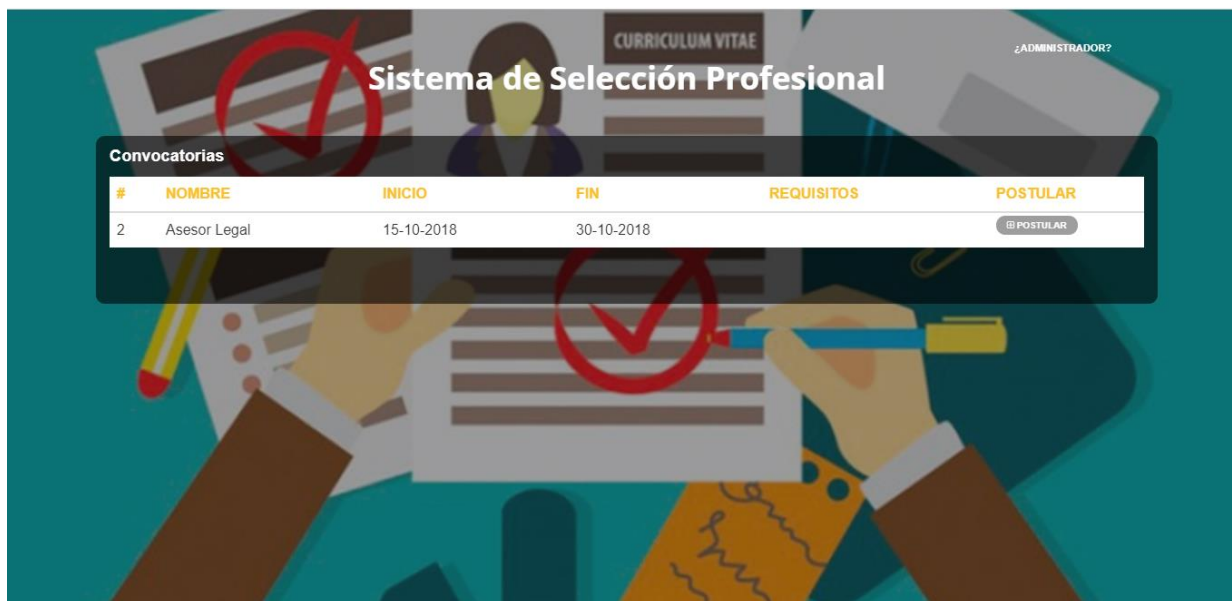
A continuación detallaremos paso a paso como funciona este Algoritmo mediante un sistema.



Menú principal del Sistema:

Pantalla donde se visualiza las convocatorias vigentes

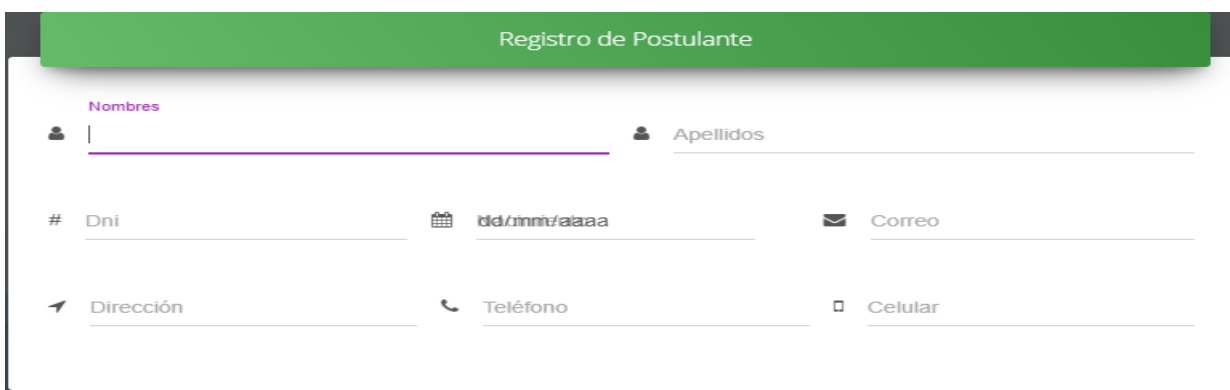
Ilustración 36: Interfaz de Pantalla Principal



Fuente: Elaboración Propia

En la Opción **POSTULAR** cada postulante ingresara y visualizará los formularios de registro de sus Datos y los requisitos del Perfil de Puesto

Ilustración 37: Interfaz de Registro de Datos Personales del Postulante



Fuente: Elaboración Propia

Registro de datos personales de postulantes, donde se ingresan datos de todos los postulantes por única vez, como son Nombres, Apellidos, Dni, Fecha de nacimiento, correo electrónico, Dirección Teléfono y Celular.

Ilustración 38: Interfaz de Registro de Formación Académica y Cursos o Capacitaciones

The screenshot shows a web interface titled "Formación Académica". It features a form with the following fields:

- Grado Instrucción:** A dropdown menu with a person icon and a downward arrow.
- Capacitaciones:** A section header.
- Capacitación:** A text input field with an envelope icon.
- Institución:** A text input field with an envelope icon.
- Inicio:** A date input field with a calendar icon and the placeholder "dd/mm/aaaa".
- Fin:** A date input field with a calendar icon and the placeholder "dd/mm/aaaa".
- Horas:** A text input field with a person icon.

Below the form is an orange button labeled "+ Agregar Capacitación". At the bottom, a table header is visible with the following columns: #, Nombre, Institución, Inicio, Fin, Horas, Acciones.

Fuente: Elaboración Propia



En este formulario el postulante Registrara su formación académica que pueden ser los siguientes:

Tabla 24: Ítems de la Evaluación Curricular

Nº	Formación Académica
1	Título de Doctor
2	Egresado de Doctorado
3	Título de Magister
4	Egresado de Maestría
5	Título Universitario
6	Bachiller Universitario
7	Título Técnico
8	Egresado de Universidad y/o Instituto
9	Estudiante de Universidad y/o Instituto
10	Secundaria Completa
11	Primaria Completa

Fuente: Elaboración Propia

Además, también registrara los cursos y/o capacitaciones que tenga con las fechas de inicio y fin y las horas respectivas, que serán agregadas con la Opción

+ Agregar Capacitación

Ilustración 39: Interfaz de Registro de Experiencia Laboral de los Postulantes

Experiencia Laboral

 Cargo

 Empresa

 dd/mm/aaaa


 dd/mm/aaaa

+ Agregar Experiencia

#	Cargo	Empresa	Inicio	Fin	Acciones



Fuente: Elaboración Propia

En este formulario el postulante registrara su experiencia laboral con el cargo que tuvo el nombre de la empresa, fecha de inicio y fecha fin (que calculara el tiempo en años experiencia) que serán agregadas con las Opción  .

Adicionalmente el sistema también tomara en cuenta la Última experiencia Laboral para la clasificación ya que es un atributo.

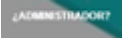
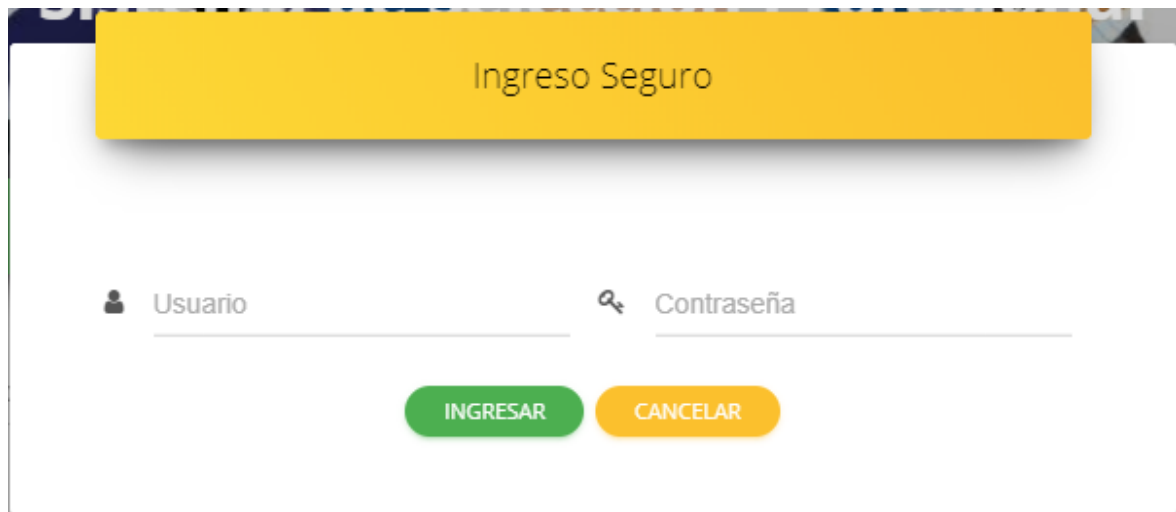
Una vez que los postulantes se hayan registrado, el encargado de la selección podrá ingresar al sistema a través de la opción  (Administrador) y visualizará el siguiente formulario.

Ilustración 40: Interfaz de Ingreso al Sistema



Fuente: Elaboración Propia

El encargado del proceso de Selección ingresara al sistema para visualizar la convocatoria y los postulantes registrados.

El encargado del proceso de selección podrá cerrar la convocatoria a través de la opción




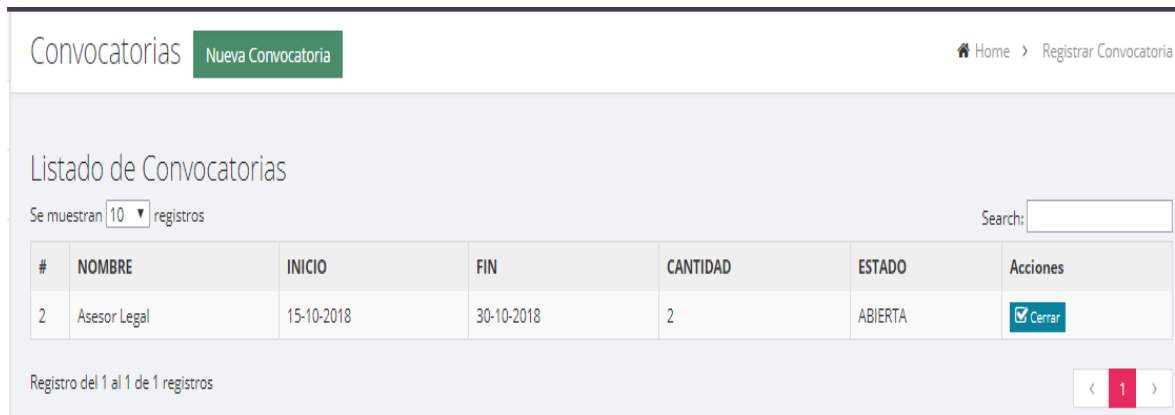
para que se comience a evaluar a través de la opción  .


Ilustración 41. Visualización de Convocatorias



Convocatorias [Nueva Convocatoria](#) Home > Registrar Convocatoria

Listado de Convocatorias

Se muestran registros Search:

#	NOMBRE	INICIO	FIN	CANTIDAD	ESTADO	Acciones
2	Asesor Legal	15-10-2018	30-10-2018	2	ABIERTA	

Registro del 1 al 1 de 1 registros

Fuente: Elaboración Propia


Al momento de dar clic en la opción evaluar  para poder ingresar

Ilustración 42: Listado de Postulantes



Listado de Postulantes

Se muestran registros Search:

DNI	APELLIDOS Y NOMBRES	NACIMIENTO	GRADO INSTRUCCIÓN	EMAIL	% SELEC.	% NO SELEC.	RESULTADO	
12345670	DIAZ DAVID	1972-05-14	TITULO UNIVERSITARIO	david_29@hotmail.com	99.488	0.512	SELECCIONADO	
45142563	SAENZ RAUL	1990-02-10	TITULO DE MAGISTER	raul@hotmail.com	98.383	1.617	SELECCIONADO	
48534851	CAMPOS ERIKA	1993-07-25	TITULO UNIVERSITARIO	erika_23@hotmail.com				

Fuente: Elaboración Propia

El encargado de la Selección Profesional al dar click en la opción evaluar, primero visualizara lo registrado por el postulante como son su curriculum, sus capacitaciones y sus experiencias laborales.



Ilustración 43: Visualización de Datos Personales del Postulante

Curriculum Capacitaciones Experiencia Psicotécnico Entrevista

Nombres * ERIKA Apellidos * CAMPOS

Dni * 48534851 Nacimiento * 1993-07-25

Email * erika_23@hotmail.com Direccion * calle el sol 397

Edad 25 Grado TITULO UNIVERSITARIO

Fuente: Elaboración Propia

Ilustración 44: Visualización de Cursos y Capacitaciones del Postulante

Curriculum **Capacitaciones** Experiencia Psicotécnico Entrevista

Nombres * ERIKA Apellidos * CAMPOS

Dni * 48534851 Direccion * calle el sol 397

Capacitaciones

Nombre	Institución	Inicio	Fin	Horas
SIGA	SBN	2018-07-10	2018-08-08	60.00

Fuente: Elaboración Propia



Ilustración 45: Visualización de Experiencia Laboral del Postulante

Curriculum Capacitaciones **Experiencia** Psicotécnico Entrevista

Nombres * ERIKA Apellidos * CAMPOS

Dni * 48534851 Direccion * calle el sol 397

Experiencias

Cargo	Empresa	Inicio	Fin	Tiempo
ANALISTA	SAGA FALABELLA	2009-01-01	2009-08-06	0.58
JEFE DE RRHH	RIPLEY	2009-09-01	2010-03-30	0.50
ASISTENTE	MINISTERIO DE SALUD	2010-05-05	2015-05-05	5.00

Fuente: Elaboración Propia

En la pestaña de Psicotécnico el encargado de la selección profesional ingresara el puntaje de las evaluaciones (aptitud académica y de las preguntas relacionadas al puesto)

Ilustración 46: Visualización de Examen Psicotécnico de los Postulantes

Curriculum Capacitaciones Experiencia **Psicotécnico** Entrevista

Nombres * ERIKA Apellidos * CAMPOS

Dni * 48534851 Direccion * calle el sol 397

Aptitud Académica *

Preguntas Relacionados al Puesto *

Fuente: Elaboración Propia

En la pestaña de Entrevista el encargado de la selección profesional ingresara el puntaje de las evaluaciones correspondientes a la entrevista.

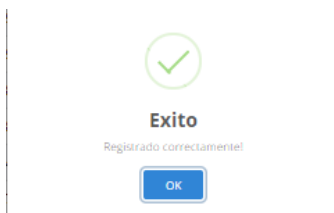


Ilustración 47: Visualización de Entrevista Personal de los Postulantes

The screenshot shows a web interface with a navigation bar at the top containing tabs for 'Currículum', 'Capacitaciones', 'Experiencia', 'Psicotécnico', and 'Entrevista'. The 'Entrevista' tab is active. Below the navigation bar, there are two columns of input fields. The left column contains: 'Nombres *' with the value 'ERIKA', 'Dni *' with the value '48534851', 'Orientación a la Calidad *' (empty), 'Orientación a resultados *' (empty), and 'Precisión en manejo de información *' (empty). The right column contains: 'Apellidos *' with the value 'CAMPOS', 'Direccion *' with the value 'calle el sol 397', 'Tolerancia al trabajo bajo presión *' (empty), and 'Sentido de urgencia *' (empty). At the bottom right of the form, there are two buttons: 'Guardar' (green) and 'Cancelar' (orange).

Fuente: Elaboración Propia

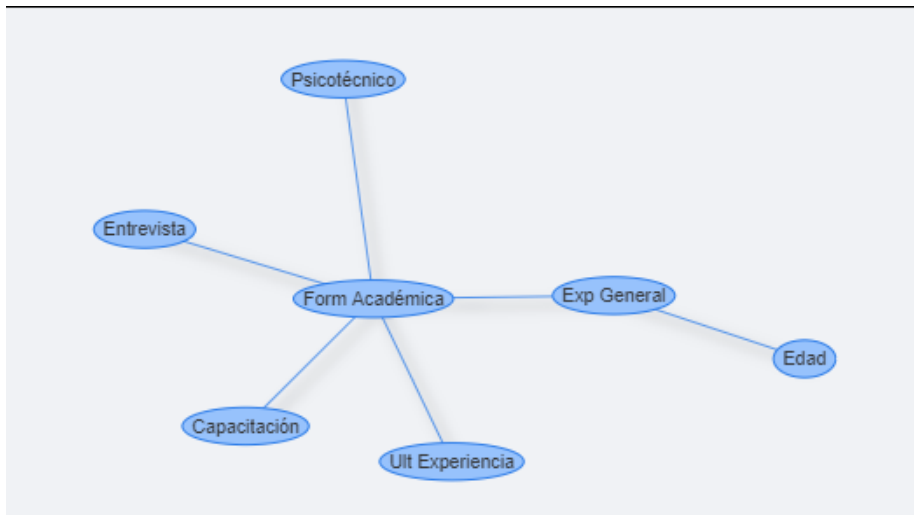
El sistema mostrara el siguiente mensaje:

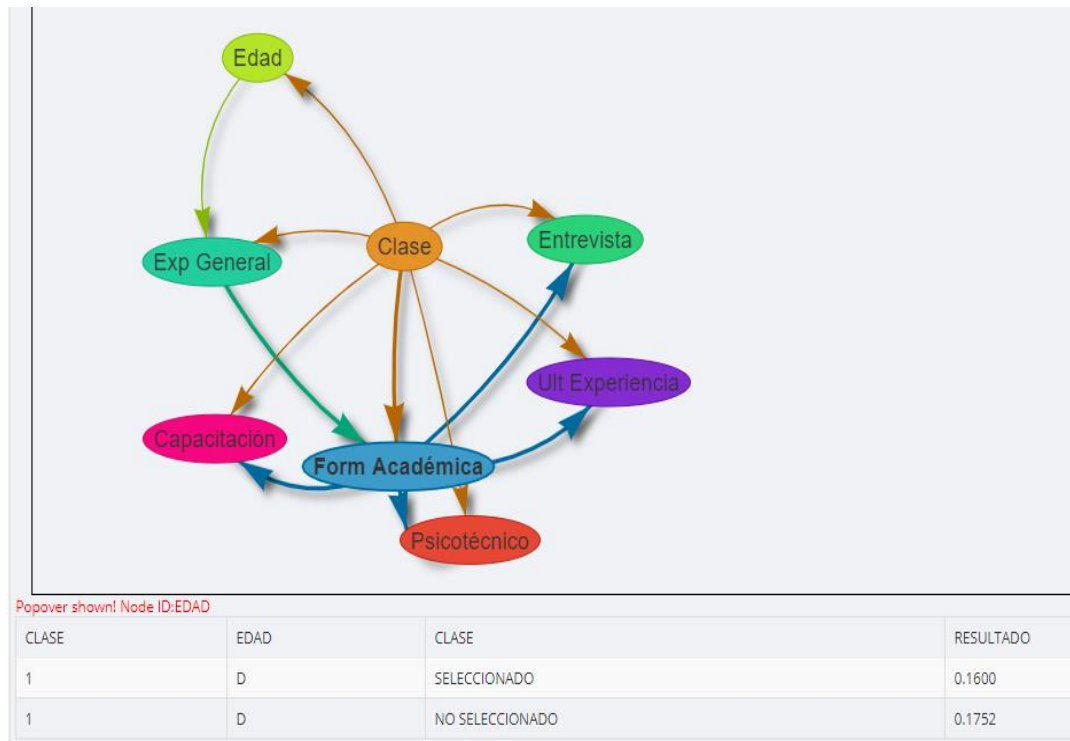


Para visualizar las redes generadas por el algoritmo implementado se deberá hacer click en la opción 



Ilustración 48: Visualización de las Redes Bayesianas utilizando el Algoritmo TAN





Fuente: Elaboración Propia

3.3.4.1.4. Pruebas y mejoras del Prototipo

Las pruebas del sistema basado en algoritmos computacionales se basaron en el enfoque de cajas negras, cuyo objetivo principal es el funcionamiento correcto, su buena rendición y su interfaz amigable al usuario.

Los resultados de las pruebas se presentan a continuación:

a. Pruebas de Requerimientos

Para poder realizar las pruebas de los requerimientos solicitados, se evaluaron estos y las funcionalidades que debe tener un sistema



Tabla 25: Pruebas de Requerimientos

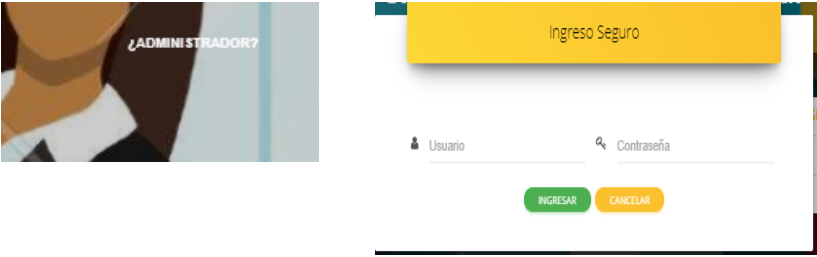
Requerimiento	¿Se cumplió el requerimiento?		Comentarios
	SI	NO	
Creación de Convocatorias	x		Correcto
Visualización de convocatorias	x		Correcto
Visualización del Perfil del Puesto	x		Correcto
Registro de datos personales	x		Correcto
Registro de Formación Académica y Cursos	x		Correcto
Registro de Experiencia Laboral	x		Correcto
Login del encargado del proceso de selección	x		Correcto
Registro de Resultado Evaluación Psicotecnia	x		Correcto
Registro de resultado de evaluación de entrevista	x		Correcto
Listado de Postulantes	x		Correcto
Generación de Red	x		Correcto
Visualización de Resultados	x		Correcto

Fuente. Elaboración Propia



b. Casos de Pruebas

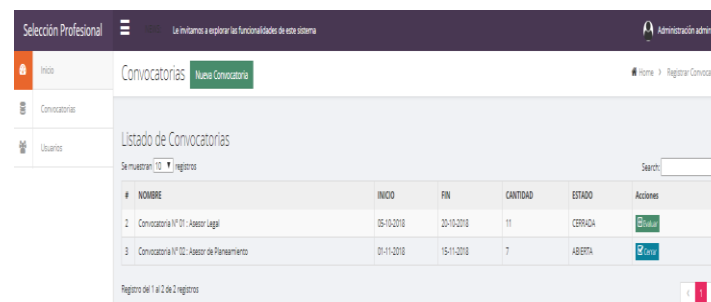
Tabla 26: *Login del Encargado de Selección Profesional*

LOGIN DEL ENCARGADO		
Objetivo de Caso de Prueba	Verificar que el administrador pueda ingresar por medio de su usuario y contraseña valida	
Identificador	TC_001_Login	
Nombre del Caso	Login del Encargado	
Pasos	Resultado Esperado	Resultado Real
1. Ir a la Opción Administrador y hacer clic	Se debe mostrar la Interfaz de Login llamado Ingreso Seguro	
2. Ingresar Usuario y Contraseña	OK	OK



3. Dar clic en el botón ingresar

Se debe mostrar la pantalla inicial de sistema



Fuente: Elaboración Propia

Tabla 27: Creación y Visualización de Convocatorias

CREACION Y VISUALIZACION DE CONVOCATORIAS

Objetivo de Caso de Prueba	Verificar la creación de convocatorias por parte del encargado de los procesos de selección
Identificador	TC0_002_Convocatorias
Nombre del Caso	Creación y visualización de convocatorias
Pre Condiciones	Ejecutar el TC_001_Login. Login del Encargado



Paso	Resultado Esperado	Resultado Real
1. Hacer clic en la Opción Nueva Convocatoria	Se debe mostrar la interfaz de Datos de Convocatoria	
2. Registrar Datos de la Convocatoria	OK	OK



3. Hacer Clic en botón Guardar
 Se debe mostrar la nueva convocatoria creada

Listado de Convocatorias

Se muestran 10 registros

Search:

#	NOMBRE	INICIO	FIN	CANTIDAD	ESTADO	Acciones
	Convocatoria N° 03 : Analista de Sistemas Informaticos	20-11-2018	07-12-2018	0	ABIERTA	<input checked="" type="checkbox"/> Cerrar

Fuente: Elaboración Propia

Tabla 28: Registro de Curriculum Vitae

REGISTRO DE CURRICULUM VITAE

Objetivo de Caso de Prueba	Verificar el registro y la validación de los datos personales y Curriculum Vitae de los postulantes al momento de ingresarlos al sistema	
Identificador	TC_003_RegistrodeCurriculumVitae	
Nombre del Caso	Registro de Curriculum Vitae	
Pre Condiciones	Ejecutar el TC_002_ Convocatorias. Creación y visualización de convocatorias	
Paso	Resultado Esperado	Resultado Real



1. Hacer Click en la opción Postular

Se debe mostrar la interfaz principal donde se registrara los datos personales y curriculum vitae del postulante

POSTULAR

POSTULAR

The screenshot displays the 'Sistema de Selección Profesional' registration interface. At the top, there is a 'Registro de Postulante' section with a 'Nombres' field and an 'Apellidos' field. Below this are fields for 'Dir.', 'Domicilio', 'Correo', 'Dirección', 'Teléfono', and 'Celular'. The interface is divided into two main sections: 'Formación Académica' and 'Experiencia Laboral'. The 'Formación Académica' section includes a dropdown for 'Grado Institución', a 'Capacitaciones' section with 'Capacitación' and 'Institución' fields, and an 'Experiencia Laboral' section with 'Empresa', 'Cargos', 'Inicio', and 'Fin' fields. There are 'Agregar Capacitación' and 'Agregar Experiencia' buttons. At the bottom, there are 'CREAR' and 'GUARDAR' buttons.



2. Digitar los datos personales del postulante

OK

The screenshot shows a form titled "Registro de Postulante" with a green header. It contains several input fields for personal data: "Nombres" and "Apellidos" (with a person icon), "Dni" (with a hash icon), "Fecha de Nacimiento" (with a calendar icon and format dd/mm/aaaa), "Correo" (with an envelope icon), "Dirección" (with a location pin icon), "Teléfono" (with a telephone icon), and "Celular" (with a mobile phone icon).

3. Digitar los datos correspondientes a la Formación Académica y Cursos y/o Capacitación

OK

The screenshot shows a form titled "Formación Académica" with a green header. It includes a dropdown menu for "Grado Instrucción", a section for "Capacitaciones" with fields for "Capacitación" and "Institución", and two date fields (format dd/mm/aaaa) for start and end dates, along with a "Horas" field. An orange button labeled "+ Agregar Capacitación" is positioned below these fields. At the bottom, there is a table header with columns: "#", "Nombre", "Institución", "Inicio", "Fin", "Horas", and "Acciones".

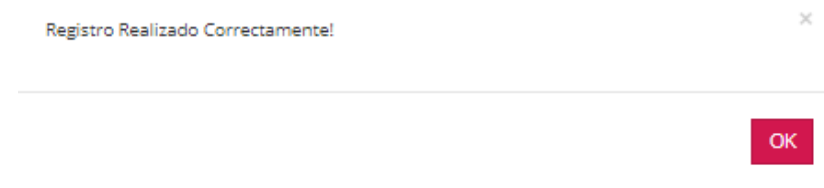


4. Digitar los datos correspondientes a la Experiencia Laboral

OK

5. Hacer Clic en botón Guardar

Se debe mostrar el siguiente mensaje “Registrado Correctamente”



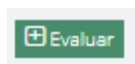
Fuente. Elaboración Propia



Tabla 29: Registro de Evaluaciones Presenciales

REGISTRO DE EVALUACIONES PRESENCIALES		
Objetivo de Caso de Prueba	Verificar el registro de las evaluaciones presenciales como el examen psicotécnico y entrevista personal	
Identificador	TC_004_RegistroEvaluacionesPresenciales	
Nombre del Caso	Registro de Evaluaciones Presenciales	
Pre Condiciones	Ejecutar el TC_003_RegistrodeCurriculumVitae Ejecutar el TC_001_Login	
Paso	Resultado Esperado	Resultado Real

1. Hacer clic en a Opción Evaluar




Se debe mostrar el listado de todos los postulantes para su posterior ingreso de las evaluaciones presenciales: psicotécnico y entrevista personal

Listado de Postulantes

Se muestran 10 registros Search:

DNI	APELLIDOS Y NOMBRES	NACIMIENTO	GRADO INSTRUCCIÓN	EMAIL	% SELEC.	% NO SELEC.	RESULTADO	
12345670	DAZ DAVID	1972-05-14	TITULO UNIVERSITARIO	david_29@hotmail.com	99.488	0.512	SELECCIONADO	
45142563	SAENZ RAUL	1990-02-10	TITULO DE MAGISTER	raul@hotmail.com	98.383	1.617	SELECCIONADO	
48534851	CAMPOS ERIKA	1993-07-25	TITULO UNIVERSITARIO	erika_23@hotmail.com				



2. Hacer clic en el icono  Se debe mostrar los datos de los postulantes e ingresar el resultado del examen Psicotécnico y Entrevista Personal

Curriculum Capacitaciones Experiencia **Psicotécnico** Entrevista

Nombres * ERIKA Apellidos * CAMPOS

Dni * 48534851 Direccion * calle el sol 397

Aptitud Académica * Preguntas Relacionados al Puesto *

Curriculum Capacitaciones Experiencia Psicotécnico **Entrevista**

Nombres * ERIKA Apellidos * CAMPOS

Dni * 48534851 Direccion * calle el sol 397

Orientación a la Calidad * Tolerancia al trabajo bajo presión *

Orientación a resultados * Sentido de urgencia *

Precisión en manejo de información *



3. Hacer clic en la opción



OK

Registro Realizado Correctamente!



Fuente. Elaboración Propia

Tabla 30: Generación de Red Bayesiana y Visualización de Resultados

GENERACIÓN DE RED BAYESIANA Y VISUALIZACION DE RESULTADOS		
Objetivo de Caso de Prueba	Verificar la generación de la red bayesiana de acuerdo a los pasos del algoritmo bayesiano y la probabilidad de cada postulante.	
Identificador	TC_005_GeneraciondeRedBayesianayResultados	
Nombre del Caso	Generación de Red Bayesiana y Visualización de Resultados	
Pre Condiciones	Ejecutar TC_004_RegistroEvaluacionesPresenciales	
Paso	Resultado Esperado	Resultado Real



1.- Registrar las evaluaciones presenciales en el sistema

Se debe mostrar los datos de los postulantes e ingresar el resultado del examen Psicotécnico y Entrevista Personal

Curriculum Capacitaciones Experiencia Psicotécnico **Entrevista**

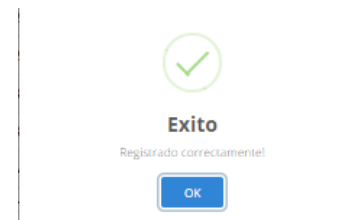
Nombres *	ERIKA	Apellidos *	CAMPOS
Dni *	48534851	Dirección *	calle el sol 397
Orientación a la Calidad *		Tolerancia al trabajo bajo presión *	
Orientación a resultados *		Sentido de urgencia *	
Precisión en manejo de información *			

Guardar Cancelar

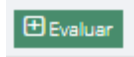
2. Hacer clic en la opción

Guardar

OK























3. Hacer clic en la Opción Evaluar
- Se debe mostrar la lista postulante con el porcentaje de la probabilidad generada por al algoritmo bayesiano



Listado de Postulantes

Se muestran 10 registros Search:

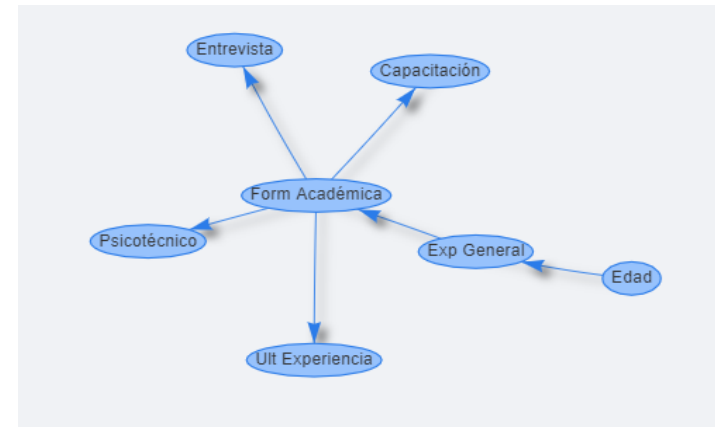
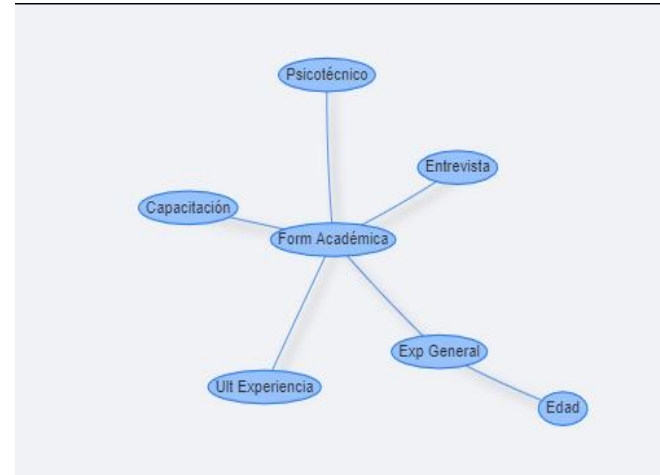
DNI	APELLIDOS Y NOMBRES	NACIMIENTO	GRADO INSTRUCCIÓN	EMAIL	% SELEC.	% NO SELEC.	RESULTADO	
08425217	MATEO YATACO FERNANDO JAVIER	1988-09-09	TITULO UNIVERSITARIO	FER_MATT76@HOTMAIL.COM	81.573	18.427	SELECCIONADO	 
08534845	RODRIGUEZ POLO VICTOR ANTONIO	1980-06-24	TITULO UNIVERSITARIO	victor_rodriguez@hotmail.com	15.261	84.739	NO SELECCIONADO	 
08844565	RAMOS MACAVILCA LUZ MARIA	1981-10-13	TITULO UNIVERSITARIO	LUZ_RAMOS13@HOTMAIL.COM	88.969	11.031	SELECCIONADO	 
08961950	VILCA ROJAS PEDRO ALBERTO	1985-01-14	TITULO UNIVERSITARIO	PEDRO_VILCAS@HOTMAIL.COM	81.423	18.577	SELECCIONADO	 
22233344	TERCERO SEGUNDO	1972-05-14	TITULO UNIVERSITARIO	segundo@hotmail.com	30.397	69.603	NO SELECCIONADO	 
25705701	CABALLERO SIFUENTES FABIOLA MARIA	1989-10-06	TITULO UNIVERSITARIO	FABI_CASI89@HOTMAIL.COM	91.894	8.106	SELECCIONADO	 
41575937	TERRAZAS PRADO DAVID ARTURO	1981-01-12	TITULO UNIVERSITARIO	TERRA_DAVID81@HOTMAIL.COM	34.574	65.426	NO SELECCIONADO	 
46587785	HUAMAN SANCHEZ JUAN MIGUEL	1984-06-12	TITULO UNIVERSITARIO	JUAN_H884@HOTMAIL.COM	16.471	83.529	NO SELECCIONADO	 
48568975	CAMPOS DIAZ ERIKA MILAGROS	1993-07-25	TITULO UNIVERSITARIO	erika_25_93@hotmail.com	86.905	13.095	SELECCIONADO	 
71245699	YOVERA CASTRO KAREN	1994-06-10	TITULO UNIVERSITARIO	karen@hotmail.com	1.030	98.970	NO SELECCIONADO	 

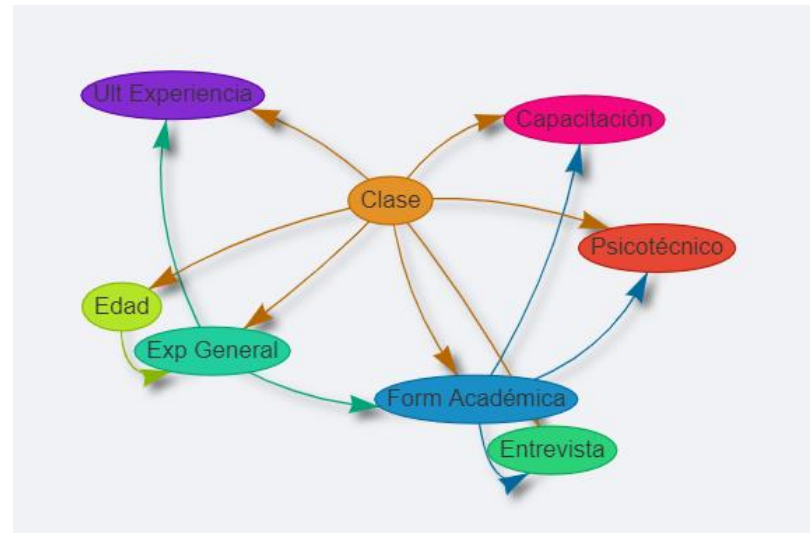


4. Hacer clic en la opción



Se debe mostrar las redes bayesianas generadas por el algoritmo bayesiano TAN y resultado de probabilidades obtenidas





Fuente: Elaboración Propia



IV

CONCLUSIONES Y RECOMENDACIONES



IV. CONCLUSIONES Y RECOMENDACIONES

4.1. Conclusiones

Después de realizar el análisis de requerimiento de los cuales 12 fueron funcionales y 2 no funcionales, se determinó que el sistema tendrá 14 funcionalidades.

Después de las características como la precisión, sensibilidad, especificidad y Margen de Error de diversos algoritmos computacionales se eligió al algoritmo bayesiano TAN (aumentado a árbol)

Al analizar los requerimientos se logró implementar el 100 % de los requerimientos funcionales y no funcionales, asimismo se logró integrar en el sistema el algoritmo bayesiano TAN.

Al realizar las pruebas se logró determinar que el indicador tiempo promedio empleado en el proceso de selección profesional de cada uno de los tres tipos de actividades consideradas en el pre-test fue de 124.17 minutos y en el post-test fue de 23.76 minutos, estableciendo 100.41 minutos de disminución. Asimismo se logró determinar que en el proceso 01 la precisión fue de 80% y la exactitud de 72 %; en el proceso 02, la precisión fue de 75% y de exactitud de 70% y en el proceso 03, la precisión fue de 83 y la exactitud de 69%

4.2. Recomendaciones

Incrementar los datos históricos utilizados como datos previos a la aplicación del algoritmo bayesiano. A mayor cantidad de datos históricos mayor exactitud.

Las redes bayesianas podrían usarse para generar más y mejores modelos del funcionamiento en diversas áreas de las organizaciones, los grupos cualquier agregado social.

V. REFERENCIAS

- Alarcon Jaimes, C. (2015). *Optimización del Clasificador Naive Bayes usando árbol de decisión*. Lima-Peru.
- AlfonsoGU. (2012 de 08 de 2012). *VELNEO*. Obtenido de <https://velneo.es/que-es-javascript/>
- Artica Chacon, M. Á., & Ocaña Sudario, N. A. (2015). *Aspecto Nutricional apoyado por un Modelo Bayesiano y su relacion en el pronostico de Morbilidad del Neonato de la gestante*. Lima- Perú.
- Atalaya Pisco, M. (2001). Nuevos Enfoques en Selecccion de Personal. *Revista de Investigacion en Psicologia*.
- Beltran Pascual, M., Muñoz Martinez, A., & Muñoz Alamillos, A. (2014). Redes Bayesianas aplicadas a problemas de credit scoring. Una aplicacion practica.
- Blanco, L. M. (2002). *Programacion en Visual Basic.Net*. Madrid España: Grupo EIDOS.
- Bonilla Gordillo, A. F., & Ojeda Schuldt, M. A. (2006). *Implementacion de Minería de Datos Basada en Redes Bayeisanas para la Toma de Decisiones en los Registros Academicos*. Guayaquil - Ecuador.
- Botelbo, T., Soprani, D., Rodrigues, C., & Frizera, A. (2017). Nuevo Enfoque para la Clasificación de Señales de EEG usando la Varianza de la diferencia entre las Clases de un Clasificador Bayesiano. *Revista Iberoamericana de Automática e Informática Industrial*, 362-371.
- Calavancha Zuñiga, N. A. (s.f.). Breve aproximacion a la tecnica de arbol de decisiones.
- Canos, L. (2008). Modelo flexibles de seleccion de personal basados en la valoracion . *Revistaa Electronica de Comunicaciones y Trabajos de ASEPUMA*.
- Coggeshall, J. (2005). *La biblia de PHP 5*. Madrid: Anaya Multimedia.
- Cordova Roque, E. G. (2017). *Analisis predictivos de muerte y sobrevida de pacientes hospitalarios mediante clasificadores supervisados*. Universidad Catolica de Santa Maria, Arequipa - Peru.
- Cortes de la Fuente, J. (2011). *La información mutua como medida de asociación y su utilidad en analisis geneticos*. Universidad Politecnica de Cataluña, Barcelona.
- CRUZ-GUERRERO, R., ALONSO-LA VERNIA, M., FRANCO-ARCEGA, A., & SIMON-MARMOLEJO, I. (27). Estudio del comportamiento de algoritmos de clasificacon segun la naturaleza de los datos. *Revista de Tecnologia Informatica*.
- Dedeu Ribera, M. (s.f.). *Etapas del Proceso de Selecccion de Personal en las Empresas Familiares*. Obtenido de <http://dissenyweb.torrentidedeu.com/es/blog/etapas-del-proceso-de-seleccio%CC%81n-de-personal-en-las-empresas-familiares.html>
- Flanagan, D. (2002). *JavaScript: The Definitive Guide (4ª Edición)*. Madrid, España: O'Reilly logo.



- Ganzabal Garcia, X. (2014). *Desarrollo y reutilizacion de componentes de Software y multimedia mediante lenguajes de guion*.
- Garrido Cobo, J., Lorente Puchades, I., & Almirall, L. J. (2013). *TFC Desarrollo de Aplicaciones Moviles*.
- Gómez Jiménez, E. (2012). *Desarrollo de Software con Netbeans 7.1*. Mexico: Alfaomega Grupo Editor S.A de C.V. Mexico.
- González, D. L., & Rodríguez, R. J. (2006). *Prototipo de Software para la clasificacion de datos mediante el Metodo Bayesiano TAN – “udTAN”*. Universidad Distrital Francisco José de Caldas, Bogota.
- Guzman Ortiz, R. L. (2018). *Sistema Informatico de control de ventas para la empresa inversiones Cuba SRL de la ciudad de Chimbote*. Universidad San Pedro. Chimbote-Peru: u.
- Hassoun, M. H. (1995). *Fundamentals of Artificial Neural Networks*.
- Hernandez Orallo, J., Ramirez Quintana, M. J., & Ferri Ramirez, C. (2004). *Introduccion a la Minería de Datos*. Valencia- España: Pearson.
- INTECO. (2009). *Curso de Introduccion a la Ingenieria del Software*. España: Instituto Nacional de Tecnologías de la Comunicacion.
- Kruchten, P. (2004). *An Ontology of architectural design decisions in software intensive systems*.
- Lázaro Gonzales, A. Z. (2015). *Propuesta de Sistema Experto para detección temprana de enfermedades neoplásicas en paciente de la Clínica San Bartolomé - Huacho, 2014*. Huacho - Perú.
- Lujan Mora, S. (2002). *Programación de aplicaciones web: historia, principios*. Madrid, España.
- Malca Bulnes, S. M. (2015). *Modelo Algoritmico para la clasificacion de una hoja de planta en base a sus características de forma y textura*. Lima - Perú.
- Melendez Valladares, S. M., Gaitan, M. E., & Perez Reyes, N. N. (2015). *Sistema web de evaluacion al desempeño docente UNAN-Managua, empleando la metodologia agil programacion extrema, en el II sementre del 2015*. Managua.
- Ministerio de Transportes y Comunicaciones. (mayo de 2017). *Plan Operativo Institucional*. Obtenido de http://transparencia.mtc.gob.pe/idm_docs/P_recientes/10365.pdf
- Oviedo Balladares, L. (2017). *Clasificador bayesiano basado en dsitrubucion normal y su aplicacion en la prediccion de tipo de ictericia neonatal*. Universidad Nacional de San Antoncio de Abad del Cusco, Cusco - Peru.
- Pedro Gabriel, R., Rodriguez, L. V., & Cruz Barboza, R. (2015). *Redes Bayesianas para la clasificacion de masas en radiografias*. Universidas Tecnologiaca de Mixteca.
- Quepuy Perleche, C. E. (2016). *Sistema de Gestión y Soporte de Toma de Decisiones basado en Algoritmos de Bayes y Cluster para mejorar procesos analíticos del área comercial de una empresa educativa*. Universidad Católica Santo Toribio de Mogrovejo, Chiclayo - Perú.



- Ramon Campoverde, K. (2016). *Desarrollo de un sistema web utilizando redes bayesianas para enseñanzas del idioma ingles en la escuela Zoila Alvarado de Jaramillo*. Universidad Nacional de Loja, Loja- Ecuador.
- Rodriguez Bucarely, C. M., Rodriguez Bucarely, P. A., Piñeyro Torres, C., Morel Pichardo, C. A., & Hilario Sanchez, E. (2008). *Visual Basic 6.0 Orientado a Base de Datos. Segunda Edicion*. Grupo Experto Bucarely.
- Sandoval, L. J. (2018). *Algoritmos de aprendizaje automatico para analisis y prediccion de datos*. Santa Tecla: ITCA-FEPADE.
- Santiesteban Rojas, J. C., Utria Perez, D., & Hernandez Reyes, C. E. (21 de 06 de 2012). *revista vinculando*. Obtenido de <http://vinculando.org/articulos/redes-bayesianas.html>
- Serrano, J., Antonio, Soria, Emilio, & D. Martin, J. (2010). *Redes Neuronales Artificiales*.
- Sommerville, I. (2006). *ingeniería del Software* (7ma. ed.). Madrid.
- Tume Ayala, V. R. (2014). *Desarrollo de Inteligencia de Negocios para analizar empresas financieras: Cajas Municipales*. Piura - Perú.
- Ureña Almagro, C. (2010). *Lenguajes de Programacion*. Universidad de Granada.
- Xiloj, G. (8 de Julio de 2015). *Visual Basic*. Obtenido de Visual Basic: <http://guillermogasparvisualbasic.blogspot.com/>



VI

ANEXOS

VI. ANEXOS

ANEXO A.

ENCUESTA DIRIGIDA A LOS EMPLEADOS DE LA OFICINA DE PERSONAL ENCARGADOS DE LA SELECCIÓN PROFESIONAL



ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS

OBJETIVO: Esta encuesta recolecta información que servirá para la implementación de un sistema basado en algoritmos computaciones como ayuda a los proceso de selección Profesional en el MTC.

INDICACIÓN:

1. Responda, cada una de las preguntas que se le presentan a continuación y marque con una “X” la repuesta que usted cree que sea la indicada a la pregunta establecida.
2. Si sus repuestas son negativas, indique ¿Por qué?, con el objetivo de ampliar la evaluación de investigación.

I- PREGUNTAS GENERALES

Nombre completo: _____

Lugar y fecha: _____

Sexo: Masculino. Femenino.

Cargo que desempeña: _____

Área: _____

Grado académico: _____

Tiempo de laborar

Desde _____ Hasta _____



II.- PREGUNTAS ESPECÍFICAS.

2.- ¿En qué beneficia la selección profesional?

3.- ¿Se tiene algún manual de reclutamiento y selección profesional?

SI NO

Porque_____

4.- ¿Cuáles son las bases de la selección profesional?

5.- ¿Cuál es la técnica de selección profesional que a su criterio permite escoger al candidato ideal?

6.- ¿Cómo identifica y selecciona al personal con los conocimientos adecuados para cubrir las necesidades de los puestos de trabajo?

7.- ¿Cuáles son los pasos más importantes que tiene el proceso de selección profesional?

8.- ¿Qué elementos se evalúa en la entrevista de selección profesional?

9.- ¿Cuántas entrevistas realiza durante el proceso de selección?



10.- ¿Es indispensable la entrevista en la selección profesional?

SI NO

PORQUE.

11.- ¿De qué manera participa la tecnología en sus procesos de selección profesional?

ANEXO B.

ENCUESTA DIRIGIDA SOBRE LA PRUEBA DEL SISTEMA AL ENCARGADO DE LA OFICINA DE PERSONAL SOBRE LAS FUNCIONALIDADES IMPLEMENTADOS DE ACUERDO A SU REQUERIMIENTO



ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS

OBJETIVO: Esta encuesta recolecta información que servirá para la implementación de un sistema basado en algoritmos computaciones como ayuda a los proceso de selección Profesional en el MTC.

INDICACIÓN:

3. Responda, cada una de las preguntas que se le presentan a continuación y marque con una “X” la respuesta que usted cree que sea la indicada a la pregunta establecida.

II- PREGUNTAS GENERALES

Nombre completo: _____

Lugar y fecha: _____

Sexo: Masculino. Femenino.

Cargo que desempeña: _____



Área: _____

Grado académico: _____

Tiempo de laborar:

Desde _____ Hasta _____

II.- PREGUNTAS

Nº	FUNCIONALIDADES	¿Funcionó correctamente?		Comentario
		SI	NO	
1	¿El encargado puede crear Convocatorias?			
2	¿El postulante puede visualizar las convocatorias?			
3	¿El postulante puede visualizar el Perfil del Puesto?			
4	¿Se puede registrar lo datos personales del postulante?			
5	¿Los postulantes pueden registrar su Formación Académica y Cursos?			
6	¿Los postulantes pueden registrar su Experiencia Laboral?			
7	¿El encargado del proceso de selección puede ingresar al sistema a través del LOGIN?			
8	¿El encargado del proceso de selección puede registrar el resultado de la Evaluación Psicotécnica?			



9	¿El encargado del proceso de selección puede registrar el resultado de la entrevista?
10	¿Se puede visualizar el listado de Postulantes por cada convocatoria?
11	¿Se visualiza la Red Bayesiana?
12	¿Se visualiza de Resultados de cada postulante?
13	¿El sistema tiene una interfaz amigable?
14	¿El sistema contiene alerta o mensaje de error?

ANEXO C.

Tabla 31: *Tabla de Resultado del Proceso de Selección 01: Asesor*

N°	Código	% Selec.	% No Selec.	Predicción	Real	Categoría
1	08425217	81.573	18.427	Seleccionado	Seleccionado	VP
2	08534845	15.261	84.739	No Seleccionado	No Seleccionado	VN
3	08844565	88.969	11.031	Seleccionado	Seleccionado	VP
4	08961950	81.423	18.577	Seleccionado	No seleccionado	FP
5	22233344	30.397	69.603	No Seleccionado	No Seleccionado	VN
6	25705701	91.894	8.106	Seleccionado	Seleccionado	VP
7	41575937	34.574	65.426	No Seleccionado	Seleccionado	FN
8	46587785	16.471	83.529	No Seleccionado	No seleccionado	FN



9	48568975	86.905	13.095	Seleccionado	Seleccionado	VP
10	71245699	1.030	98.970	No Seleccionado	No seleccionado	VN
11	72398687	17.613	82.387	No Seleccionado	No seleccionado	VN

Fuente: Elaboración Propia

ANEXO D:

Tabla 32 : Tabla de Resultado del Proceso de Selección 02. Asesor de Planeamiento

N°	Código	% Selec.	% No Selec.	Predicción	Real	Categoría
1	06013061	41.083	58.917	No Seleccionado	Seleccionado	FN
2	06013062	7.325	92.675	No Seleccionado	No Seleccionado	VN
3	06052182	40.711	59.289	No Seleccionado	Seleccionado	FN
4	07192012	90.449	9.551	Seleccionado	Seleccionado	VP
5	07626682	91.621	8.379	Seleccionado	Seleccionado	VP
6	42927101	74.980	25.020	Seleccionado	No Seleccionado	FP
7	44370535	1.145	98.855	No Seleccionado	No Seleccionado	VN
8	45524977	76.851	23.149	Seleccionado	Seleccionado	VP
9	46587748	1.446	99.554	No Seleccionado	No Seleccionado	VN
10	73239710	0.999	99.001	No Seleccionado	No Seleccionado	VN

Fuente: Elaboración Propia



ANEXO E:
Tabla 33 : Tabla de Resultado del Proceso de Selección 03: Analista de Sistemas

N°	Código	% Selec.	% No Selec.	Predicción	Real	Categoría
1	02880108	54.173	45.827	Seleccionado	No Seleccionado	VP
2	09385062	4.915	95.085	No Seleccionado	No Seleccionado	VN
3	29306110	27.145	72.855	No Seleccionado	Seleccionado	FN
4	29109269	97.824	2.176	Seleccionado	Seleccionado	VP
5	40777432	0.678	99.322	No Seleccionado	No Seleccionado	VN
6	41368585	94.494	5.506	Seleccionado	Seleccionado	VP
7	41517408	30.223	69.777	No Seleccionado	Seleccionado	FN
8	42854538	70.660	29.340	Seleccionado	No Seleccionado	FP
9	46040428	22.568	77.432	No Seleccionado	Seleccionado	FN
10	46410089	86.524	13.476	Seleccionado	Seleccionado	VP
11	46417227	15.755	84.245	No Seleccionado	No Seleccionado	VN
12	46838838	2.379	97.721	No Seleccionado	No Seleccionado	VN
13	72398687	84.002	15.998	Seleccionado	Seleccionado	VP

Fuente: Elaboración Propia


ANEXO F:**CODIGO FUENTE DE LA CONEXIÓN A LA BASE DE DATOS**

```
class clsAccesoDatos{

//Codigo de Tabla
public $gIdTabla;
//Codigo de Cliente
public $gIdCliente;
// Total de Paginas
public $gNroPaginas;
// Total de Registros
public $gNroRegistros;
// Pagina que se muestra
public $gPagActual;
// Conexion BD
private $gCnx;
// Para ejecutar procedures
public $gStmt;
public $gError;
public $gMsg;
// Nro de Registro a Mostrar
private $gNumReg = 20;
//Servidor de Base de Datos
private $gServidor = "localhost";
//Nombre de Base de Datos
// private $gBaseDatos = "fv";
private $gBaseDatos = "bayes2";
//Nombre de Usuario
//Tipo de Base Datos
private $gTipoBD = 2; //1=SQLSERVER, 2=MYSQL, 3=POSTGRESQL
public $gTipoConex = 1; //1=PDO, 2 = PDOSICA
private $gFB;
// Constructor de la clase
function __construct(){
    global $ggTipoBD;
    $this->gTipoBD=$ggTipoBD;
}
```



```
if($this->gTipoBD==1){
    $this->gServidor = "localhost";
}
if($this->gTipoBD==2){
    $this->gServidor = "localhost";
    $user='root';
    $pass="";
}
if($this->gTipoBD==3){
    $this->gServidor = "localhost";
    //$user='mampuerto';//sica
    //$pass='tBZPD*w#$eAE';//sicaweb2013
    $user='postgres';
    $pass='postgres';
}
// Crea una Conexion SQLSERVER 2000
//echo "Antes de COncetar";
try {
    //print_r(PDO::getAvailableDrivers());
    if(isset($cnx) && $cnx){
        $this->gCnx = $cnx;
    }else{
        if($this->gTipoBD==1){
            $this->gCnx = new PDO("mssql:host=".$this->gServidor.";dbname=".$this-
            >gBaseDatos,$user,$pass);
        }
        if($this->gTipoBD==2){
            $this->gCnx = new PDO("mysql:host=".$this->gServidor.";port=3306;dbname=".$this-
            >gBaseDatos,$user,$pass);
            //$this->gCnx->setAttribute (PDO::MYSQL_ATTR_USE_BUFFERED_QUERY, true);
        }
        if($this->gTipoBD==3){
            if($this->gTipoConex==1){
                $this->gCnx = new PDO("pgsql:host=".$this->gServidor.";port=5432;dbname=".$this-
                >gBaseDatos,$user,$pass);
            }else{
                echo "Antes de Post";
            }
        }
    }
}
```



```

    $this->gCnx = new PDOSICA("pgsql:host=". $this->gServidor.";port=5432;dbname=". $this-
    >gBaseDatos,$user,$pass);
    echo $this->gCnx;
    echo "Inicia:.....". $this->gCnx->errorInfo();
    }
    }
    }
    // $this->gCnx->setAttribute( PDO::ATTR_ERRMODE, PDO::ERRMODE_WARNING );
    // $PDO->setAttribute(PDO_ATTR_PERSISTENT, true);
} catch (PDOException $e) {
    echo "Error:\n" . $e->getMessage();
}
//mysql_query("SET NAMES 'utf8'");

```

ANEXO G :

CODIGO DE CALCULO DE PROBALIDAD FINAL

```

$resultado=$objEvaluacion->insertarResultadoFinal($idperconvocatoria, $archivo_xml,
$cantidad_xml);
$resultadofinal=$objEvaluacion->listarResultadoFinal($idperconvocatoria);
$seleccionado_inicial=1;
$noseleccionado_inicial=1;
$seleccionado_final=0;
$noseleccionado_final=0;
foreach ($resultadofinal as $dato) {
    if($dato["tipo"]=="SELECCIONADO"){
        $seleccionado_inicial = $seleccionado_inicial*$dato["resultado"];
    }
    if($dato["tipo"]=="NO SELECCIONADO"){
        $noseleccionado_inicial = $noseleccionado_inicial*$dato["resultado"];
    }
}
$seleccionado_final = (
$seleccionado_inicial/($seleccionado_inicial+$noseleccionado_inicial) ) *100;
$noseleccionado_final = (
$noseleccionado_inicial/($seleccionado_inicial+$noseleccionado_inicial) ) *100;

```

```
$decision = "";
if($seleccionado_final>=$noseleccionado_final){
    $decision="SELECCIONADO";
}else{
    $decision="NO SELECCIONADO";
}
$actualizarresultado=$objEvaluacion->actualizarResultado($idperconvocatoria,
$seleccionado_final, $noseleccionado_final, $decision);

    $response['inserted'] = true;
}
$objEvaluacion->finalizarTransaccion();

$response['RPTA'] = $resultado;
//$response['XML'] = $archivo_xml;
//$response['ordenado'] = $c_ordenado;

header('Content-Type: application/json');
echo json_encode($response);
}
```