



---

**UNIVERSIDAD CATÓLICA LOS ÁNGELES  
CHIMBOTE**

**FACULTAD DE INGENIERÍA  
ESCUELA PROFESIONAL DE INGENIERÍA DE  
SISTEMAS**

**PROPUESTA DE IMPLEMENTACIÓN DE MINERÍA DE  
DATOS PARA EL COLEGIO NACIONAL DE LA  
LIBERTAD – HUARAZ; 2020.**

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

**AUTOR**

**RONDO ROPA, ELÍAS MANUEL**

**ORCID: 0000-0002-2064-0393**

**ASESORA**

**SUXE RAMÍREZ, MARÍA ALICIA**

**ORCID: 0000-0002-1358-4290**

**CHIMBOTE – PERÚ**

**2020**

## **EQUIPO DE TRABAJO**

### **AUTOR**

Rondo Ropa, Elías Manuel

ORCID: 0000-0002-2064-0393

Universidad Católica Los Ángeles de Chimbote, Estudiante de Pregrado,  
Chimbote, Perú

### **ASESORA**

Suxe Ramírez, María Alicia

ORCID: 0000-0002-1358-4290

Universidad Católica Los Ángeles de Chimbote, Facultad de Ingeniería,  
Escuela Profesional de Ingeniería de Sistemas, Chimbote, Perú

### **JURADO**

Castro Curay, José Alberto

ORCID: 0000-0003-0794-2968

Ocaña Velásquez, Jesús Daniel

ORCID: 0000-0002-1671-429X

Torres Ceclén, Carmen Cecilia

ORCID: 0000-0002-8616-7965

**JURADO EVALUADOR DE TESIS Y ASESOR**

MGTR. ING. CIP. JOSÉ ALBERTO CASTRO CURAY

PRESIDENTE

DR. ING. CIP. JESÚS DANIEL OCAÑA VELÁSQUEZ

MIEMBRO

MGTR. ING. CIP. CARMEN CECILIA TORRES CECLÉN

MIEMBRO

DRA. ING. CIP. MARÍA ALICIA SUXE RAMÍREZ

ASESORA

## **DEDICATORIA**

A mi madre que siempre me dio su apoyo para alcanzar este objetivo, ella es el modelo que he decido imitar, a mi esposa que siempre estuvo ahí para apoyarme y ayudarme en los momentos más difíciles y estresantes, a mis hijos que son el motivo que me impulsa a seguir adelante y que espero ser un ejemplo para ellos.

***Elías Manuel Rondo Ropa***

## **AGRADECIMIENTO**

A Dios por darme vida y salud, a mi familia por apoyarme en todo momento, mi docente y asesor del curso por el asesoramiento constante y a todas las personas que estuvieron involucradas en la presente tesis como a los profesionales que aportaron con sus conocimientos para lograr mis objetivos y metas profesionales.

***Elías Manuel Rondo Ropa***

## RESUMEN

La presente tesis fue desarrollada bajo la línea de investigación: Domótica y automatización de la Escuela Profesional de Ingeniería de Sistemas de la Universidad Católica los Ángeles de Chimbote; tuvo como objetivo: Proponer la implementación de Minería de Datos para el Colegio Nacional de la Libertad en la ciudad de Huaraz, con la finalidad de determinar patrones de comportamiento de la deserción escolar, la metodología de investigación fue del tipo descriptiva de nivel cuantitativa con diseño no experimental de corte transversal, el universo se delimitó en 210 y la muestra se seleccionó a 43 estudiantes del 5to año de educación secundaria, para la recolección de datos se utilizó el instrumento denominado cuestionario mediante la técnica de la encuesta, los cuales arrojaron los siguientes resultados, para la primera dimensión, sobre el nivel de satisfacción del proceso actual de detección de posibles casos de riesgo, se observó que el 81.40% Sí acepta el proceso actual, para la segunda dimensión, sobre la necesidad de implementación de Minería de Datos para la detección de posibles casos de riesgo, se observó que el 100.00%, Sí está de acuerdo con la implementación, dichos resultados son concordantes con las hipótesis específicas, quedando demostrada y justificada la hipótesis general, se concluye que si existe la necesidad de contar con la propuesta de implementación de Minería de Datos que permita identificar los patrones de comportamiento de deserción escolar, teniendo que el alcance de la investigación es local donde se contempla a todos los estudiantes de la Institución Educativa.

**Palabras clave:** Aprendizaje Automático, Deserción escolar, Inteligencia Artificial, Minería de datos, Patrones de comportamiento.

## ABSTRACT

This thesis was developed under the line of research: Domotics and automation of the Professional School of Systems Engineering of the Universidad Católica Los Ángeles de Chimbote; had as objective: To propose the implementation of data mining for the National School of Freedom in the city of Huaraz, in order to determine behavior patterns of school dropout, the research methodology was of the descriptive type of quantitative level with design Non-experimental cross-sectional, the universe was delimited in 210 and the sample was selected to 43 students of the 5th year of secondary education, for the data collection the instrument called questionnaire was used by means of the survey technique, which yielded the following Results, for the first dimension, on the level of satisfaction of the current process of detection of possible risk cases, it was observed that 81.40% Yes accept the current process, for the second dimension, on the need to implement data mining to the detection of possible cases of risk, it was observed that 100.00%, Yes they agree with the implementation, said res The results are consistent with the specific hypotheses, the general hypothesis being demonstrated and justified, it is concluded that if there is a need to have the data mining implementation proposal that allows identifying the behavior patterns of school dropout, having the scope of The research is local where all the students of the Educational Institution are considered.

**Keywords:** Machine Learning, Dropout, Artificial Intelligence, Data Mining, Behavior Patterns.

## ÍNDICE DE CONTENIDO

EQUIPO DE TRABAJO .....	ii
JURADO EVALUADOR DE TESIS Y ASESOR.....	iii
DEDICATORIA .....	iv
AGRADECIMIENTO .....	v
RESUMEN .....	vi
ABSTRACT.....	vii
ÍNDICE DE CONTENIDO .....	viii
ÍNDICE DE TABLAS .....	x
ÍNDICE DE GRÁFICOS .....	xi
I. INTRODUCCIÓN.....	1
II. REVISIÓN DE LA LITERATURA.....	6
2.1. Antecedentes .....	6
2.1.1. Antecedentes a nivel internacional .....	6
2.1.2. Antecedentes a nivel nacional.....	8
2.1.3. Antecedentes a nivel regional .....	11
2.2. Bases teóricas .....	13
2.2.1. Rubro de la empresa .....	13
2.2.2. La empresa Investigada .....	14
2.2.3. Las Tecnologías de información y comunicaciones (TIC).....	22
2.2.4. Teoría relacionada con la Tecnología de la investigación .....	24
2.2.4.1. Definiciones .....	24
2.2.4.2. Casos de éxito .....	36
2.2.4.3. Metodologías de Minería de Datos. ....	37
III. HIPÓTESIS .....	41
3.1. Hipótesis General .....	41
3.2. Hipótesis específicas .....	41
IV. METODOLOGÍA.....	42
4.1. Tipo de la investigación .....	42
4.2. Nivel de la investigación de la tesis .....	42
4.3. Diseño de la investigación .....	43



4.4.	Universo y muestra .....	44
4.5.	Definición operacional de las variables en estudio .....	45
4.6.	Técnicas e instrumentos de recolección de datos.....	47
4.6.1.	Técnica.....	47
4.6.2.	Instrumentos.....	47
4.7.	Plan de análisis .....	47
4.8.	Matriz de consistencia.....	48
4.9.	Principios éticos .....	50
V.	RESULTADOS .....	52
5.1.	Resultados de la encuesta a los alumnos.....	52
5.1.1.	Resultados de la dimensión 1: Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar.....	52
5.1.2.	Resultados de la dimensión 2: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento. ....	62
5.1.3.	Resultados por dimensión.....	72
5.1.3.1.	Resultado general de la dimensión 1.....	72
5.1.3.2.	Resultado general de la dimensión 2.....	74
5.1.4.	Resumen general.....	76
5.2.	Análisis de resultados.....	79
5.3.	Propuesta de mejora .....	81
5.3.1.	Propuesta técnica .....	82
5.3.2.	Diagrama de Gantt .....	96
5.3.3.	Propuesta económica .....	97
VI.	CONCLUSIONES .....	98
VII.	RECOMENDACIONES.....	100
VIII.	REFERENCIAS BIBLIOGRÁFICAS .....	102
ANEXOS .....		107
ANEXO NRO. 1:	CRONOGRAMA DE ACTIVIDADES .....	108
ANEXO NRO. 2:	PRESUPUESTO.....	109
ANEXO NRO. 3:	CUESTIONARIO DE ALUMNOS.....	110
ANEXO NRO. 4:	CONSENTIMIENTO INFORMADO.....	113

## ÍNDICE DE TABLAS

Tabla Nro. 1: Matriz de Operacionalización de la Variable .....	45
Tabla Nro. 2: Matriz de Consistencia .....	48
Tabla Nro. 3: Técnicas de detección de casos de riesgo.....	52
Tabla Nro. 4: Importancia del estudio escolar.....	53
Tabla Nro. 5: Familiares que desertaron en sus estudios.....	54
Tabla Nro. 6: Círculo de amistades peligrosas .....	55
Tabla Nro. 7: Referencias sobre la situación económica.....	56
Tabla Nro. 8: Identificación de la composición familiar .....	57
Tabla Nro. 9: Incremento de la delincuencia .....	58
Tabla Nro. 10: Prevención de embarazos a temprana edad .....	59
Tabla Nro. 11: Detección temprana de acoso sexual.....	60
Tabla Nro. 12: Detección temprana de prevención del bullying .....	61
Tabla Nro. 13: Necesidad de implementación de las TIC .....	62
Tabla Nro. 14: Tecnologías basadas en inteligencia artificial.....	63
Tabla Nro. 15: Modelo de Sistemas Expertos .....	64
Tabla Nro. 16: Creación de bases de datos .....	65
Tabla Nro. 17: Automatización de la detección de casos de riesgo .....	66
Tabla Nro. 18: Información oportuna para toma de decisiones.....	67
Tabla Nro. 19: Anticipación ante patrones de comportamiento .....	68
Tabla Nro. 20: Sistemas Expertos.....	69
Tabla Nro. 21: Confiabilidad y calidad de la información.....	70
Tabla Nro. 22: Anticipación ante casos de riesgo .....	71
Tabla Nro. 23: Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar.....	72
Tabla Nro. 24: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento.....	74
Tabla Nro. 25: Resumen general de dimensiones.....	76
Tabla Nro. 26: Variables resultantes de predicción.....	90
Tabla Nro. 27: Historial de cálculos .....	94
Tabla Nro. 28: Presupuesto de la Implementación .....	97

## ÍNDICE DE GRÁFICOS

Gráfico Nro. 1: Ubicación Geográfica de Colegio Nacional de la Libertad.....	15
Gráfico Nro. 2: Organigrama del Colegio Nacional de la Libertad – Huaraz .....	20
Gráfico Nro. 3: Ejemplo de un árbol de decisiones.....	32
Gráfico Nro. 4: Ejemplo de Métodos Bayesianos. ....	32
Gráfico Nro. 5: Ejemplo de Redes Neuronales.....	33
Gráfico Nro. 6: Resultado general de la dimensión 1 .....	73
Gráfico Nro. 7: Resultado general de la dimensión 2.....	75
Gráfico Nro. 8: Resumen general de las dimensiones .....	77
Gráfico Nro. 9: Resumen porcentual de las dimensiones .....	78
Gráfico Nro. 10: Proceso de la Metodología KDD. ....	83
Gráfico Nro. 11: Matriz de datos generales .....	84
Gráfico Nro. 12: Matriz de datos Académicos .....	85
Gráfico Nro. 13: Matriz de integración. ....	86
Gráfico Nro. 14: Datos atípicos del atributo edad .....	88
Gráfico Nro. 15: Esquema para la generación del modelo predictivo de deserción escolar. ....	92
Gráfico Nro. 16: Árbol de decisión (con Rattle de R).....	93
Gráfico Nro. 17: Evaluación del modelo .....	93
Gráfico Nro. 19: Diagrama de Gantt Implementación .....	96

## I. INTRODUCCIÓN

En los últimos años en el mundo actual se ha incrementado la deserción escolar en los alumnos principalmente en el nivel secundario por diferentes factores influyentes, hoy en día existen diferentes técnicas y métodos de percepción para determinar los patrones de comportamiento que intervienen en el aprovechamiento académico, el cual permite identificar anticipadamente los potenciales desertores escolares, actualmente es posible agrupar los factores a través de dos grandes marcos interpretativos carácter extraescolares e intraescolares (1) (2).

Actualmente en Perú respecto a la educación se analizan posibilidades y criterios para detectar los patrones o factores que perjudican el rendimiento académico de los estudiantes en los diversos niveles, aunque aún no existen modelos de predicción programados que permitan analizar estos casos de manera inmediata por medio de sistemas expertos. Por parte del gobierno, lamentablemente no existen prácticas y procedimientos manuales o informáticos que permitan analizar y prevenir la deserción teniendo en cuenta el aprovechamiento académico, solo contando con investigaciones superficiales y análisis fundados en métodos y modelos estadísticos (3) (4) (5).

El Colegio Nacional de la Libertad es una Institución Educativa descentralizada de la UGEL Huaraz, la Dirección Regional de Educación y el Ministerio de Educación, considerada como una Institución Educativa que brinda servicios de educación básica regular en los niveles de primaria en el turno mañana, secundaria en el turno tarde y educación básica alternativa en el turno noche dirigido a los pobladores de su jurisdicción y alrededores, uno de los principales problemas identificados en la Institución Educativa es el abandono de los estudios, conocido también como deserción escolar, ésta problemática se presenta mayormente en los alumnos de educación secundaria, el proceso utilizado actualmente que permite identificar los posibles casos de

riesgo no se encuentra implementado mediante el uso de las tecnologías de la información y comunicación (TIC) el cual no permite facilitar información en tiempo real. Por ello el presente trabajo de investigación propone la implementación de Minería de Datos que permita identificar patrones de comportamiento de la deserción escolar logrando intervenir de manera oportuna y revertir posibles decisiones drásticas de los alumnos que conlleve al abandono de sus estudios.

Debido a ésta situación, se planteó el siguiente enunciado del problema: ¿De qué manera la propuesta de implementación de Minería de Datos para el Colegio Nacional de la Libertad en la Ciudad de Huaraz permitirá determinar patrones de comportamiento de la deserción escolar?

Con la finalidad de poder dar solución a esta problemática se definió el siguiente objetivo general: Proponer la implementación de Minería de Datos para el Colegio Nacional de la Libertad en la ciudad de Huaraz, con la finalidad de determinar patrones de comportamiento de la deserción escolar.

Y para poder lograr cumplir con el objetivo propuesto anteriormente se definieron los siguientes objetivos específicos:

1. Recolectar información de acuerdo a los requerimientos que ayuden a determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad.
2. Utilizar la metodología adecuada, para el análisis y diseño de la solución de minería de datos que permita cumplir con los requerimientos establecidos en el Colegio Nacional de la Libertad.

3. Diseñar la implementación de Minería de Datos, que permita la determinación de patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad.

El presente trabajo de investigación tiene como justificación académica, el emplear los conocimientos obtenidos durante el proceso de formación académica en la Universidad Católica Los Ángeles de Chimbote, junto a la experiencia obtenida en el desarrollo de las practicas pre profesional y las responsabilidades que desempeño en el campo laboral, valdrá de cimiento para solucionar los requerimientos establecidos de la institución educativa en análisis y estudio.

Como justificación operativa, la implementación de Minería de Datos permitirá determinar, mediante el aprendizaje automático basado en la ciencia de los algoritmos, la posibilidad de extraer patrones de comportamiento en la deserción escolar los cuales facilitará a los directivos el análisis e intervención de manera anticipada por medio de la predicción.

Como justificación económica, la institución educativa contará, de manera permanente y continuán, con la información actualizada de manera anticipada de los posibles casos de deserción escolar que se podrían presentar mediante el análisis de patrones de comportamiento detectados permitiendo que la institución educativa y por consecuencia el estado peruano no pierda a un ciudadano que en el futuro pagará impuestos y generara ingresos activando la economía gracias a los conocimientos adquiridos y la profesionalización que obtendrá al termino de sus estudios escolares.

Como justificación tecnológica, la implementación de Minería de Datos logrará la utilización de las tecnologías de la información y comunicación (TIC) mediante las diversas herramientas tecnológicas las cuales estarán puestas a disposición en beneficio de la institución educativa brindando

solución a los requerimientos y la problemática identificada permitiendo que se actué con anticipación a los posibles casos presentados.

Como justificación institucional, la problemática investigada es la deserción escolar del cual muchos autores coinciden que es un problema socioeducativo refiriéndose al número de estudiantes que han abandonado sus estudios en cualquier momento del año escolar siendo un proceso en el que se encuentran involucrados múltiples factores y causas, el colegio Nacional de la Libertad se encuentra inmerso al igual que todos los colegios tanto privados como públicos siendo este último el de mayor índice en las estadísticas dentro de los últimos años.

El alcance de la investigación es local teniendo en cuenta que el Colegio Nacional de la Libertad es una institución educativa que ofrece educación escolarizada en los niveles de primaria y secundaria contando con 35 alumnos por sección; para el presente trabajo de investigación se comprende la participación de los alumnos del quinto año de educación secundaria con sus respectivos docentes de aula, con la posibilidad de ser implementado tanto en los demás grados como en otras Instituciones Educativas públicas.

La presente investigación, en su metodología, utiliza un enfoque cuantitativo de tipo descriptiva, así mismo su diseño es no experimental y por las características de su ejecución es de corte transversal o también conocidos como diseños de investigación transeccional.

El resumen general de resultados se puede observar que, en la primera dimensión sobre el nivel de satisfacción sobre el proceso actual de detección de posibles casos de riesgo se observó que el 81,40% de los encuestados manifestaron que, Sí están satisfechos con el método actual de detección de posibles casos de abandono escolar, mientras que, el 18,60% de los encuestados manifestaron que, No están satisfechos con el sistema actual, y

respecto a la dimensión 2: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento, se observa que, el 100,00% de los encuestados manifestaron que, Sí están de acuerdo con la implementación de Minería de Datos que permita determinación de patrones de comportamiento de la deserción escolar, dichos resultados son concordantes con las hipótesis específicas y quedando demostrada y justificada la hipótesis.

El análisis concluyó que los resultados coinciden con los antecedentes y bases teóricas mencionadas anteriormente y teniendo en cuenta el proceso actual de identificación de posibles casos de riesgo, permiten afirmar que existe la necesidad de contar con la implementación de Minería de Datos que permita al personal docente, auxiliares y personal administrativo de la Institución educativa, identificar los patrones de comportamiento y posibles casos de riesgo de deserción escolar.



## II. REVISIÓN DE LA LITERATURA

### 2.1. Antecedentes

#### 2.1.1. Antecedentes a nivel internacional

En el año 2016, la autor Oñate A. (6), realizó una tesis titulada “Análisis de la Deserción y Permanencia Académica en la Educación Superior aplicando Minería De Datos”, de la Universidad Nacional de Colombia, el objetivo es obtener información para identificar las características y perfiles de los estudiantes admitidos y clasificación de los estudiantes en los cuatro semestres académicos que presentaron bloqueo por bajo rendimiento académico, la metodología de investigación es CRISP-DM, obtuvo como resultados que en la evaluación de la predicción de pérdida de la condición académica utilizando los datos de entrada del proceso de admisión, fueran entrenados y probados en diferentes modelos, para clasificar estudiantes con bloqueo académico en las primeras cuatro matriculas académicas, se utilizó la información inicial del proceso de admisión y el historial académico de la matrícula anterior, concluye que el análisis de los datos pudo evidenciar que efectivamente hay diferentes tipos de desempeño de acuerdo el perfil socioeconómico del estudiante y el historial académico, demostrando que es factible hacer predicciones y que esta investigación puede ser una herramienta muy útil para la toma de decisiones, finalmente recomienda que este trabajo puede ser utilizado para la toma de decisiones, por parte del programa de permanencia y graduación de la Universidad y puede ser utilizado como punto de partida para futuras investigaciones de minería de datos en la educación.

En el año 2016, el autor Vásquez J. (7), realizó una tesis titulada “Modelo predictivo para estimar la deserción de estudiantes en una institución de educación superior”, de la Universidad de Chile en Santiago de Chile, el objetivo es construir modelos predictivos para la detección temprana de las deserciones en el programa de Ingeniería en Información y Control de Gestión de la Facultad de Economía y Negocios de la Universidad de Chile, la metodología de investigación fue Knowledge Discovery in Data Bases (KDD), obtuvo como resultados que se pudo identificar los predictores más importantes para cada semestre que son el rendimiento en la PSU, el número de padres vivos, la evaluación de los alumnos que realizaban a los profesores de manera semestral y el rendimiento académico universitario, concluye que, los seis modelos generados permitirán a la escuela predecir de manera semestral el conjunto de alumnos que podría desertar, y de esta manera aplicar políticas educacionales sobre este grupo de estudiantes para reducir los efectos negativos la deserción, finalmente recomienda que es importante agrupar los alumnos según los antecedentes familiares y el rendimiento que tienen por semestre, el hecho de que 5 de los 6 mejores modelos estén compuestos por técnicas de clusterización responde a que los alumnos tienden a agruparse y dependiendo de sus características, algunos de ellos desertan y los otros se mantienen en la carrera.

En el año 2015, el autor Córdova J. (8), realizó una tesis titulada “Aplicación de Técnicas de Minería de Datos para predecir la deserción de los estudiantes que pertenecen al Colegio Fisco-Misional “San Francisco” de la Ciudad de Ibarra”, de la Universidad Regional Autónoma de los Andes de Ibarra Ecuador, el objetivo es implantar la Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes que pertenecen

al Colegio "San Francisco" de la Ciudad de Ibarra, la metodología de investigación fue el paradigma mixto, es decir, la combinación de los enfoques cuantitativos y cualitativos, obtuvo como resultados que, el modelo predictivo obtenido, a través de la técnica de clustering ha ofrecido resultados satisfactorios, ya que permitió conocer las características primordiales de un posible desertor, las mismas que principalmente están relacionadas con el nivel académico del estudiante, concluye que la validación de la propuesta que se realizó con la evaluación participativa y de monitoreo, logró el análisis de cada uno de los resultados obtenidos con la participación de la institución; de acuerdo a ello se tomó las decisiones pertinentes para la solución del problema planteado, finalmente recomendó que para poder crear un modelo de forma correcta y ordenada es importante, hacer uso de una metodología para el desarrollo de proyectos de minería de datos, siendo en la actualidad CRISP DM una de las más utilizadas, ya que la misma propone las fases necesarias, para generar un modelo de calidad.

#### 2.1.2. Antecedentes a nivel nacional

En el año 2018, el autor Yamao, E. (9), realizó una investigación titulada "Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la Escuela Profesional de Ingeniería de Computación y Sistemas", de la Universidad de San Martín de Porres de Lima, el objetivo es predecir el rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la Escuela Profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres, la metodología del enfoque de investigación fue cuantitativo, en sus resultados, se ha logrado demostrar la importancia de las técnicas de minería de datos para realizar análisis estadístico y predicción a base de la data

relacionada con los ingresantes. No se han aplicado para este estudio técnicas de inteligencia artificial como redes neuronales que podrían ser consideradas como una continuación de este estudio para ver la posibilidad de alcanzar modelos con mayor capacidad de predicción, se concluye que se logró predecir a través de técnicas estadísticas y minería de datos el rendimiento académico de los estudiantes ingresantes a la Escuela Profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres, finalmente se recomendó que la data obtenida de las bases de datos de la universidad provenía de sistemas transaccionales por la que en muchos casos el formato y el tipo de datos no eran los ideales y fue necesario realizar varios procesos de limpieza y transformación de datos.

En el año 2016, el autor Piscoya L. (10), realizó una investigación titulada “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la Región De Lambayeque”, de la Universidad Señor de Sipán de Lambayeque, el objetivo es aplicar técnicas de minería de datos para predecir la deserción estudiantil de la educación básica regular Lambayecana, la metodología de investigación es de tipo tecnológica y diseño experimental, en sus resultados se logró recopilar y analizar la información brindada por la Ugel Chiclayo, se hizo el análisis de los archivos ofimáticos con los datos históricos de todos los alumnos matriculados, donde se determinó que se tuvo que homogenizar los campos, donde inicialmente comprendía 4831254 registros obtenidos están entre los años 2006 hasta 2015, se concluye que se construyó una aplicación web para evaluar los resultados obtenidos, el sistema se diseñó en php obteniendo una interfaz donde es capaz de interactuar con el servidor a fin de ejecutar los modelos ya sean reales o simulaciones donde se extrae

el histórico de cada colegio analizado la data warehouse, para que el usuario realice las pruebas pertinentes, finalmente se recomendó un licenciamiento en cuanto para el software SPSS. Para el tratamiento de los datos, los tratamientos de valores nulos en datos de esta naturaleza deben ser tratados con el mayor detalle posible, una matriz consolidada permitió identificar los valores faltantes en la base de datos que podían ocasionar daños en los cálculos de la serie y la considerar un formato homogéneo.

En el año 2015, los autores Gálvez M. y Flores K. (11), realizaron una investigación titulada “Modelo predictivo de deserción universitaria de la carrera de Ingeniería Informática”, de la Universidad Ricardo Palma de Lima, el objetivo es implementar un modelo predictivo de deserción estudiantil de pregrado en la escuela ingeniería informática de la Universidad Ricardo Palma, la metodología empleada es CRISP-DM consta de cuatro niveles de abstracción, organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos, en sus resultados, se identifican los principales factores que desencadenan la deserción en los alumnos de la escuela informática, se aprecia que las variables que determinan la deserción de los alumnos están divididas en 3 niveles, teniendo en el primer nivel la variable más predominante en el resultado, se concluye que el factor económico y académico son los más predominantes en el problema de la deserción, al analizar las diferentes técnicas que se utilizan en la minería de datos específicamente para la predicción, se observa que los arboles de decisión resultan ser buenos clasificadores, según los resultados obtenidos, finalmente recomendó que en este trabajo permitió identificar que la universidad no recolecta la suficiente información referente a la caracterización del estudiante al momento de ingresar a estudiar, que permitan establecer modelos

de predicción, para obtener un mejor resultado se deberá de trabajar con datos reales de los alumnos, para poder lograr los niveles de predicción necesarios para la validación positiva de los modelos.

### 2.1.3. Antecedentes a nivel regional

En el año 2017, el autor Duran E. (12), realizó una investigación titulada “Data Mart como parte de una solución de inteligencia de negocios, para el soporte de la toma de decisiones de la gestión académica de la Facultad de Ciencias en la UNASAM”, de la Universidad Nacional Santiago Antúnez de Mayolo de Huaraz, el objetivo es implementar un Data Mart, como parte de una solución de inteligencia de negocio, para mejorar el proceso de Toma de Decisiones en La Gestión Académica de Facultad de Ciencias en la UNASAM, la metodología de investigación fue el método experimental, en sus resultados se logró mejorar significativamente el proceso de Toma de Decisiones en la Gestión Académica de Facultad la Ciencias de la UNASAM, también se logró la implementación de un Data Mart, como parte de una solución de inteligencia de negocio influye significativamente en el diseño de la toma de decisiones de la gestión académica, concluye que según los resultados el 33.3% de los tomadores de decisiones muestran un nivel muy bueno se encuentra en el nivel muy bueno y el 66,7% de los tomadores de decisiones muestran un nivel de bueno. El  $p=0,000$ ; por lo que queda demostrado la influencia del Data Mart sobre la elección de decisiones de la gestión académica, finalmente recomendó el empleo de esta metodología DWEP para el desarrollo de Data Mart o Data Warehouse, porque es una metodología robusta y muy útil que se caracteriza por ser incremental e iterativo.

En el año 2016, el autor Torres L. (13), realizó una investigación titulada “Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz – 2016”, de la universidad Cesar Vallejo de la Filial Huaraz, el objetivo es descubrir patrones de deserción estudiantil e identificar los factores que influyen en la deserción relacionados con los datos operativos de los estudiantes, la metodología de investigación tuvo un enfoque empírico, sus resultados establecen un patrón de comportamiento para los estudiantes del cuarto ciclo que obtienen un promedio menor o igual a 12 y que desertan en el quinto ciclo, indicando que el promedio obtenido en el cuarto ciclo es un factor que influye en la deserción estudiantil, se concluye que se cumplieron los objetivos planteados para la presente investigación, pues se estableció un repositorio de datos para su análisis mediante minería de datos, se descubrieron algunos patrones de deserción estudiantil y sus factores asociados, y se estableció un procedimiento para el estudio de la deserción estudiantil con minería de datos educacional, finalmente recomendó que se debe ampliar la cobertura del cuestionario de la carpeta del postulante, con la finalidad de recopilar los datos relacionados con el entorno familiar, social y económico de los estudiantes, para ser utilizados en estudios futuros y realizar acciones de seguimiento del rendimiento escolar de los estudiantes, con la finalidad de realizar labores de prevención de la deserción estudiantil, toda vez que el promedio obtenido en los ciclos de estudio es un factor de deserción.

En el año 2015, el autor Díaz S. (14), realizó una investigación titulada “Sistema integral bajo el enfoque de minería de datos y redes neuronales para la predicción y control de la contaminación atmosférica por PM10 en la ciudad de Chimbote”, de la universidad

nacional del santa de Chimbote, el objetivo es implementar un sistema integral, basado en el uso de minería de datos con redes neuronales para la predicción de los niveles de contaminación atmosférica producido por PM10 en la ciudad de Chimbote, que permita anticiparse a situaciones de riesgo, la metodología de investigación es de tipo documental, básica, descriptiva y explicativa, en sus resultados se logró que las demostraciones realizadas a nivel de satisfacción de usuario y tiempo, los resultados caen en la región de aceptación; por lo tanto podemos decir que la hipótesis planteada es válida, también se logró demostrar un nivel porcentual de predicción del agente contaminante PM10, en un 91.00%, después de la aplicación del sistema integral y redes neuronales, en la ciudad de Chimbote, se concluye que se demostró a través de pruebas estadísticas que el Sistema Integral realmente cumple el objetivo de predecir los Índices de Calidad del Aire, respecto a la contaminación atmosférica por PM10, finalmente recomendó que se debe de realizar una capacitación a todos los usuarios involucrados en el uso del sistema, para explicarles las funcionalidades y poder levantar todas sus inquietudes.

## 2.2.Bases teóricas

### 2.2.1. Rubro de la empresa

El Colegio Nacional de la Libertad de la ciudad de Huaraz cuenta con personería jurídica de Institución Pública de derecho público interno y de duración indefinida inscrito formalmente en la Superintendencia Nacional de Administración Tributaria (SUNAT) desde el 24 de julio de 2006, contando actualmente con el estado de contribución activo RUC: 20531003765 (15).



## 2.2.2. La empresa Investigada

### - Información general

Es una Institución Educativa descentralizada de la UGEL Huaraz, la Dirección Regional de Educación y el Ministerio de Educación, considerado como centro de educación básica regular y básica alternativa, cuenta con los niveles de primaria en el turno mañana, secundaria en el turno tarde y la nocturna en el turno noche (16).

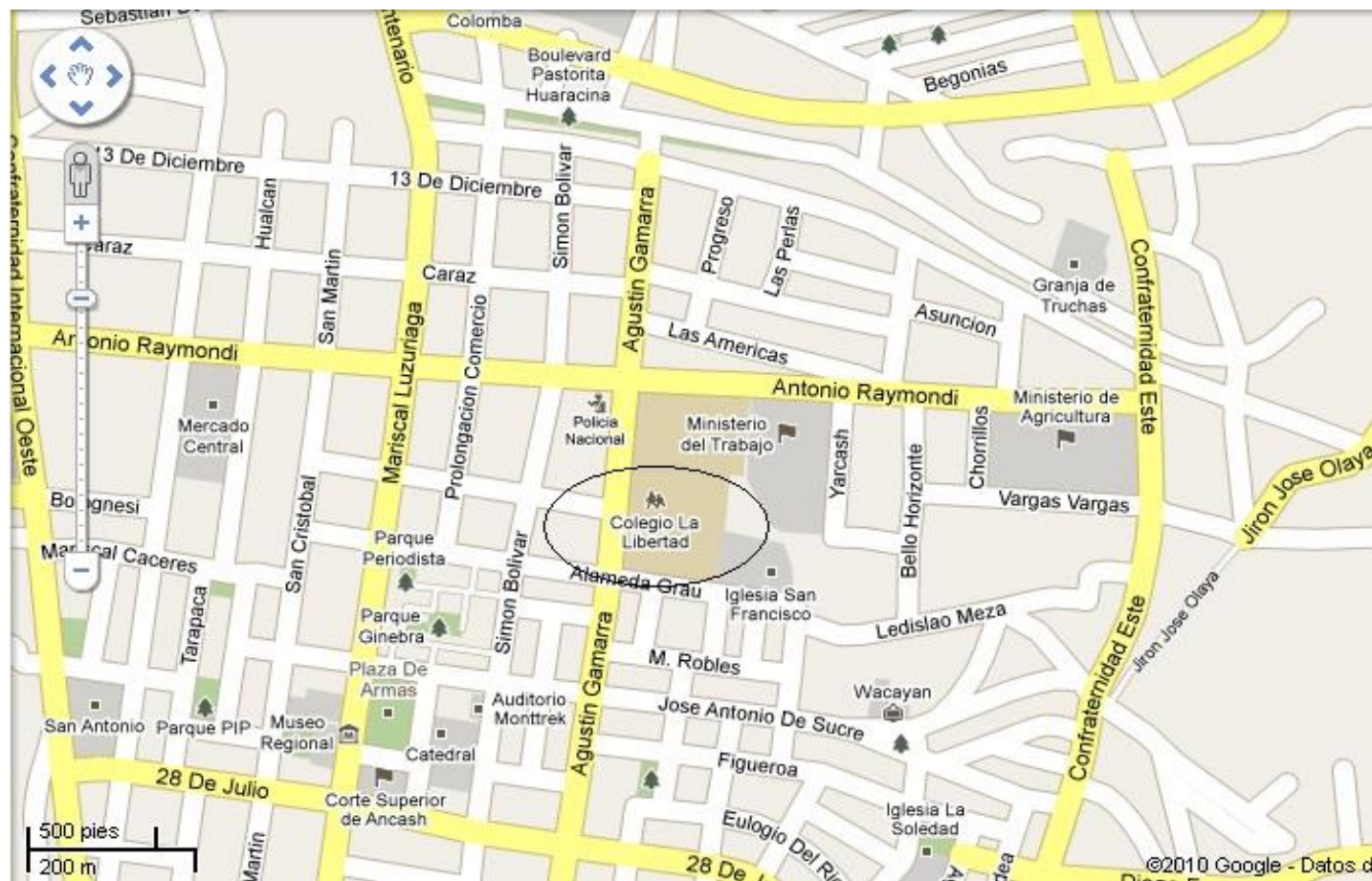
#### Denominación de la Organización:

Nombre : Colegio Nacional de la Libertad  
Sigla : CLL Huaraz.  
Ubicación : Provincia de Huaraz en la Región Ancash.  
Categorización : Institución Educativa Nacional.

#### Datos Geográficos de la Institución Educativa

Localización : Se encuentra ubicado en la Av. Agustín Gamarra N° 463 y Alameda Miguel Grau N° 519 Barrio de San Francisco en el Distrito y Provincia de Huaraz en la Región Ancash.  
Superficie : Cuenta con un área física conformada y saneada de aproximadamente de 1250 m<sup>2</sup>, de la cual existe un área techada de 428 m<sup>2</sup>.

Gráfico Nro. 1: Ubicación Geográfica de Colegio Nacional de la Libertad.



Fuente: Sitio web de Colegio Nacional de la Libertad (17).

- Historia

El Colegio De La Libertad alma matar de la cultura ancashina fue creada mediante Ley firmada el 31 de enero y promulgada el 1 de febrero de 1828 por el Mariscal José de la Mar, de cuyas aulas han egresado ilustres personalidades que tuvieron y tienen un rol protagónico de las diversas esferas de la sociedad regional, nacional y mundial, luego de crearse el célebre “Colegio De La Libertad”, tuvo como primer Rector a don José María Robles Arnao, luego le sucedieron en el cargo don Manuel Castillo y don Julián de Morales y Maguiña. En el año 1928, en las fiestas conmemorativas por el Primer Centenario, se entonó por primera vez, el himno de nuestra institución educativa; escrito por Alejandro Dextre Sierra y la música, por Antonio Guzmán Arenas (18).

En sus primeros años, formó ciudadanos con ideales nacionalistas y de liberación; posteriormente, la educación se realizó dentro de los lineamientos de la corriente conductista; con contenidos y métodos propios de una sociedad en vías de industrialización, hasta el año 1969 aproximadamente. A partir del año 1970, se impone el aprendizaje cognoscitivo; luego años más tarde, tiene lugar el enfoque cognitivo que busca el desarrollo de capacidades y valores por medio de metodologías activas. Este último paradigma, centrado en la formación del estudiante, es el que toma nuestra institución como principio de gestión; y se caracteriza por brindar una educación abierta al cambio, solidaria, participativa, democrática, analítica, creadora e innovadora (18).

Desde su creación la infraestructura de nuestra institución educativa, sufrió los estragos causados por los fenómenos naturales, como el aluvión del 13 de diciembre de 1941, el terremoto del 31 de mayo de 1970, y por último un fuerte ventarrón derribó el simbólico árbol de poroto el 25 de abril del 2003, que contaba con más de cien años de vida en los jardines de nuestro colegio, en la actualidad la institución educativa cuenta con una infraestructura educativa moderna que constituye una condición importante y favorable para el aprendizaje de nuestros estudiantes, espacios educativos que son aprovechados para fines pedagógicos y orientados a los temas transversales del currículo nacional actual que orientan el actuar y la convivencia de los actores educativos de nuestra Institución (18).

El presente año la Institución Educativa Emblemática “Colegio De La Libertad” de Huaraz el 19 de setiembre celebra sus 192 Aniversario con el compromiso de seguir trabajando por la mejora de los aprendizajes y la concreción de nuestra visión institucional que considera a nuestros estudiantes como personas competentes, que vivencian valores, respetuosos de la diversidad sociocultural y conciencia ambiental; con capacidades creativas, críticas, reflexivas y emprendedoras para enfrentar los desafíos en una sociedad democrática de manera autónoma y asertiva; con docentes y padres de familia comprometidos en la búsqueda de la excelencia y del bien común (18).

- Objetivos organizacionales

Dentro de los principales objetivos considerados por la institución tenemos:

- Promover el uso de técnicas y estrategias de lectura y estudio, mediante actividades lúdicas y recreativas para mejorar las capacidades comunicativas de los estudiantes.
- Desarrollar actividades matemáticas contextualizadas que permitan la construcción de nociones y conceptos para mejorar el desarrollo de capacidades matemáticas.
- Mejorar la convivencia escolar, mediante la mediación y el diálogo entre pares, para garantizar una convivencia armónica en la Institución Educativa.
- Optimizar la elaboración de la programación curricular, mediante la promoción del trabajo colegiado por grados y áreas, para lograr un desempeño docente eficiente.
- Garantizar el cuidado y limpieza del ambiente escolar, mediante acciones de sensibilización utilizando diversas estrategias, para preservar la infraestructura institucional.
- Desarrollar en los estudiantes capacidades, conocimientos y actitudes emprendedoras y empresariales, mediante la ejecución de proyectos productivos que ayuden a fomentar la cultura del

emprendimiento aprovechando los recursos propios, de la I.E. y de la localidad, para generar alternativas de autoempleo.

- Desarrollar la gestión ambiental dentro de la familia, la I.E. y comunidad, mediante la ejecución de proyectos educativos ambientales integrados, para la conservación y protección de los recursos naturales dentro y fuera de la Institución Educativa.

- **Visión**

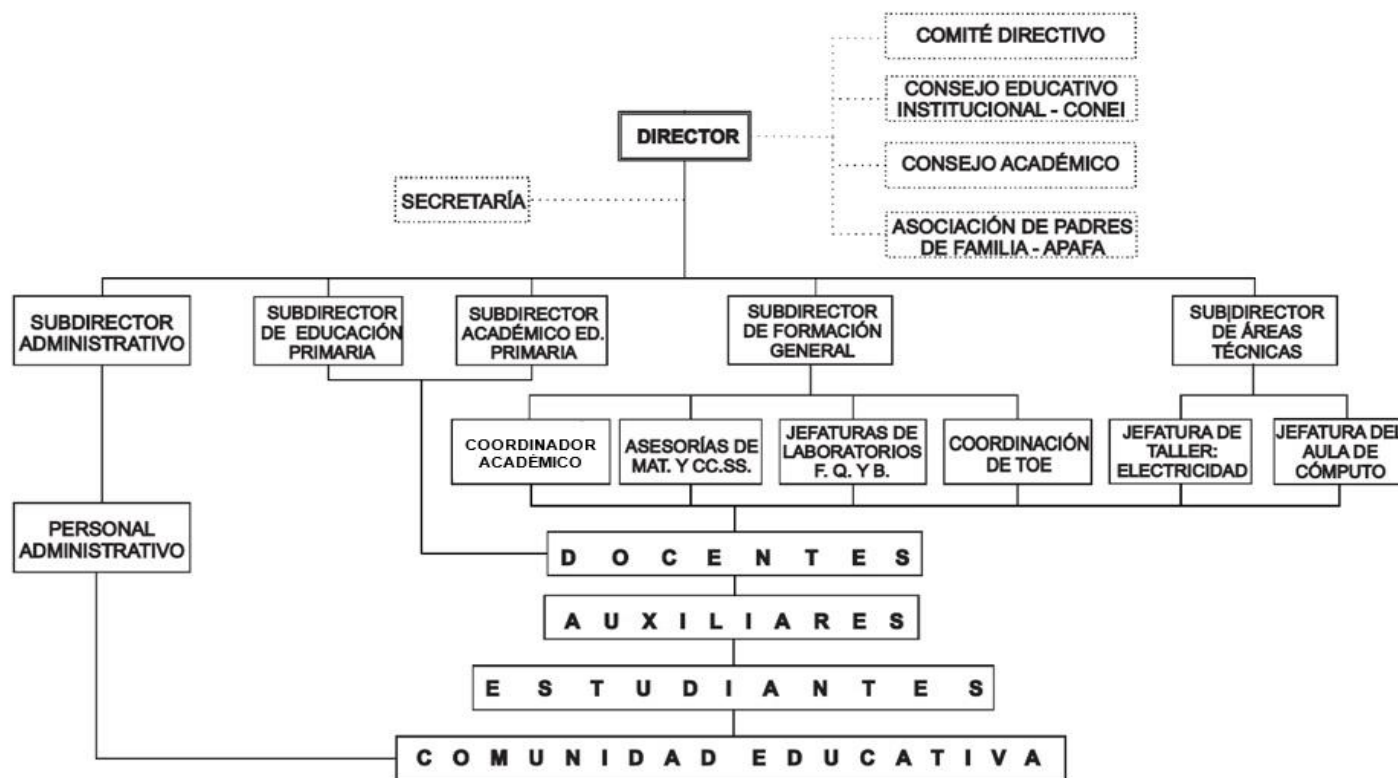
2020 nuestros estudiantes son personas competentes, que vivencian valores, respetuosos de la diversidad sociocultural y conciencia ambiental; con capacidades creativas, críticas, reflexivas y emprendedoras para enfrentar los desafíos en una sociedad democrática de manera autónoma y asertiva; con docentes y padres de familia comprometidos en la búsqueda de la excelencia y del bien común.

- **Misión**

Somos una institución educativa que brinda un servicio educativo de calidad científica, tecnológica y humanística. Formamos integralmente a nuestros estudiantes fortaleciendo sus habilidades y destrezas, para el desarrollo de sus capacidades, con un enfoque por competencias, contando para ello con docentes calificados y comprometidos con el quehacer educativo; promoviendo con el ejemplo la práctica de valores, las actitudes democráticas y el compromiso con el bienestar de su comunidad.

- Organigrama

Gráfico Nro. 2: Organigrama del Colegio Nacional de la Libertad – Huaraz



Fuente: Colegio Nacional de la Libertad (16).

- TIC que utiliza la empresa investigada

La institución educativa para el área administrativa cuenta con cuatro computadoras de escritorio, 2 impresoras multifuncional, una fotocopidora y un multimedia, para las dos aulas de innovación tecnológica cuentan con 25 y 32 computadoras respectivamente, 2 impresora multifuncional dos multimedia, para las aulas se cuenta con 23 computadoras portátiles con sus respectivos accesorios, 15 equipos multimedia. Respecto al software instalados en los ordenadores, cuentan con la instalación del sistema operativo Windows XP, 2007, 2008 y 2010 según sus características, además cuentan con la suite de ofimática en sus diferentes versiones, también cuentan con sistemas congeladores de protección de información e instalación, por último, la instalación de diferentes programas del Ministerios de educación y demás programas necesario. Finalmente se cuenta con un servicio dedicado de internet el cual permite el acceso a las diferentes unidades y aulas virtuales que cuenta la Institución Educativa.

- Infraestructura tecnológica

La infraestructura del Colegio Nacional de la Libertad cuenta con el apoyo humano, los recursos, la evaluación, la relación entre la infraestructura tecnológica y la planificación académica, el acceso que tienen los estudiantes y los docentes a la tecnología que ofrece la Institución y el papel que tiene el Gobierno Nacional como facilitador en el acceso de estas nuevas tecnologías en pro de la Educación. A continuación se lista aquellos elementos vinculados directamente con la tecnología y los equipos informáticos: La I.E. cuenta con 115



computadoras distribuido en tres laboratorios, las cuales 21 de ellas no están en uso debido a su deterioro por falta de mantenimiento continuo, 15 computadoras portátiles, 6 switch de red, 3 router de red, 5 impresoras láser, 1 impresora a colores, 8 cámaras fotográficas, 6 cámaras de video, 3 scanner, 12 proyectores de video, 6 pantallas de proyección, 2 fax, 10 monitores de TV, 8 reproductores de DVD, 2 reproductores de Blu Ray, según el último inventario realizado en la institución educativa.

### 2.2.3. Las Tecnologías de información y comunicaciones (TIC)

#### - Definición

Las tecnologías de la información y comunicación TIC abarcan un pensamiento muy amplio y a su vez muy voluble, con relación a una lista de servicios, aplicaciones y tecnologías, que usan numerosos tipos de equipos computo (hardware), programas informáticos (software) y prioritariamente los que se utilizan para la comunicación por medio de la red de redes. Las TIC rotan en relación a cuatro tecnologías básicas: la informática, la microelectrónica, los medios multimedia y las telecomunicaciones siendo el más relevante, todos ellos rotan de forma interactiva e inter-conexionada, lo que facilita conseguir nuevas alternativas de comunicación y fomentar las que se encuentran en forma apartada. Las nuevas alternativas son los medios electrónicos que realizan procesos como crear, almacenar, recuperar y transmitir la información de forma cuantitativa, con velocidad y en grandes cantidades y lo realizan modificando múltiples códigos en un contexto de hipertexto (fusión de hipertexto con multimedia). Ante las múltiples

formas de definir las TIC, todas coinciden en contemplar a las tecnologías como instrumento técnico que rotan en relación a la información o comunicación entre ellas, mejor dicho, de alguna forma implícita las contemplan como medios que se usan para que se concrete el proceso de transmisión de la información y la comunicación (19).

- Historia

Existen infinidad de cambios y adaptaciones que la sociedad y el mundo ha podido experimentar particularmente en las últimas décadas y siglo pasado. El avance y la evolución de las TIC ha revolucionado de manera acelerada no sólo el manejo de las empresas sino también la forma de vida del ser humano en forma general. La sociedad, las escuelas, los hábitos y todo los demás que nos contextualiza en general ha experimentado una gran revolución tecnológica. Actualmente contamos con la posibilidad de volvermos a comunicar con personas de quienes perdimos por completo referencia alguna y no sabíamos nada desde hace muchos años y a la vez tenemos la oportunidad de conocer a nuevas personas que gracias a la evolución de las tecnologías de comunicación jamás las hubiéramos conocido e interactuado. Las redes sociales más populares y públicas como son: Facebook, Twitter e Instagram como también las privadas que son utilizadas para poder relacionarse con otras personas, han apoyado en el desarrollo global de la TIC. Actualmente nuestra forma de comprar y hacer negocios ha cambiado de igual modo, presentando nuevas alternativas. En general las TIC han influenciado y redescubierto en nuestros propios gustos, en nuestras aficiones o simpatías y en nuestros hábitos de consumo (20).

- Las TIC más utilizadas en la empresa investigada

El Colegio Nacional de la Libertad es una Institución Educativa que cuenta con una sitio web donde existe información básica, el uso de las redes sociales para la interacción con los alumnos, cuenta con un correo institucional un centro de cómputo implementado con el hardware y los software con licencia para el correcto uso de los ordenadores para el desarrollo de las clases de computación y una área adicional de uso exclusivo de los docentes para el desarrollo de una clases y de consulta mediante el uso de la internet.

#### 2.2.4. Teoría relacionada con la Tecnología de la investigación

##### 2.2.4.1. Definiciones

Principales definiciones relacionadas al presente trabajo de investigación:

- Base de Datos

Es un conjunto de información denominada datos que son almacenados en la memoria del disco duro, están organizados mediante una estructura de datos, cada base de datos ha sido diseñada para satisfacer los requisitos de información de una empresa u otro tipo de organización, antes de existir las bases de datos se trabajaba con sistemas de ficheros lo que surgieron al informatizar el manejo de los archivadores manuales para proporcionar un acceso más eficiente a los datos almacenados en los mismos. Una base de datos se puede

percibir como un gran almacén de datos que se define y se crea una sola vez, y que se utiliza al mismo tiempo por distintos usuarios. En una base de datos todos los datos se integran con una mínima cantidad de duplicidad. De este modo, la base de datos no pertenece a un solo departamento, sino que se comparte por toda la organización. Además, la base de datos no sólo contiene los datos de la organización, también almacena una descripción de dichos datos. Esta descripción es lo que se denomina metadatos, se almacena en el diccionario de datos o catálogo y es lo que permite que exista independencia de datos lógica-física (21).

- Minería de Datos

Es un proceso de análisis y búsqueda en bases de datos de gran capacidad (Big Data) para identificar y clasificar información útil que pueda servir para la toma de decisiones en una organización, entendido también como la tecnología y software empleado para identificar patrones de comportamiento dentro de la base de datos almacenada, el punto fundamental de la minería de datos es que los patrones identificados apoyen en mejorar la toma de decisiones (22).

Ventajas: La ventaja principal de la minería de datos es la infinidad de escenarios empresariales u organizacionales en los que se puede implementar, a modo de ejemplo tenemos (22):

- La Predicción: Basado principalmente en el pronóstico que permite tomar decisiones que cambian el futuro de la organización (22).
- Probabilidad: Permite la selección de información calificada para el análisis o la estadística que permita un manejo o tratamiento personalizado (22).
- Análisis de secuencias: Directamente relacionado a comportamiento de la información teniendo en cuenta su continuidad e interacción con el resto de la información (22).

Etapas de la minería de datos: Dentro de un proceso de minería de datos podemos encontrar cinco fases (22):

- Objetivo y recolección de datos: Lo primero de todo es centrarnos de en qué tipo de información queremos obtener. Imaginemos el ejemplo que un supermercado quiere conocer a qué hora del día es donde más asistencia de clientes hay. Este sería el objetivo y la información que quiere obtener el comercio en este caso (22).
- Procesamiento y gestión de los datos: Una vez que sabemos los datos que queremos recopilar ponemos a trabajar a los datos. Esta quizás sea la fase más complicada del proceso. Pues requiere seleccionar la muestra representativa sobre la que se va a realizar el análisis. Una vez

escogida la muestra se debe analizar qué tipo de variables o modelo de regresión se va a realizar sobre la muestra (22).

- Selección del modelo: Está muy relacionado con la anterior fase. Se trata de crear un modelo o Algoritmo que nos arroje el mejor resultado posible. Para ello hay que hacer un análisis exhaustivo de las variables a incluir en el modelo. Esto se convierte en una tarea complicada ya que dependerá del tipo de información a analizar. Por ello, los mineros de datos llevan a cabo distintos exámenes del algoritmo como: regresión lineal, árbol de decisión, series temporales, red neuronal, etc (22).
  
- Análisis y revisión de resultados: Básicamente es analizar los resultados para comprobar si arrojan una explicación lógica. Explicación que facilite la toma de decisiones en base a la información suministrada por los resultados (22).
  
- Actualización del modelo: El último paso del proceso sería la actualización del modelo. Es muy importante que se vaya haciendo con el paso del tiempo para que no quede obsoleto. Las variables del modelo podrían pasar a ser no significativas y por tanto se requiere un control periódico del mismo (22).

- **Árbol de decisión**

Es un diagrama que contiene un nodo raíz donde se encuentran todas las observaciones, nodos internos que albergan a los nodos de división y los nodos hoja que contiene la clasificación final para un conjunto de observaciones. Los arboles de decisión son parte de las técnicas de minería de datos. Un árbol representa una segmentación de los datos, que se crea mediante la aplicación, de una serie de reglas simples. Cada regla asigna una observación, a un segmento basada en el valor de una entrada. Una regla se aplica después de otra, dando como resultado una jerarquía de segmentos dentro de segmentos. La jerarquía se llama árbol y cada segmento se llama nodo, así los nodos internos de un árbol representan validaciones sobre los atributos, las ramas representan las salidas de las validaciones, y los “nodos hoja” representan las clases (23).

- **Algoritmo CART**

El algoritmo Árboles de Clasificación y de Regresión (Classification And Regression Trees – CART), realiza particiones binarias, con el objetivo que la media de cada rama sea diferente y, por tanto, discrimine con suficiente precisión, un número adecuado de particiones, para asignar a cada hoja un valor cercano a la media de los elementos que caen en ella. Este algoritmo genera arboles de fácil interpretación con resultados óptimos, lo que se considera una ventaja, al crear modelos predictivos (23).

- Herramienta RPART

La herramienta Árboles de regresión y particionamiento recursivos (Recursive Partitioning And Regression Trees – RPART), es una función implementada para obtener la construcción del árbol su principal característica es que permite crecer el árbol deteniéndose cuando cierto criterio se alcanza teniendo en cuenta que detiene el crecimiento cuando el decremento de la desviación va por debajo de cierto umbral delimitado Este algoritmo utiliza, el proceso “Top-Down” para construir, el árbol de arriba hacia abajo, cuenta con variables nominales y cuantitativas y son apropiado en su aplicación (23).

- Modelo OLAP

Es una solución utilizada en el campo de la llamada inteligencia empresarial cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (o cubos OLAP) que contienen datos resumidos de grandes bases de datos o Sistemas Transaccionales (OLTP). Se usa en informes de negocios de ventas, marketing, informes de dirección, minería de datos y áreas similares, la razón de usar OLAP para las consultas es la rapidez de respuesta, una base de datos relacional almacena entidades en tablas discretas si han sido normalizadas (24).



- Inteligencia de Negocios

Se caracteriza por la implementación de una gran variedad de tecnologías, software con plataformas, especificaciones de aplicaciones y diversos procesos, su principal objetivo es contribuir a la toma de decisiones que permitan la mejora del desempeño de una empresa o conjunto organizacional y promover la ventaja competitiva. La Inteligencia de Negocios faculta a las organizaciones a la toma de mejores decisiones de manera rápida y eficiente y precisa esto quiere decir que es necesario analizar desde tres perspectivas las cuales son: Hacer mejores decisiones rápidamente, convertir o procesar datos en información y utilizar la aplicación adecuada para su administración (25).

- Inteligencia Artificial

Es el campo científico relacionado a la informática centrado básicamente en la creación de aplicaciones, programas o procedimiento o mecanismos que permitan descifrar comportamientos que son considerados inteligentes, mejor dicho “las máquinas piensan como seres humanos”, regularmente son capaces de analizar grandes cantidades de datos (big data) identificando patrones y tendencias de comportamiento los cuales permiten formular predicciones automáticamente con mucha rapidez y alta precisión. (26).

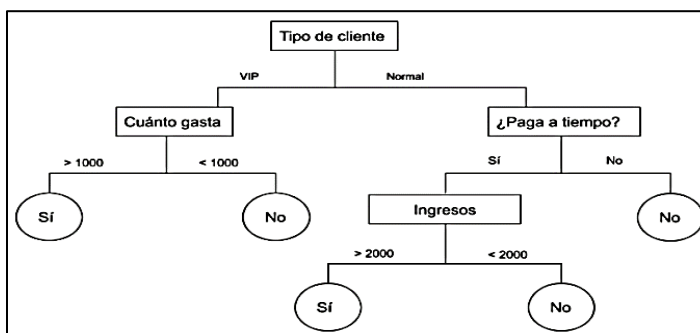
- Sistema Experto

Son aplicaciones computacionales sistematizadas de manera interactiva y confiable, con la capacidad de realizar toma de decisiones y afrontar problemas o dificultades comunes o complejos. La Inteligencia Artificial simula este proceso, pero cuando se trata de sistemas expertos cambia el sentido basándose a la resolución del problema o dificultad en un dominio específico. Los sistemas expertos fueron los antecesores de los sistemas actuales de inteligencia artificial, aprendizaje profundo y aprendizaje automático, identificando por medio de una variedad de características que proporciona eficiencia, precisión y resolución imaginativa de problemas. Son también operadores confiables que no deben emitir ningún tipo de error, son muy flexibles y el mecanismo de compilación de información es muy efectivo (27).

- Técnicas de Minería de Datos

Arboles de Decisiones: Son estructuras de forma de árbol que representan conjuntos de decisiones que generan reglas para la clasificación de un conjunto de datos, las cuales explican el comportamiento de una variable con relación a otras, y pueden traducirse fácilmente en reglas de negocio, son utilizados con finalidad predictiva y de clasificación por ejemplo se pueden emplear para optimizar respuestas de campañas, identificar clientes potenciales o realizar evaluación de riesgos (28).

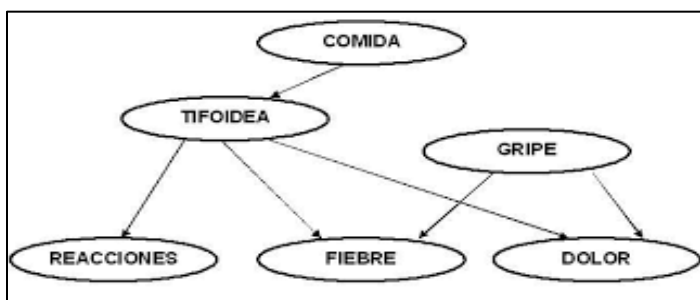
Gráfico Nro. 3: Ejemplo de un árbol de decisiones.



Fuente: Sitio web de Conecta Software (29).

Métodos Bayesianos: Consiste en representar todos los posibles sucesos en los que se está interesado mediante una representación simbólica de probabilidades condicionales de transición entre sucesos, puede codificarse a partir del conocimiento de un experto o puede ser inferido a partir de los datos. Permite establecer relaciones causales y efectuar predicciones, estos métodos proporcionan una metodología práctica para la inferencia y predicción y, en última instancia, para tomar decisiones que involucran cantidades inciertas (28).

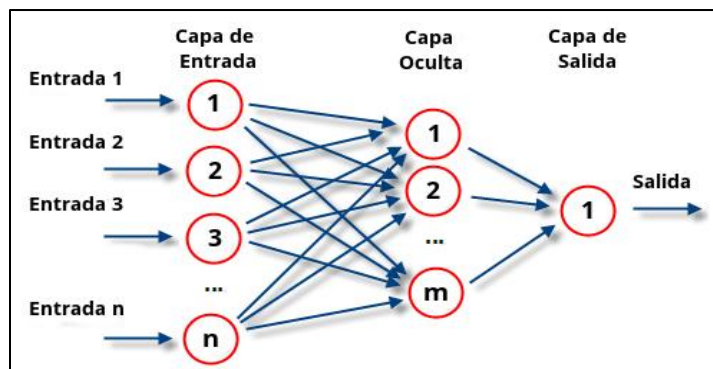
Gráfico Nro. 4: Ejemplo de Métodos Bayesianos.



Fuente: INAOEP (30).

Redes Neuronales Artificiales: Constituyen una técnica inspirada en los trabajos de investigación, iniciados en 1930, que pretendían modelar computacionalmente el aprendizaje humano llevado a cabo a través de las neuronas en el cerebro, son una nueva forma de analizar la información con una diferencia fundamental con respecto a las técnicas tradicionales: son capaces de detectar y aprender patrones y características dentro de los datos, se comportan de forma parecida a nuestro cerebro aprendiendo de la experiencia y el pasado y aplicando tal conocimiento a la resolución de problemas nuevos (31).

Gráfico Nro. 5: Ejemplo de Redes Neuronales.



Fuente: Wikipedia.org (32).

- Librerías de Minería de Datos

A continuación, se mencionan algunas de las librerías más importantes:

- Xelopes (Extended Library For Prudys Embedded Solution): Es una librería bajo la

licencia pública GNU para el desarrollo de aplicaciones de minería de datos. Esta librería está implementada para que sea eficiente para la mayoría de los algoritmos de aprendizaje, por eso, es importante destacar que el usuario puede desarrollar aplicaciones particulares de minería de datos. Sus principales características son las siguientes: Acceso a datos, Modelos de redes neuronales, Métodos de agrupamiento, Métodos de reglas de asociación, Árboles lineales, Árboles no lineales (33).

- Mlc++ (Machine Learning Library In C++): Es un conjunto de librerías que fueron desarrolladas por la Universidad de Standford. La mayoría de las versiones son bajo dominio de investigación, a excepción de la versión 1.3.x, que se distribuye bajo licencia de dominio público. Las principales características son: Acceso a datos, Transformaciones de datos, Métodos de aprendizaje mediante objetos (33).
- Suites: Posee las mismas capacidades que el procesamiento de datos, los modelos de análisis, el diseño de experimentos o el soporte gráfico para la visualización de resultados. En este caso, Suites destaca porque existe una interfaz que facilita la interacción entre el usuario y la herramienta (33).

- R-Project: Es un entorno de trabajo basado en los entornos de programación S y S-PLUS desarrollados a principios de los años noventa del pasado siglo por Bill Venables y David M. Como señalan Venables et al. (2011), es un entorno integrado de facilidades informáticas para la manipulación de datos, el cálculo y la generación de gráficos. R-Project pretende convertirse en un sistema internamente coherente que se caracterizaría por un desarrollo basado en la contribución relativamente altruista de la comunidad científica (33).
  
- Spss Clementine: Es uno de los sistemas de minería de datos más conocidos. Posee una herramienta visual desarrollada por ISL que tiene una arquitectura cliente / servidor. Este sistema se caracteriza por lo siguiente: Acceso a datos, Procesamiento de Datos, Técnicas de Aprendizaje, Técnicas de evaluación de modelos, Visualización de resultados, Exportación (33).
  
- Weka (Waikato Environment For Knowledge Analysis): Es una herramienta visual de libre distribución desarrollada por los investigadores de la Universidad de Waikato en Nueva Zelanda. Sus principales características son: Acceso a datos desde archivo en formato ARFF, Pre procesamiento de datos, Modelos de Aprendizaje, Visualización del entorno (33).

- Kepler: Sistema desarrollador y transformado en una herramienta comercial distribuida por Dialogis. Posee múltiples modelos de análisis. Sus principales herramientas de aprendizaje son las siguientes: Árboles de decisión, Redes neuronales, Regresión no lineal, Aplicaciones estadísticas (33).
- Odms (Oracle Data Mining Suite): Está diseñado sobre una arquitectura cliente servidor; ofrece una gran versatilidad en cuanto al acceso a grandes volúmenes de información. Se caracteriza principalmente por: Acceso a datos en diversos formatos: almacenes de datos, bases de datos relacionales como SQL, Oracle entre otros, Pre procesado de datos: muestreo de datos, patrones de datos, Modelos de aprendizaje: redes neuronales, regresión lineal, Herramientas de visualización (33).
- Yale: Herramienta de aprendizaje automático implementado en Java por la Universidad de Dortmund. El sistema incluye operaciones para lo siguiente: Importación y pre-procesamiento de datos, Aprendizaje automático, Validación de modelos (33).

#### 2.2.4.2.Casos de éxito

- En el año 2018, el autor Yamao, Eiriku (9), realizó una investigación titulada “Predicción del rendimiento

académico mediante minería de datos en estudiantes del primer ciclo de la Escuela Profesional de Ingeniería de Computación y Sistemas”, de la Universidad de San Martín de Porres de Lima, la metodología del enfoque de investigación fue cuantitativo.

- En el año 2016, el autor Piscoya L. (10), realizó una investigación titulada “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la Región De Lambayeque”, de la Universidad Señor de Sipán de Lambayeque, la metodología de investigación es de tipo tecnológica y diseño experimental.
- En el año 2015, los autores Gálvez M. y Flores K (11), realizaron una investigación titulada “Modelo predictivo de deserción universitaria de la carrera de Ingeniería Informática”, de la Universidad Ricardo Palma de Lima, la metodología empleada es CRISP-DM consta de cuatro niveles de abstracción, organizados de forma jerárquica en tareas que van desde el nivel más general hasta los casos más específicos.

#### 2.2.4.3. Metodologías de Minería de Datos.

A continuación, se mencionan los Procesos tradicionales de minería de datos más utilizadas:

- KDD (Knowledge Discovery Databases): El descubrimiento de conocimiento en bases de datos es un término común usado en la minería de datos, es un



proceso iterativo conformado por una serie de procedimientos para la generación del conocimiento, las fases de KDD son: Selección, donde se crea un conjunto de datos objetivo para el proceso de descubrimiento, el procesamiento o limpieza, donde se analiza la calidad de los datos, remoción de ruido, datos desconocidos, nulos duplicados, técnicas de estadística para el reemplazo, la transformación o reducción de datos se encarga de disminuir el número de variables o encontrar representaciones invariantes de los mismos, la etapa de minería de datos se encarga de la búsqueda de patrones usando tareas de descubrimiento como clasificación, clustering, patrones secuenciales y asociaciones. Creando modelos ya sean predictivos o descriptivos y la interpretación o evaluación se interpretan los patrones descubiertos, se puede retornar a las anteriores etapas para posteriores iteraciones (34).

- CRISP-DM (Cross Industry Standard Process for Data Mining): Es una metodología de referencia de las más utilizadas en proyectos de minería de datos, es un conjunto de tareas definidas en cuatro niveles (fases, tareas generales, tareas específicas e instancias del proceso), organizadas de forma jerárquica, el nivel superior está organizado por seis etapas y se caracteriza por enfatizar en los detalles de cada uno; dividiéndose en diferentes actividades y tareas: En la etapa de comprensión del negocio se definen los objetivos y requerimientos del proyecto desde el punto de vista del negocio para elaborar un plan con el fin de alcanzar los objetivos planteados, la comprensión de los datos consta

en la familiarización y recolección de datos para verificar la calidad y detectar si hay datos duplicados, inconsistentes, erróneos, incompletos, entre otros, la etapa de preparación de los datos se construye el conjunto de datos que se va a usar. Realizando limpieza de datos, creación de variables adicionales, integración de diferentes conjuntos de datos incluyendo también cambios de formato, la etapa de modelado selecciona las técnicas a usar para construir un modelo. Como existen diferentes técnicas sobre los datos para la solución de un mismo problema, en ocasiones es necesario volver a la etapa de preparación de los datos para la minería, en la evaluación, el modelo que se obtuvo es revisado para constatar que logró los objetivos del proyecto. Esta fase finaliza al aceptar los resultados del modelo y el despliegue del modelo, el conocimiento que se obtuvo es presentado de manera que sea fácil de entender e interpretar por el usuario final, ya sea por medio de un reporte, un aplicativo, etc (34).

- Catalyst: Se le conoce como P3TQ (Product, Place, Price, Time, Quantity) es conformada por 2 modelos: Modelo de negocio (MII) que es una guía para desarrollar y construir un modelo para resolver un problema de negocio, tiene cinco etapas: Dato (Conjunto de datos para explorar y encontrar patrones), Oportunidad (Problema u oportunidad de negocio), Prospectiva (Descubrir donde la minería de datos ofrece valor a la organización), Definido (Especificación del modelo de minería de datos con un propósito específico, Estratégico (Para dar soporte a un escenario planeado

por la organización) y Modelo de explotación de información (MIII) que es una guía para la ejecución y realización de modelos de minería de datos basados en el modelo (MII), ofrece una guía para la explotación de la información a través de las siguientes etapas: Preparación de los datos (Actividades que revisan calidad de los datos, características de variables, tamaño de datos, etc), Selección de herramientas y modelado inicial, Refinar el modelo seleccionado, Implementar el modelo, Comunicación de resultados (34).

### **III. HIPÓTESIS**

#### 3.1.Hipótesis General

La propuesta de implementación de Minería de Datos permite determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – Huaraz; 2020.

#### 3.2.Hipótesis específicas

- 3.2.1. La recolección de información de acuerdo a los requerimientos, ayuda a determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad.
- 3.2.2. La utilización de la metodología adecuada, para el análisis y diseño de la solución de minería de datos, permite cumplir con los requerimientos establecidos en el Colegio Nacional de la Libertad.
- 3.2.3. El diseño de la implementación de Minería de Datos, permite la determinación de patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad.

## **IV. METODOLOGÍA**

El presente trabajo de investigación tendrá un alcance o tipo de investigación descriptiva con enfoque cuantitativo.

### **4.1. Tipo de la investigación**

El Presente trabajo de investigación es de tipo descriptiva, el objetivo principal consiste en describir las circunstancias, comportamientos o eventos que se presenten por la implementación de procedimientos relacionados de manera detallada, teniendo en cuenta cómo se comportan y se manifiestan al ser sometidos a un análisis, en otras palabras, para el presente trabajo de investigación pretende determinar patrones mediante la investigación aplicada con el recojo de información de manera independiente los cuales utilizando conocimientos sobre minería de datos y diseño de sistemas expertos permitirá en una futura implementación reconocer los patrones de comportamiento de la deserción escolar (35).

### **4.2. Nivel de la investigación de la tesis**

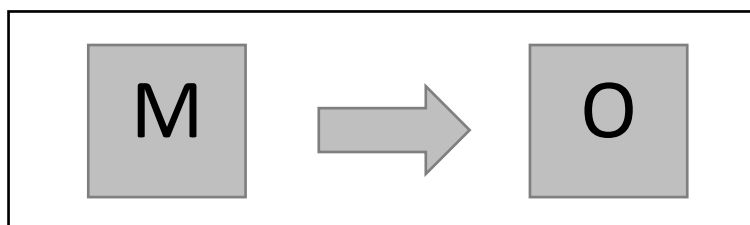
Por las características, el presente trabajo de investigación tiene un enfoque cuantitativo, estas características numéricas pueden estar presentadas en diferentes grados de intensidad y tienen carácter numérico. Esta permite distribuir a los individuos de acuerdo con ciertas características por medio de las cuales pueden distinguirse de otros individuos que no las poseen. Utiliza la recolección de datos para poder probar las hipótesis planteadas, teniendo como base la medición numérica y el análisis estadístico, los cuales permitirán establecer los patrones de comportamiento y así poder probar teorías (35).

#### 4.3. Diseño de la investigación

El diseño es no experimental, teniendo en cuenta que los estudios se realizarán sin la manipulación deliberada de las variables y se percibe los fenómenos en su ambiente natural aplicando la técnica de la observación, la característica de su ejecución será de corte transversal porque se realizará la evaluación en un tiempo determinado dentro del presente año.

- Investigación No Experimental: consiste específicamente en observar fenómenos en su contexto natural tal como se dan, para seguidamente ser analizados, se ejecuta sin la necesidad de manipular o altera de manera convenientemente o deliberada las variables independientes para probar el efecto que provoca sobre las demás variables, (35).
- Corte transversal: También conocidos como diseños de investigación transeccional, tiene como característica principal la de recogen datos en un determinado momento y en un único tiempo. Su principal propósito es describir las variables y analizar observando su incidencia e interrelación en un determinado momento y tiempo. En otras palabras, es como capturar una fotografía de momento preciso de que algo sucede (35).

El diseño de la investigación se gráfica de la siguiente manera:



Dónde: M = Muestra y O = Observación

#### 4.4.Universo y muestra

4.4.1. Universo: Es el conjunto de todos los casos que concuerdan con determinadas especificaciones haciendo la totalidad de un fenómeno en estudio, incluye la totalidad de unidades de análisis o entidades de población que integran dicho fenómeno y que debe cuantificarse para dicho estudio. Además, debe ubicarse claramente en el entorno según sus características de contenido, lugar y el tiempo (35).

Para la presente tesis de investigación se ha delimitado el universo en una cantidad de 210 alumnos del quinto año de educación secundaria.

4.4.2. Muestra: Es una sub agrupación o subconjunto de elementos que integran el conjunto definido como universo, pocas veces es posible medir a todo el universo, por lo se selecciona una muestra con la cual se pretende que sea el reflejo fiel del conjunto. Se categorizan en dos ramas que son las muestras probabilísticas y no probabilísticas (35).

Para la presente tesis de investigación, la muestra ha sido seleccionada utilizando el método de muestreo no probabilístico y fue por conveniencia la cual está constituida por 43 alumnos del quinto año de educación secundaria.

#### 4.5. Definición operacional de las variables en estudio

Tabla Nro. 1: Matriz de Operacionalización de la Variable

Variable	Definición Conceptual	Dimensiones	Indicadores	Escala de Medición	Definición Operacional
Minería de Datos para determinar patrones de comportamiento de la deserción escolar	Minería de Datos: Proceso de análisis y búsqueda en bases de datos para ubicar y lograr clasificar información útil. El punto fundamental es que los patrones identificados aporten en la mejorar de toma de decisiones (22).	Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar	<ul style="list-style-type: none"> <li>- Detección casos de riesgo.</li> <li>- Importancia del estudio.</li> <li>- Familiares desertores.</li> <li>- Amistades peligrosas.</li> <li>- Situación económica.</li> <li>- Composición familiar.</li> <li>- Incremento de delincuencia.</li> <li>- Embarazos a temprana edad.</li> <li>- Acoso sexual.</li> <li>- Prevención de bullying.</li> </ul>	Ordinal	<ul style="list-style-type: none"> <li>- SI</li> <li>- NO</li> </ul>



	<p>Deserción Escolar es un concepto que utilizado generalmente para referirse a aquellos estudiantes que abandonan o dejan de asistir a clase quedando de manera momentánea fuera del sistema educativo (36).</p>	<p>Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento.</p>	<ul style="list-style-type: none"> <li>- Implementación de las TIC.</li> <li>- Inteligencia Artificial.</li> <li>- Sistemas Expertos.</li> <li>- Base de datos.</li> <li>- Mejora detección de patrones.</li> <li>- Información oportunamente.</li> <li>- Patrones de comportamiento.</li> <li>- Reportes automatizados</li> <li>- Calidad de la información.</li> <li>- Anticipación a posibles casos.</li> </ul>		
--	---	---	--	--	--

Fuente: Elaboración Propia.

#### 4.6. Técnicas e instrumentos de recolección de datos.

##### 4.6.1. Técnica

Encuesta: Es el proceso en su conjunto, desde el establecimiento del cuestionario en función de los objetivos de investigación hasta la codificación de las respuestas obtenidas a partir de la muestra, donde el cuestionario es la herramienta específicamente diseñada para la administración de las preguntas, pudiendo estar organizado o no en escalas o índices, que a veces se extraen de la reproducción de ítems provenientes de test estandarizados y baremados (35).

##### 4.6.2. Instrumentos

Cuestionario: Consiste en un conjunto de preguntas respecto de una o más variables a medir, el instrumento estandarizado que se utiliza para el recojo de datos durante el trabajo de campo de algunas investigaciones cuantitativas, fundamentalmente, las que se llevan a cabo con metodologías de encuestas. En pocas palabras, es la herramienta que permite al científico social plantear un conjunto de preguntas para recoger información estructurada sobre una muestra de personas (35).

#### 4.7. Plan de análisis

Una vez que se obtuvieron los datos recopilados se empieza con la tabulación de los resultados de cada pregunta en el programa Microsoft Excel versión 2016, realizando el análisis de datos por cada una de las preguntas establecidas por cada dimensión, permitiendo así resumir los datos en gráficos que muestra el porcentaje de impacto de las mismas.

#### 4.8. Matriz de consistencia

Tabla Nro. 2: Matriz de Consistencia

Problema	Objetivo general	Hipótesis general	Variables	Metodología
¿De qué manera la propuesta de implementación de Minería de Datos para el Colegio Nacional de la Libertad en la Ciudad de Huaraz permitirá determinar patrones de comportamiento de la deserción escolar?	Proponer la implementación de Minería de Datos para el Colegio Nacional de la Libertad en la ciudad de Huaraz, con la finalidad de determinar patrones de comportamiento de la deserción escolar.	La propuesta de implementación de Minería de Datos permite determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – Huaraz; 2020.	Minería de Datos.	Tipo: Descriptiva Nivel: Cuantitativa Diseño: No Experimental y de Corte Transversal.
	Objetivos específicos	Hipótesis específicas		
	1. Recolectar información de acuerdo a los requerimientos que ayuden a determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad.	1. La recolección de información de acuerdo a los requerimientos, ayuda a determinar patrones de comportamiento de la deserción escolar en el		

	<p>2. Utilizar la metodología adecuada, para el análisis y diseño de la solución de minería de datos que permita cumplir con los requerimientos establecidos en el Colegio Nacional de la Libertad.</p> <p>3. Diseñar la implementación de Minería de Datos, que permita la determinación de patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad</p>	<p>Colegio Nacional de la Libertad.</p> <p>2. La utilización de la metodología adecuada, para el análisis y diseño de la solución de minería de datos, permite cumplir con los requerimientos establecidos en el Colegio Nacional de la Libertad.</p> <p>3. El diseño de la implementación de Minería de Datos, permite la determinación de patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad</p>		
--	---	--	--	--

Fuente: Elaboración propia.

#### 4.9.Principios éticos

Durante el desarrollo del presente trabajo de investigación denominada “Propuesta de Implementación de Minería de Datos para el Colegio Nacional de la Libertad – Huaraz; 2020”. Se ha considerado en forma estricta el cumplimiento de los principios éticos de la Universidad Católica Los Ángeles de Chimbote (ULADECH), que permitan asegurar la originalidad de la Investigación y son los siguientes (37):

- Protección a las personas. La persona en toda investigación es el fin y no el medio, por ello necesitan cierto grado de protección, el cual se determinará de acuerdo al riesgo en que incurran y la probabilidad de que obtengan un beneficio, a continuación, detallo (37).
- Beneficencia y no maleficencia. Se debe asegurar el bienestar de las personas que participan en las investigaciones. En ese sentido, la conducta del investigador debe responder a las siguientes reglas generales: no causar daño, disminuir los posibles efectos adversos y maximizar los beneficios (37).
- Justicia. El investigador debe ejercer un juicio razonable, ponderable y tomar las precauciones necesarias para asegurarse de que sus sesgos, y las limitaciones de sus capacidades y conocimiento, no den lugar o toleren prácticas injustas. Se reconoce que la equidad y la justicia otorgan a todas las personas que participan en la investigación derecho a acceder a sus resultados. El investigador está también obligado a tratar equitativamente a quienes participan en los procesos, procedimientos y servicios asociados a la investigación (37).

- Integridad científica. La integridad o rectitud deben regir no sólo la actividad científica de un investigador, sino que debe extenderse a sus actividades de enseñanza y a su ejercicio profesional. La integridad del investigador resulta especialmente relevante cuando, en función de las normas deontológicas de su profesión, se evalúan y declaran daños, riesgos y beneficios potenciales que puedan afectar a quienes participan en una investigación (37).
  
- Consentimiento informado y expreso. En toda investigación se debe contar con la manifestación de voluntad, informada, libre, inequívoca y específica; mediante la cual las personas como sujetos investigadores o titular de los datos consienten el uso de la información para los fines específicos establecidos en el trabajo de investigación (37).

## V. RESULTADOS

### 5.1. Resultados de la encuesta a los alumnos

#### 5.1.1. Resultados de la dimensión 1: Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar.

Tabla Nro. 3: Técnicas de detección de casos de riesgo

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de los métodos de detección actual de deserción escolar, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	10	23,26
No	33	76,74
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Conoces algún compañero que tenga la intención de dejar sus estudios escolares y este siendo apoyado por el colegio para cambiar de opinión?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 3, se observa que, el 76,74% de los encuestados manifestaron que, No se ha implementado métodos de detección de posibles casos de deserción escolar en la I.E., mientras que, el 23,26% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 4: Importancia del estudio escolar.

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la importancia del estudio escolares, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	35	81,40
No	8	18,60
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Tus docentes te repiten constantemente sobre la importancia de terminar tus estudios escolares dentro de la edad apropiada?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 4, se observa que, el 81,40% de los encuestados manifestaron que, Sí es constante el dialogo sobre la importancia de los estudios escolares en la I.E., mientras que, el 18,60% de los encuestados manifestaron todo lo contrario.



Tabla Nro. 5: Familiares que desertaron en sus estudios.

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de contar con familiares que desertaron en sus estudios, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	25	58,14
No	18	41,86
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿En algún momento tu colegio averiguo si tú o tus compañeros tienen familiares que no hayan culminado sus estudios escolares?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 5, se observa que, el 58,14% de los encuestados manifestaron que, Sí han identificado y consolidado información sobre alumnos con familiares que desertaron en sus estudios en la I.E., mientras que, el 41,86% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 6: Círculo de amistades peligrosas

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la identificación de amistades peligrosas de los alumnos de manera interna o externa a la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	17	39,53
No	26	60,47
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Tus docentes han llegado a conocer en algún momento a tus amistades del colegio y de tu entorno social fuera del colegio?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 6, se observa que, el 60,47% de los encuestados manifestaron que, No se identifica de manera constante las amistades peligrosas de los alumnos en la I.E., mientras que, el 39,53% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 7: Referencias sobre la situación económica

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de contar con referencias relacionada a la situación económica de los alumnos en la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	23	53,49
No	20	46,51
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿El colegio sabes si tú o quizás tus compañeros estudian y trabajan para aportar con ingresos económicos para los gastos tu hogar?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 7, se observa que, el 53,49% de los encuestados manifestaron que, Sí se cuenta con referencias respecto a la situación económica de los alumnos en la I.E., mientras que, el 46,51% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 8: Identificación de la composición familiar

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la identificación de las familias integradas o desintegradas de los alumnos en la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	34	79,07
No	9	20,93
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Tus docentes conocen con quien vives, si es solo con tu papá, solo con tu mamá, con tu papá y tu mamá o con algún otro familiar que no sean tus padres?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 8, se observa que, el 79,07% de los encuestados manifestaron que, Sí se ha identificado las familias integradas o desintegradas de los alumnos en la I.E., mientras que, el 20,93% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 9: Incremento de la delincuencia

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la consolidación de la información sobre el incremento de la delincuencia en la ciudad, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	27	62,79
No	16	37,21
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿En la actualidad el colegio brinda información preventiva ante el incremento de la delincuencia alrededor de la Institución educativa o en la ciudad?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 9, se observa que, el 62,79% de los encuestados manifestaron que, Sí se ha consolidado información sobre el incremento de la delincuencia en la I.E., mientras que, el 37,21% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 10: Prevención de embarazos a temprana edad

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la implementación de programas en favor de la prevención de embarazos a temprana edad, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	36	83,72
No	7	16,28
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Has recibido en tu colegio alguna charla u orientación sobre la prevención de embarazos en la etapa escolar?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 10, se observa que, el 83,72% de los encuestados manifestaron que, Sí se ha implementado programas en favor de la prevención de embarazos a temprana edad en la I.E., mientras que, el 16,28% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 11: Detección temprana de acoso sexual

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la detección temprana de acoso sexual en todas las formas, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	36	83,72
No	7	16,28
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Has recibido en tu colegio alguna charla u orientación relacionada a la prevención de acoso sexual dentro o fuera de la institución?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 11, se observa que, el 83,72% de los encuestados manifestaron que, Sí se ha implementado métodos o técnicas en favor de la detección temprana del acoso sexual en la I.E., mientras que, el 16,28% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 12: Detección temprana de prevención del bullying

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de los métodos o técnicas en favor de la detección temprana del bullying, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	17	39,53
No	26	60,47
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Conoces algún caso de maltrato físico o psicológico constante a uno o más compañeros, donde el colegio haya intervenido para solucionarlo?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 12, se observa que, el 60,47% de los encuestados manifestaron que, No se ha implementado métodos o técnicas en favor de la detección temprana del bullying en la I.E., mientras que, el 39,53% de los encuestados manifestaron todo lo contrario.



5.1.2. Resultados de la dimensión 2: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento.

Tabla Nro. 13: Necesidad de implementación de las TIC

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la necesidad de implementar nuevas tecnologías de la información y comunicación (TIC) en la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	42	97,67
No	1	2,33
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Te gustaría que tu colegio cuenta con nuevas tecnologías que permitan mejorar cada día en la atención a los alumnos?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 13, se observa que, el 97,67% de los encuestados manifestaron que, Sí es necesario la implementación de nuevas técnicas de la información y comunicación (TIC) en la I.E., mientras que, el 2,33% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 14: Tecnologías basadas en inteligencia artificial.

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de implementar aplicaciones tecnológicas basadas en inteligencia artificial en la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	43	100,00
No	-	-
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Estarías de acuerdo con implementar de nuevas aplicaciones tecnológicas inteligentes que apoyen a tu colegio?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 14, se observa que, el 100,00% de los encuestados manifestaron que, Sí está de acuerdo con la implementación de aplicaciones tecnológicas basadas en inteligencia artificial en la I.E.

Tabla Nro. 15: Modelo de Sistemas Expertos

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la elaboración de un modelo de sistemas expertos basado en el comportamiento de la información el cual apoyaría en la detección temprana de posibles casos de deserción escolar en la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	42	97,67
No	1	2,33
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Te gustaría que exista un sistema informático que sea capaz de apoyar a tus docentes, detectando de manera anticipada algún posible caso de abandono de estudios por parte de tus compañeros?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 15, se observa que, el 97,67% de los encuestados manifestaron que la elaboración de un modelo de sistemas expertos basado en el comportamiento de la información, Sí apoyaría en la detección temprana de posibles casos de deserción escolar en la I.E., mientras que, el 2,33% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 16: Creación de bases de datos

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la implementación de métodos de recolección o generación de datos que se almacenen en un ordenador, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	41	95,35
No	2	4,65
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Tu colegio almacena información en computadoras de datos personales, registros de calificaciones u otra información de los alumnos o docentes?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 16, se observa que, el 95,35% de los encuestados manifestaron que, Sí se ha implementado métodos de recolección o generación de datos que han sido almacenados en un ordenador en la I.E., mientras que, el 4,65% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 17: Automatización de la detección de casos de riesgo

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca del mejoramiento automatizado de detección temprana de los posibles casos de deserción escolar en la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	42	97,67
No	1	2,33
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Estarías de acuerdo de que tu colegio cuente con la posibilidad de influir en cambiar de decisión ante algún indicio detectado en tu persona?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 17, se observa que, el 97,67% de los encuestados manifestaron que, Sí están a favor del mejoramiento automatizado de detección temprana de los posibles casos de deserción escolar en la I.E., mientras que, el 2,33% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 18: Información oportuna para toma de decisiones

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la necesidad de automatizar y disponer de información de manera oportuna para la toma de decisiones, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	43	100,00
No	-	-
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Crees que es importante que tu colegio cuente con información automatizada que le permita tomar mejores decisiones en un corto plazo?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 18, se observa que, el 100,00% de los encuestados manifestaron que, Sí es necesaria la automatización y disponibilidad de información de manera oportuna para la toma de decisiones en la I.E.

Tabla Nro. 19: Anticipación ante patrones de comportamiento

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de que si sería factible contar con la posibilidad de anticiparse ante el comportamiento de posibles casos de deserción escolar en la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	42	97,67
No	1	2,33
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Estarías de acuerdo que tu colegio se anticipe detectando y apoyando a algún posible caso que se presente de abandono de estudios por parte de los alumnos?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 19, se observa que, el 97,67% de los encuestados manifestaron que, Sí sería factible contar con la posibilidad de anticiparse ante el comportamiento de posibles casos de deserción escolar en la I.E., mientras que, el 2,33% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 20: Sistemas Expertos

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de la necesidad de contar con la disponibilidad de reportes e informes automatizados para la toma de decisiones en la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	42	97,67
No	1	2,33
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Consideras que sería bueno que tu colegio tenga la posibilidad de contar con información inmediata sin mucha demora para tomar alguna decisión importante?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 20, se observa que, el 97,67% de los encuestados manifestaron que, Sí es necesario de contar con la disponibilidad de reportes e informes automatizados para la toma de decisiones en la I.E., mientras que, el 2,33% de los encuestados manifestaron todo lo contrario.



Tabla Nro. 21: Confiabilidad y calidad de la información.

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de que la I.E. debe contar con información confiable y de calidad almacenada en un ordenador, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	42	97,67
No	1	2,33
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Para ti es importante que tu colegio cuente con información confiable y de calidad que se encuentre clasificada y almacenada en una computadora?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 21, se observa que, el 97,67% de los encuestados manifestaron que, la Institución Educativa Sí debe contar con información confiable y de calidad almacenada en un ordenador, mientras que, el 2,33% de los encuestados manifestaron todo lo contrario.

Tabla Nro. 22: Anticipación ante casos de riesgo

Frecuencias y respuestas distribuidas de los alumnos encuestados, acerca de que si estaría de acuerdo con la implementación de un modelo de comportamiento de información para la intervención oportuna ante posibles casos de deserción escolar en la I.E., respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	43	100,00
No	-	-
Total	43	100,00

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para responder a la siguiente pregunta: ¿Es importante que tu colegio cuente con un sistema que le permita analizar comportamientos de información ingresada, para la intervención oportuna de posibles casos de abandono de estudios?

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 22, se observa que, el 100,00% de los encuestados manifestaron que, Sí están de acuerdo con la implementación de un modelo de comportamiento de información para la intervención oportuna ante posibles casos de deserción escolar.

### 5.1.3. Resultados por dimensión

#### 5.1.3.1. Resultado general de la dimensión 1

Tabla Nro. 23: Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar.

Frecuencias y respuestas distribuidas de los estudiantes encuestados, acerca de la dimensión 1, en donde se aprueba el método actual para la detección de posibles casos de abandono escolar, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

Alternativas	n	%
Si	35	81,40
No	8	18,60
Total	43	100,00

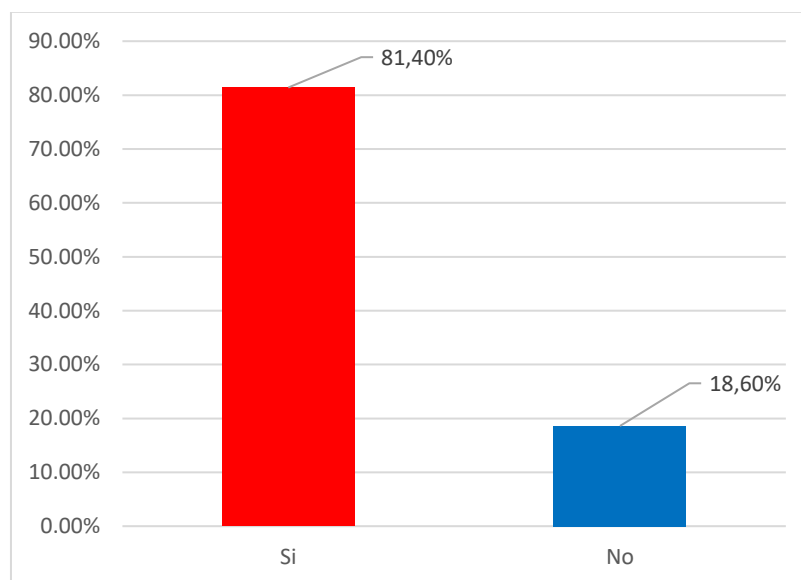
Fuente: Instrumento de recolección de datos para medir la dimensión 1: Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar, basado en 10 preguntas, aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz.

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 23, se observa que, el 81,40% de los encuestados manifestaron que, Sí están satisfechos con el método actual de detección de posibles casos de abandono

escolar, mientras que, el 18,60% de los encuestados manifestaron que, No están satisfechos con el método actual de detección de posibles casos de abandono escolar.

Gráfico Nro. 6: Resultado general de la dimensión 1



Fuente: Tabla Nro. 23: Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar.

### 5.1.3.2.Resultado general de la dimensión 2

Tabla Nro. 24: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento.

Frecuencias y respuestas distribuidas de los estudiantes encuestados, acerca de la dimensión 2, en donde se aprueba la implementación de un sistema experto, respecto a la propuesta de implementación de minería de datos para determinar patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad – HUARAZ; 2020.

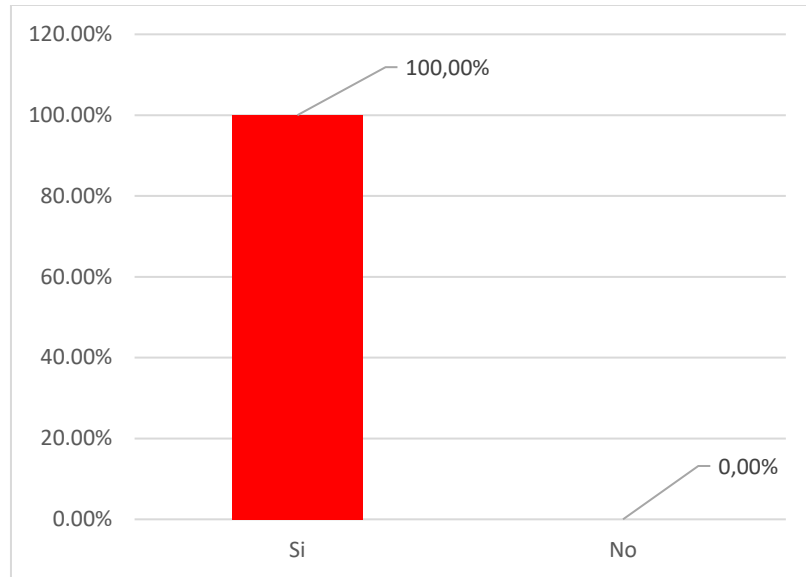
Alternativas	n	%
Si	43	100,00
No	-	-
Total	43	100,00

Fuente: Instrumento de recolección de datos para medir la dimensión 1: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento, basado en 10 preguntas, aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz.

Aplicado por: Rondo, E.; 2020.

En la Tabla Nro. 24, se observa que, el 100,00% de los encuestados manifestaron que, Sí están de acuerdo con la implementación de un sistema experto que permita determinación de patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad.

Gráfico Nro. 7: Resultado general de la dimensión 2



Fuente: Tabla Nro. 24: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento.

#### 5.1.4. Resumen general

Tabla Nro. 25: Resumen general de dimensiones

Frecuencias y respuestas distribuidas, para determinar los niveles correspondientes a la dimensión 1: Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar, y la dimensión 2: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento, respecto a la implementación de Minería de Datos para el Colegio Nacional de la Libertad – Huaraz; 2020.

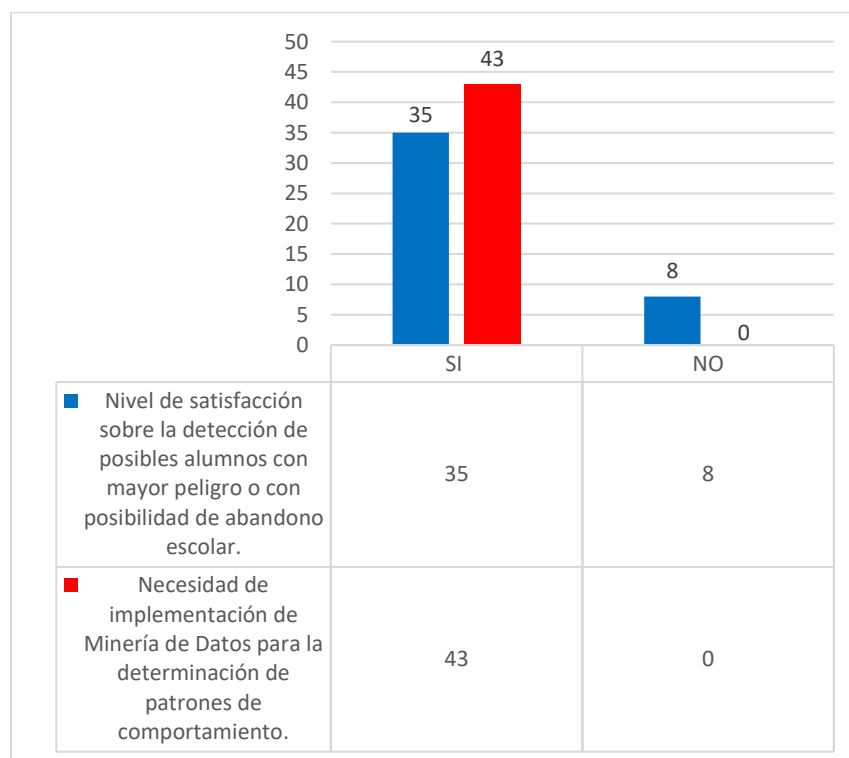
Dimensiones	Alternativas de Respuestas				Muestra	
	Si	%	No	%	n	%
Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar.	35	81,40	8	18,60	43	100
Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento.	43	100,00	-	-	43	100

Fuente: Instrumento de recolección de datos aplicado a los alumnos del Colegio Nacional de la Libertad – Huaraz, para medir la dimensión 1 y la dimensión 2, las cuales fueron definidas para esta investigación.

Aplicado por: Rondo, E.; 2020.

Una vez obtenidos los resultados, en la Tabla Nro. 25, se puede observar que, en lo que respecta a la dimensión 1: Nivel de satisfacción sobre método de detección de posibles casos de abandono escolar, el 81,40% de los encuestados manifestaron que, Sí están satisfechos con el método actual de detección de posibles casos de abandono escolar, mientras que, el 18,60% de los encuestados manifestaron que, No están satisfechos con el sistema actual, y respecto a la dimensión 2: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento, se observa que, el 100,00% de los encuestados manifestaron que, Sí están de acuerdo con la implementación de Minería de Datos que permita la determinación de patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad.

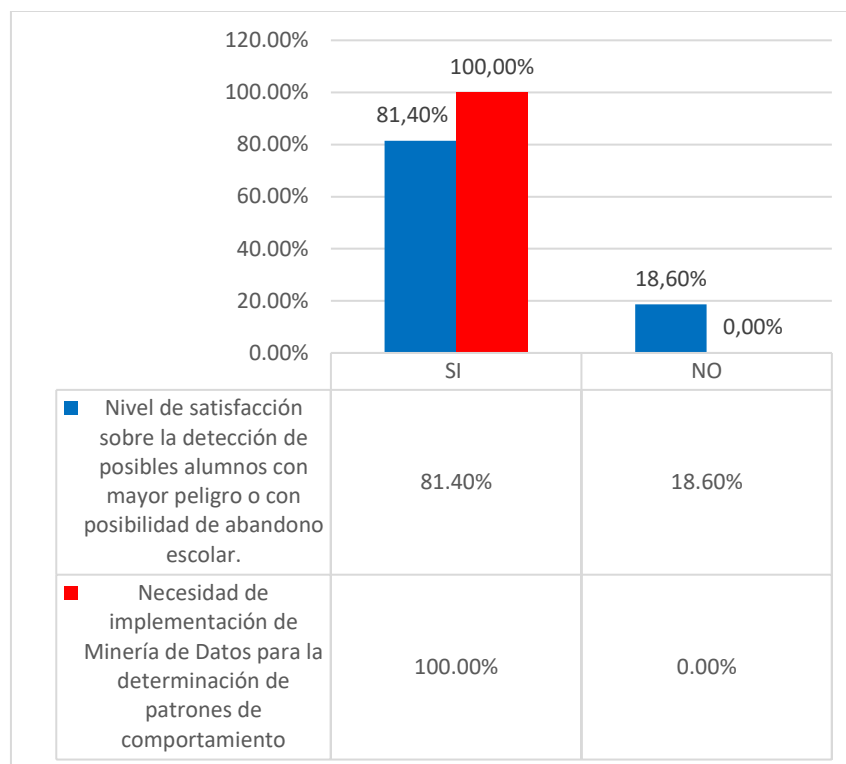
Gráfico Nro. 8: Resumen general de las dimensiones



Fuente: Tabla Nro. 25: Resumen general de dimensiones



Gráfico Nro. 9: Resumen porcentual de las dimensiones



Fuente: Tabla Nro. 25: Resumen general de dimensiones

## 5.2. Análisis de resultados

La presente investigación tuvo como objetivo general: Proponer la implementación de Minería de Datos para el Colegio Nacional de la Libertad en la ciudad de Huaraz, con la finalidad de determinar patrones de comportamiento de la deserción escolar; en el cual se obtuvieron dos dimensiones que determinan la satisfacción del proceso actual y la necesidad de implementar nuevos procesos frente a la detección de posibles casos riesgo, en consecuencia una vez realizada la aplicación del instrumento que permita conocer la percepción de los encuestados y la interpretación de todos los resultados se proceden a analizarlos detenidamente en los siguientes párrafos:

Referente a la primera dimensión: Nivel de satisfacción sobre el proceso actual de detección de posibles casos de riesgo en la Tabla Nro. 4 se observa que el 81.40% de los encuestados manifestaron que Si se dialoga constantemente con los estudiantes sobre la importancia los estudios escolares; mientras el 18.60% expresaron lo contrario. Estos resultados tienen una similitud con los resultados obtenidos por Piscoya L. (10), en su trabajo de investigación titulada “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la Región De Lambayeque”, quien indica que se realizó la selección de las técnicas predictivas de Minería de Datos, determinando que el modelo a utilizar sería uno de series de tiempo y redes neuronales, dada la naturaleza de los datos analizados en el datawarehouse, se realizó un breve análisis de las técnicas u algoritmos que intervenían en este tipo de modelo como se muestra en la tabla 13 del trabajo de investigación citado en los antecedentes a nivel nacional; así mismo coinciden con la investigación realizada por Gálvez M. y Flores K. (11), en sus tesis titulada “Modelo predictivo de deserción universitaria de la carrera de Ingeniería Informática”, de la

Universidad Ricardo Palma de Lima, quienes en su investigación y para un cuestionario semejante obtuvieron un alto indicador de ausencia y supervisión por parte de sus padres. Así mismo teóricamente coincide con el autor Marqués M. (21), quien opina que una base de datos puede servir como un gran almacén de datos, que son utilizados al mismo tiempo por varios usuarios que es diseñado para satisfacer los requisitos de información actualizada de manera permanente y continua de una manera anticipada de los posibles casos de deserción escolar que se pueden presentar mediante comportamientos detectados. Además, no sólo contiene los datos de la empresa u organización, también almacena una descripción de dichos datos, esta descripción es lo que se denomina metadatos, que se almacena en el diccionario de datos o catálogo, se concluye que existe un porcentaje favorable de satisfacción del proceso de identificación de posibles casos de riesgo de deserciones mediante el análisis del comportamiento de los alumnos.

Referente a la segunda dimensión: Necesidad de implementación de Minería de Datos para la determinación de patrones de comportamiento; en la Tabla Nro. 15, se observa que, el 97,67% de los encuestados manifestaron que, Sí sería factible contar con la posibilidad de anticiparse ante el comportamiento de posibles casos de deserción escolar en la I.E., mientras que, el 2,33% de los encuestados manifestaron todo lo contrario; Estos datos obtenidos tienen similitud con los resultados obtenidos en la investigación de Córdova J. (8), en su tesis denominado: “Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes que pertenecen al Colegio Fisco-Misional “San Francisco” de la Ciudad de Ibarra”, también tiene una similitud con la investigación realizada por Torres L. (13), quien realizó una investigación titulada “Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz – 2016”, ambos autores

quienes respectivamente en sus trabajos de investigación y para un cuestionario semejante, obtuvieron un alto porcentaje a favor de contar con la posibilidad de anticiparse ante el comportamiento de posibles casos de deserción escolar con el apoyo de análisis de información para la toma de decisiones. Teóricamente estos resultados también se asemejan con el estudio realizado por el autor López J. (22) menciona que la minería de datos es un proceso de análisis y búsqueda de gran capacidad (Big Data) para clasificar y ubicar información útil que pueda servir para la toma de decisiones, también denominado como la tecnología y software empleado para identificar patrones de comportamiento dentro de la base de datos almacenada, se concluye que existe la necesidad en su totalidad de implementar de minería de datos que permita identificar patrones de los posibles casos de riesgo de deserción mediante el comportamiento de la información de los alumnos de manera automatizada simplificando los procesos.

### 5.3. Propuesta de mejora

Se propone el uso de las tecnologías de la información y comunicación con la implementación de Minería de Datos basado en la ciencia de los algoritmos, que permita determinar los posibles casos de riesgo de deserción escolar; para ello, luego de realizar una comparación de metodologías de Minería de Datos, resulta que el Descubrimiento de Conocimiento en Bases de Datos (Knowledge Discovery in Databases o KDD) es la metodología más adecuada para llevar a cabo la implementación de manera ordenada y sistematizada del modelo predictivo que se pretende.

### 5.3.1. Propuesta técnica

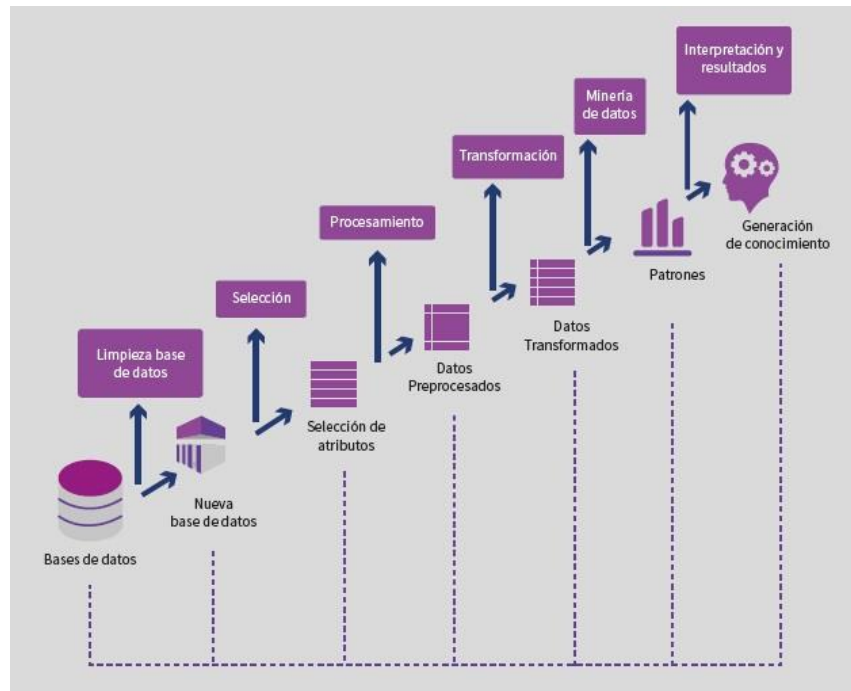
#### Modelo Predictivo de Deserción Escolar Basado en Árboles de Decisión y Minería de Datos

Esta propuesta de mejora muestra la construcción de un modelo predictivo de deserción escolar, para pronosticar la probabilidad de que un estudiante abandone sus estudios escolares, mediante técnicas de clasificación, basadas en árboles de decisión. Con la aplicación de la metodología Knowledge Discovery in Databases – KDD, se logró construir un árbol con cuatro niveles de profundidad y reglas, que evalúan a los posibles casos de deserción. Llegando a concluir que las variables nivel y notas tienen mayor influencia en la deserción.

**Objetivo:** Crear un modelo predictivo de deserción escolar, para determinar la probabilidad, que un estudiante abandone el colegio, teniendo en cuenta, el rendimiento académico y variables de su entorno personal, en el presente trabajo de investigación se describe la metodología utilizada para la generación del modelo predictivo, los resultados de la creación del modelo y su aplicación.

**Metodología:** La metodología se basa en el Descubrimiento de Conocimiento en Bases de Datos (Knowledge Discovery in Database – KDD), dividido en cinco principales etapas que son: Selección, Procesamiento, Transformación, Minería de Datos y Evaluación, aplicando el algoritmo de Árbol de clasificación y regresión (Classification and Regression Tree – CART) de la herramienta R - Árboles de regresión y particionamiento recursivos (Recursive Partitioning And Regression Trees – RPART), logrando construir un árbol con cuatro niveles de profundidad y reglas.

Grafico Nro. 10: Proceso de la Metodología KDD.

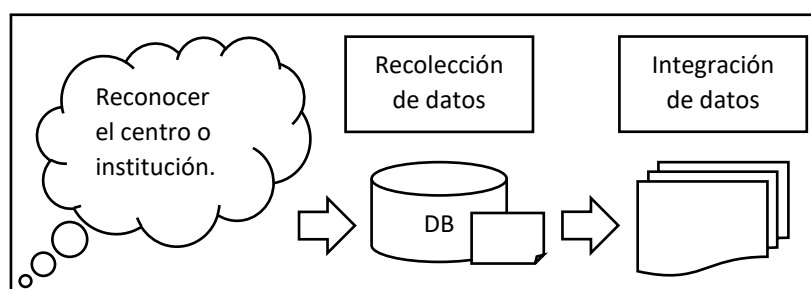


Fuentes: Universidad de Rosario (38).

Aplicación de las Cinco Etapas de la Metodología:

### 1. Selección de datos

Se procede a recolectar la información necesaria abarcando diferentes criterios de manera conveniente y utilizando diferentes métodos de recolección de datos, los cuales son almacenados en un motor de base de datos apropiado. Para el presente trabajo de investigación se utilizará una simulación de base de datos.



Fuente: Elaboración propia.

Los datos simulados corresponden a un grupo de estudiantes del quinto año de educación secundaria de Colegio Nacional de la Libertad y se encuentran clasificados en dos matrices:

Matriz de datos generales, contiene los datos personales con los atributos: identificador, con identificador numérico del 1 al 20, para mantener la confidencialidad de los participantes, además, género, fecha de nacimiento, lugar nacimiento, dirección de residencia.

Grafico Nro. 11: Matriz de datos generales

	A	B	C	D	E
1	<b>ID</b>	<b>Genero</b>	<b>FechaNac</b>	<b>LugarNac</b>	<b>Dirección</b>
2	0001	Masculino	08/04/2002	Huaraz	Independencia
3	0002	Masculino	13/02/2002	Huaraz	Independencia
4	0003	Masculino	08/06/2003	Yungay	Huaraz
5	0004	Femenino	03/02/2004	Huaraz	Independencia
6	0005	Masculino	24/11/2004	Carhuaz	Huaraz
7	0006	Femenino	22/06/2004	Yungay	Independencia
8	0007	Femenino	24/12/2003	Huaraz	Huaraz
9	0008	Femenino	21/05/2004	Caraz	Independencia
10	0009	Femenino	05/05/2004	Carhuaz	Independencia
11	0010	Femenino	10/11/2004	Huaraz	Independencia
12	0011	Masculino	23/07/2004	Recuay	Huaraz
13	0012	Femenino	10/10/2003	Caraz	Huaraz
14	0013	Femenino	17/02/2003	Recuay	Huaraz
15	0014	Femenino	04/04/2004	Huaraz	Independencia
16	0015	Femenino	06/06/2004	Recuay	Huaraz
17	0016	Masculino	10/05/2002	Huaraz	Independencia
18	0017	Femenino	03/05/2004	Caraz	Huaraz
19	0018	Femenino	09/05/2004	Recuay	Huaraz
20	0019	Masculino	02/07/2004	Caraz	Huaraz
21	0020	Masculino	19/07/2004	Huaraz	Independencia

Fuente: Elaboración propia.

Matriz de datos Académicos, contiene datos académicos con los atributos: identificador, periodo académico, grado o año en curso, cursos, promedio de notas obtenidas en el primer y segundo semestre.

Grafico Nro. 12: Matriz de datos Académicos

	A	B	C	D	E	F	G	H
1	ID	Periodo	Grado/Año	Curso	1° Sem	2° Sem	Promedio	Condición
2	0001	2020	Quinto	Raz. Matemático	14	08	11	Aprobado
3	0001	2020	Quinto	Raz. Verbal	15	14	15	Aprobado
4	0001	2020	Quinto	Comunicación	09	15	12	Aprobado
5	0001	2020	Quinto	Literatura	13	16	15	Aprobado
6	0001	2020	Quinto	Química	13	11	12	Aprobado
7	0001	2020	Quinto	Física	14	16	15	Aprobado
8	0001	2020	Quinto	Historia	15	08	12	Aprobado
9	0001	2020	Quinto	Matemática	10	08	09	Desaprobado
10	0001	2020	Quinto	Educ. Artística	08	19	14	Aprobado
11	0001	2020	Quinto	Educ. Física	08	09	09	Desaprobado
12	0002	2020	Quinto	Raz. Matemático	19	15	17	Aprobado
13	0002	2020	Quinto	Raz. Verbal	15	17	16	Aprobado
14	0002	2020	Quinto	Comunicación	15	18	17	Aprobado
15	0002	2020	Quinto	Literatura	13	11	12	Aprobado
16	0002	2020	Quinto	Química	09	18	14	Aprobado
17	0002	2020	Quinto	Física	09	15	12	Aprobado
18	0002	2020	Quinto	Historia	17	19	18	Aprobado
19	0002	2020	Quinto	Matemática	08	10	09	Desaprobado
20	0002	2020	Quinto	Educ. Artística	08	16	12	Aprobado
21	0002	2020	Quinto	Educ. Física	09	10	10	Desaprobado

Fuente: Elaboración propia.

Matriz de integración, inicialmente se contó con 200 registros para el análisis, los atributos de la Matriz de datos generales y académicos, fueron integrados en una hoja de cálculo, además se agregó los atributos Edad, Promedio de Notas del primer y segundo bimestre con prefijos que representan los curso (RM=Razonamiento Matemático, RV=Razonamiento Verbal, CO=Comunicación, LI=Literatura, QU=Química, FI=Física, HI=Historia, MA=Matemática, EA=Educación Artística, EF=Educación Física) Promedio Global y Repitente.



Grafico Nro. 13: Matriz de integración.

	A	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	
1	ID	Edad	RM		RV		CO		LI		QU		FI		HI		MA		EA		EF		Pro.	Repi	
2			P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	P1	P2	Gen.	tente	
3	0001	17	14	08	15	14	09	15	08	09	07	08	08	06	15	08	10	08	08	19	08	09	10	14	SI
4	0002	18	19	15	15	17	15	18	13	11	09	18	09	15	17	19	08	10	08	16	09	10	14	SI	
5	0003	16	10	19	16	10	19	11	11	18	09	14	18	15	08	16	15	16	12	10	14	12	14	NO	
6	0004	16	16	11	15	11	18	15	12	18	11	10	11	12	18	13	09	17	12	14	17	17	14	NO	
7	0005	15	13	13	10	13	19	19	14	16	18	11	11	12	11	11	15	08	09	18	15	08	13	NO	
8	0006	15	09	09	09	16	14	16	09	15	16	10	10	19	16	15	11	17	11	17	19	08	13	NO	
9	0007	16	16	13	10	19	14	09	14	14	18	10	13	14	17	08	11	17	12	16	08	09	13	NO	
10	0008	15	08	09	10	09	09	19	18	14	11	17	16	17	15	16	11	19	09	17	17	15	14	NO	
11	0009	15	19	17	15	14	13	15	14	17	11	13	11	15	18	16	19	11	09	09	18	15	14	NO	
12	0010	15	11	14	16	11	09	09	12	09	11	16	09	16	17	19	19	10	17	15	17	11	13	NO	
13	0011	15	14	10	11	08	19	16	12	17	08	17	18	18	13	16	18	17	15	14	18	12	15	NO	
14	0012	16	08	13	10	12	17	18	12	09	17	10	15	18	08	12	17	08	14	17	15	11	13	NO	
15	0013	17	09	18	18	18	11	10	15	18	16	11	11	11	11	10	12	15	17	18	08	12	13	SI	
16	0014	15	09	17	11	14	15	19	17	16	19	13	16	14	14	08	14	10	19	14	10	16	14	NO	
17	0015	15	18	11	08	08	19	13	16	11	15	19	08	14	17	16	15	08	09	17	11	12	13	NO	
18	0016	17	12	08	19	10	17	10	15	13	19	13	09	08	13	19	09	12	11	10	15	15	13	SI	
19	0017	15	09	10	12	12	08	12	10	12	09	12	08	10	09	12	09	09	11	08	13	13	10	NO	
20	0018	15	10	17	10	19	19	08	18	15	10	18	09	14	09	12	09	08	18	16	13	09	13	NO	
21	0019	15	10	13	19	16	17	11	08	15	18	15	13	18	10	15	10	11	13	18	09	10	13	NO	
22	0020	15	13	13	17	17	18	12	18	12	14	08	09	10	14	10	14	10	10	15	19	11	13	NO	
--																									

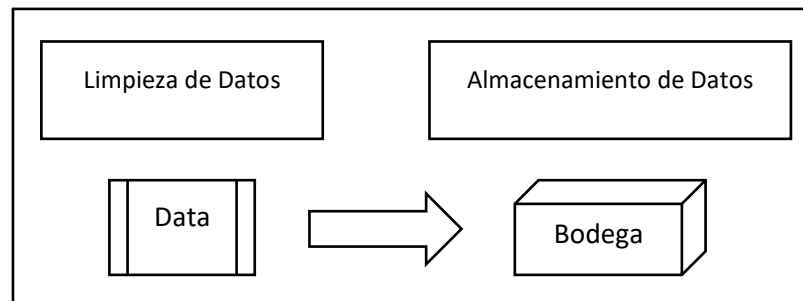
Fuente: Elaboración propia.

Los atributos se determinaron de la siguiente manera:

- Edad, es la Fecha de nacimiento menos la fecha de inicio del presente periodo 2020.
- P1 y P2, es el promedio de notas por curso en el primer y segundo bimestre respectivamente.
- Promedio general, es el promedio de promedios teniendo en cuenta el peso porcentual de cada uno.
- Repitente, se determina en base a que si el estudiante se encuentra en el mismo año por segunda o más veces.

## 2. Procesamiento de datos

Por medio del uso de diferentes métodos o técnicas de validación se realiza el descubrimiento, corrección o eliminación de registros permitiendo identificar datos incompletos, incorrectos, inexactos, no pertinentes, entre otros contando con la posibilidad de convertir a un formato adecuado para su respectivo almacenamiento en un sistema de gestión de datos adecuado.



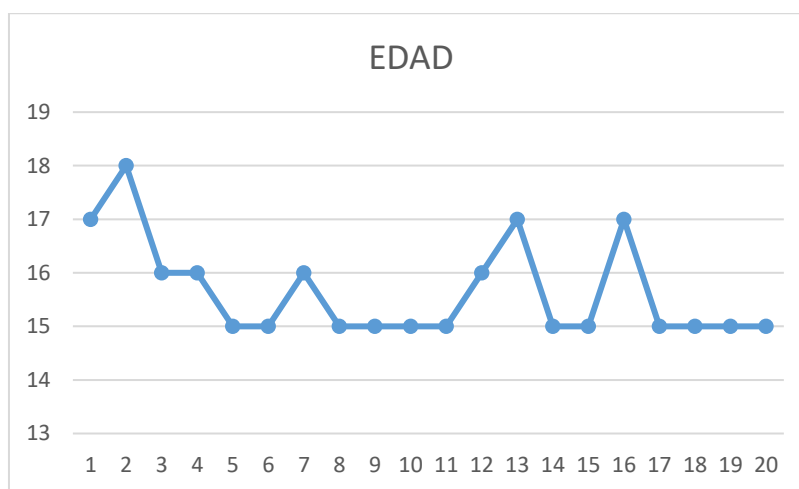
Fuente: Elaboración propia.

Durante el proceso se detecta la existencia de datos atípicos respecto a la edad apropiada para cursar el quinto año de educación secundaria, siendo cuatro los datos identificados donde un estudiante tienen 18 años y tres estudiantes tienen 17 años.

Se representa el resultado en la siguiente gráfica de la variable.

Edad = (17, 18, 16, 16, 15, 15, 16, 15, 15, 15, 15, 16, 17, 15, 15, 17, 15, 15, 15, 15).

Grafico Nro. 14: Datos atípicos del atributo edad



Fuente: Elaboración propia.

En el atributo de promedios se realiza la consulta para identificar registros en blanco los cuales son considerados atípicos por el tipo de atributo y de manera masiva se registra el valor cero según ejemplo siguiente:

Dice:

P1 = (14,15,09,,07,08,15,10,08,)

P2 = (,14,15,09,08,,,08,19,09)

Debe decir:

P1 = (14,15,09,00,07,08,15,10,08,00)

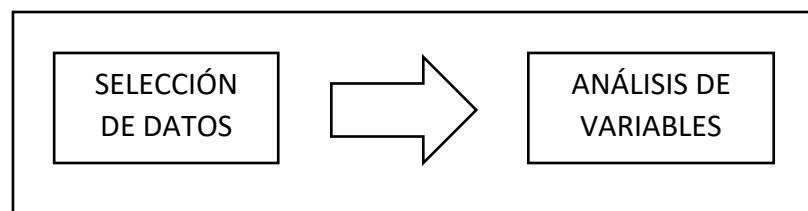
P2 = (00,14,15,09,08,00,00,08,19,09)

Además, se reemplazó, los caracteres especiales como la ñ y tildes, en el atributo nombre, lugar de nacimiento y ciudad de residencia evitando resultados erróneos, para la Bodega de

datos, se obtuvo una base de datos limpia con datos almacenados de los estudiantes, resultando tres tablas denominadas: Datos Generales con 20 registros, Datos Académicos con 200 registros y Datos Integrados con 20 registros los cuales fueron almacenados correctamente en hojas de cálculo.

### 3. Transformación de Datos

Este proceso nos permite determinar las variables de predicción al momento de realizar el almacenamiento en una base de datos relacional teniendo en cuenta que dichas variables pueden ser obtenidas mediante el análisis del entorno y pueden ser específicas o complejas según su medición.



Fuente: Elaboración propia.

Para la selección de datos, se tomó los registros almacenados en hojas de cálculo de los estudiantes los cuales son ingresados a una base de datos relacional de acuerdo al motor de base de datos elegida, mediante este proceso se obtiene las variables de predicción de forma condicional con sus respectivos atributos que se clasifican en tres tipos:

- Nominal
- Ordinal
- Cuantitativo.

Tabla Nro. 26: Variables resultantes de predicción.

<b>ATRIBUTO</b>	<b>TIPO DE ATRIBUTO</b>
IDENTIFICADOR	CUANTITATIVO
GENERO	NOMINAL
EDAD	CUANTITATIVO
LUGAR DE NACIMIENTO	NOMINAL
DIRECCIÓN ACTUAL	NOMINAL
PERIODO	ORDINAL
GRADO/AÑO	ORDINAL
CURSO	NOMINAL
PROMEDIO 1° SEMESTRE	CUANTITATIVO
PROMEDIO 2° SEMESTRE	CUANTITATIVO
CONDICIÓN	NOMINAL
REPITENTE	NOMINAL

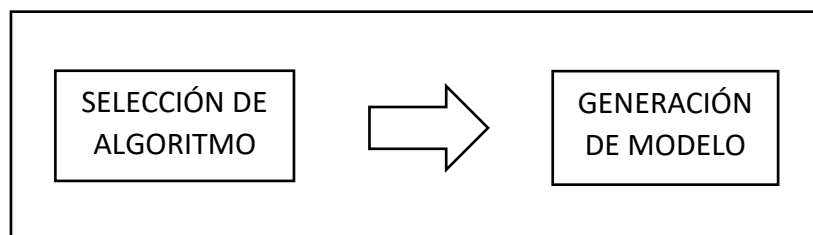
Fuente: Elaboración propia.

#### 4. Minería de datos

A continuación, se presentan las técnicas empleadas, las cuales pueden ser de tipo supervisado o no supervisado.

- Identificación: Evidenciar la existencia de objetos, eventos y actividades en el conjunto de datos (análisis factorial, discriminante, regresivo, de correlaciones).
- Clasificación: Particionar los datos de acuerdo a las clases o etiquetas asignadas al conjunto de datos (ej: tablas de decisión, reglas de decisión, clasificadores basados en casos, redes neuronales, clasificadores bayesianos y clasificadores basados en acoplamientos).

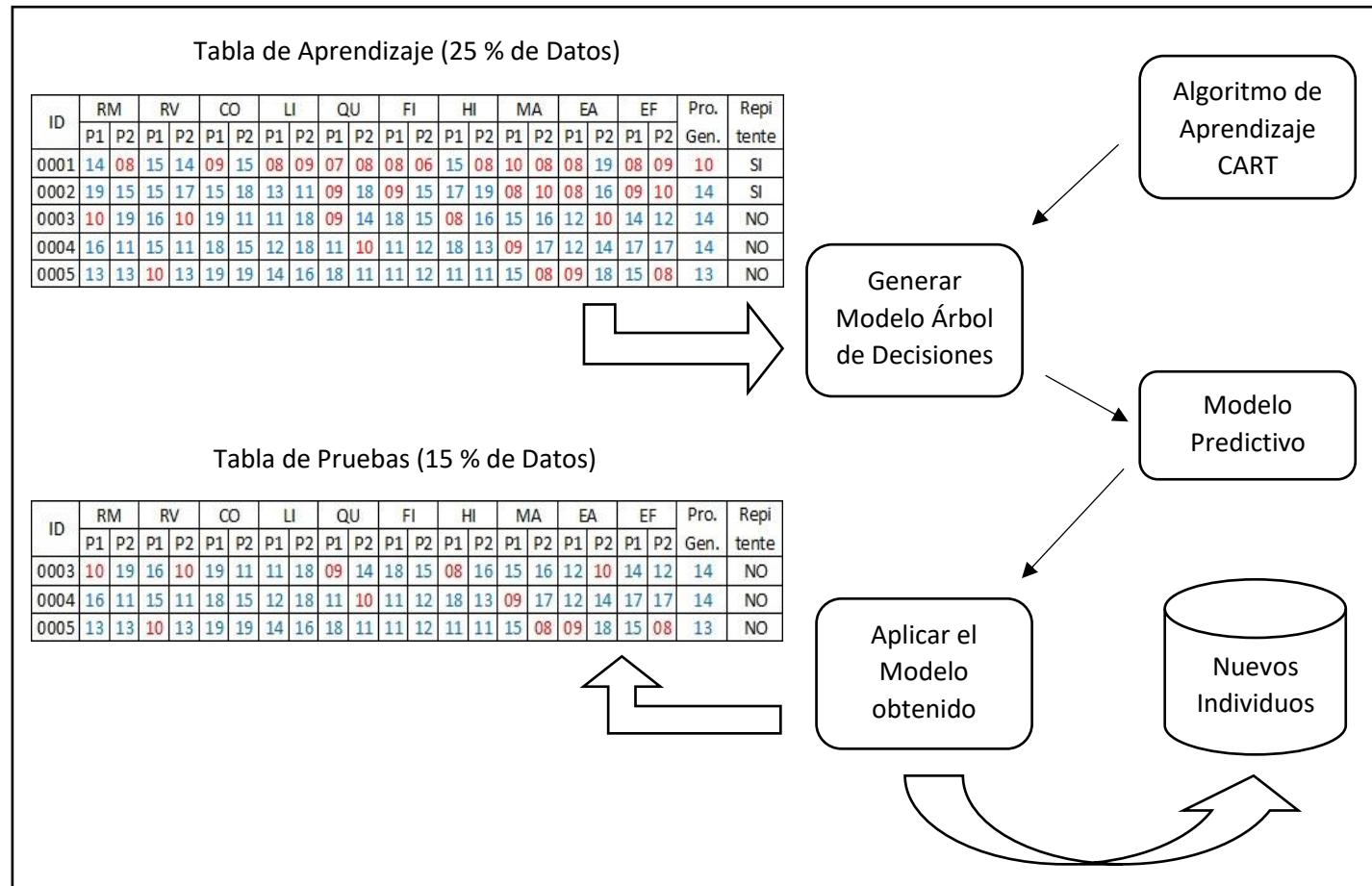
- Agrupación: Permitir la maximización de similitudes y minimización de diferencias entre objetos, mediante la aplicación de algún criterio de agrupación.
- Asociación: Tener presente que las reglas de asociación buscan descubrir conexiones existentes entre objetos identificados.
- Predicción: Descubrir el comportamiento de ciertos atributos en el futuro. (Regresión y series temporales, análisis discriminante, métodos bayesianos, algoritmos genéricos, árboles de decisión, redes neuronales).



Fuente: Elaboración propia.

A continuación, se muestra el proceso para la generación del modelo predictivo en sus diferentes fases. Teniendo en cuenta que, para la construcción del modelo, en la tabla de aprendizaje, se tomó el 25% de los datos (cinco registros), donde se visualiza dos registros que han repetido una o más veces y tres registros que no han repetido el año escolar y en la tabla de pruebas, se tomó el 15% (tres registros), donde se visualiza tres registros que no han repetido el año escolar y servirán de modelo.

Grafico Nro. 15: Esquema para la generación del modelo predictivo de deserción escolar.



Fuente: Elaboración propia.

Grafico Nro. 16: Árbol de decisión (con Rattle de R)

```

Resumen del modelo Árbol de decisión de Clasificación (construido con 'rpart'):
n= 378

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

1) root 378 119 NO (0.68518519 0.31481481)
2) NIVEL=DECIMO,NOVENO,OCTAVO,SEPTIMO,SEXTO 267 8 NO (0.97003745 0.02996255)
4) NOTA.15=1,0,4,5,6,0,6,3,6,5,6,6,6,6,6,9,7,0,7,1,7,3,7,5,7,6,7,7,7,8,8,0,8,1,8,2,8,3,8,4,8,5,8,6,8,8,8,9,9,0,9,2,9,3,9,4 229 0 NO (1.00000000 0.00000000) *
5) NOTA.15=7,2,7,4,7,9,8,7,9,1 38 8 NO (0.78947368 0.21052632)
10) NOTA.25=3,4,6,7,6,8,7,2,7,4,7,5,7,7,7,8,7,9,8,2,8,3,8,4,8,6,8,8,9,0,9,1,9,2,9,5,9,4 33 3 NO (0.90909090 0.09090909) *
11) NOTA.25=6,7,6,8,8,8,7 8 0 SI (0.00000000 1.00000000) *
3) NIVEL=CUARTO,PRIMERO,QUINTO,SEGUNDO,TERCERO 111 0 SI (0.00000000 1.00000000) *

Classification tree:
rpart(formula = DESERTOR ~., data = crsfdataset[crsftrain, c(crsfinput,
crsfTarget)], method = "class", par = list(split = "information"),
control = rpart.control(minsplit = 1, minbucket = 4, maxdepth = 3,
use surrogate = 0, maxsurrogate = 0))

Variables actually used in tree construction:
[1] NIVEL NOTA.15 NOTA.25

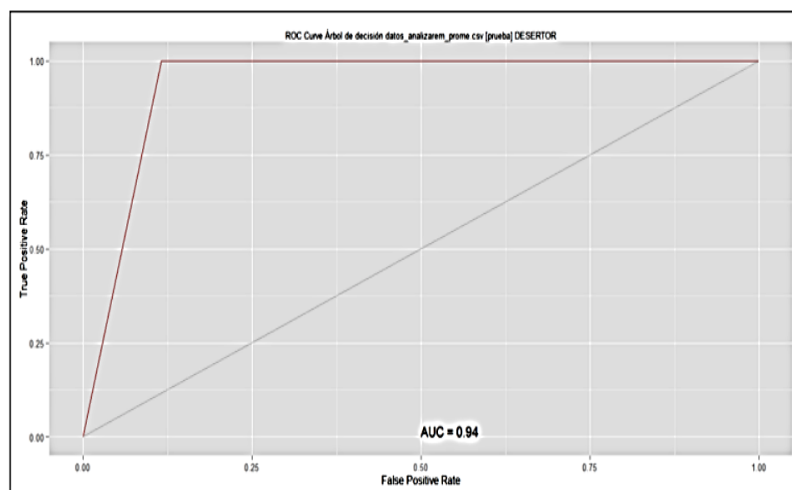
Root node error: 119/378 = 0.31481
n= 378 |
    
```

Fuente: Elaboración propia.

### 5. Evaluación

Para la evaluación del modelo, se construyó la curva Receiver Operating Characteristic (ROC) para medir su efectividad donde se muestra al modelo con un 94% de efectividad en la predicción.

Grafico Nro. 17: Evaluación del modelo



Fuente: Elaboración propia.



**Resultados:** Para probar el modelo, se utilizaron los datos de la tabla de pruebas, mostrando los siguientes.

Tabla Nro. 27: Historial de cálculos

ID	Gen.	Edad	Periodo	Grado/Año	Pro. Gen.	Repite	Prob.
0001	M	17	2020	Quinto	10	SI	81.00%
0002	M	18	2020	Quinto	14	SI	65.00%
0003	M	16	2020	Quinto	14	NO	20.00%
0004	F	16	2020	Quinto	14	NO	20.00%
0005	M	15	2020	Quinto	13	NO	12.00%

Fuente: Elaboración propia.

Los resultados del análisis del árbol de decisión, encontraron que la profundidad de la estructura del árbol era de cuatro niveles de profundidad. El primer nodo se formó con la variable repitente, el segundo nodo se formó con la variable promedio 1° semestre, el tercer nodo con la variable promedio 2° semestre y el cuarto nodo, con la variable condición.

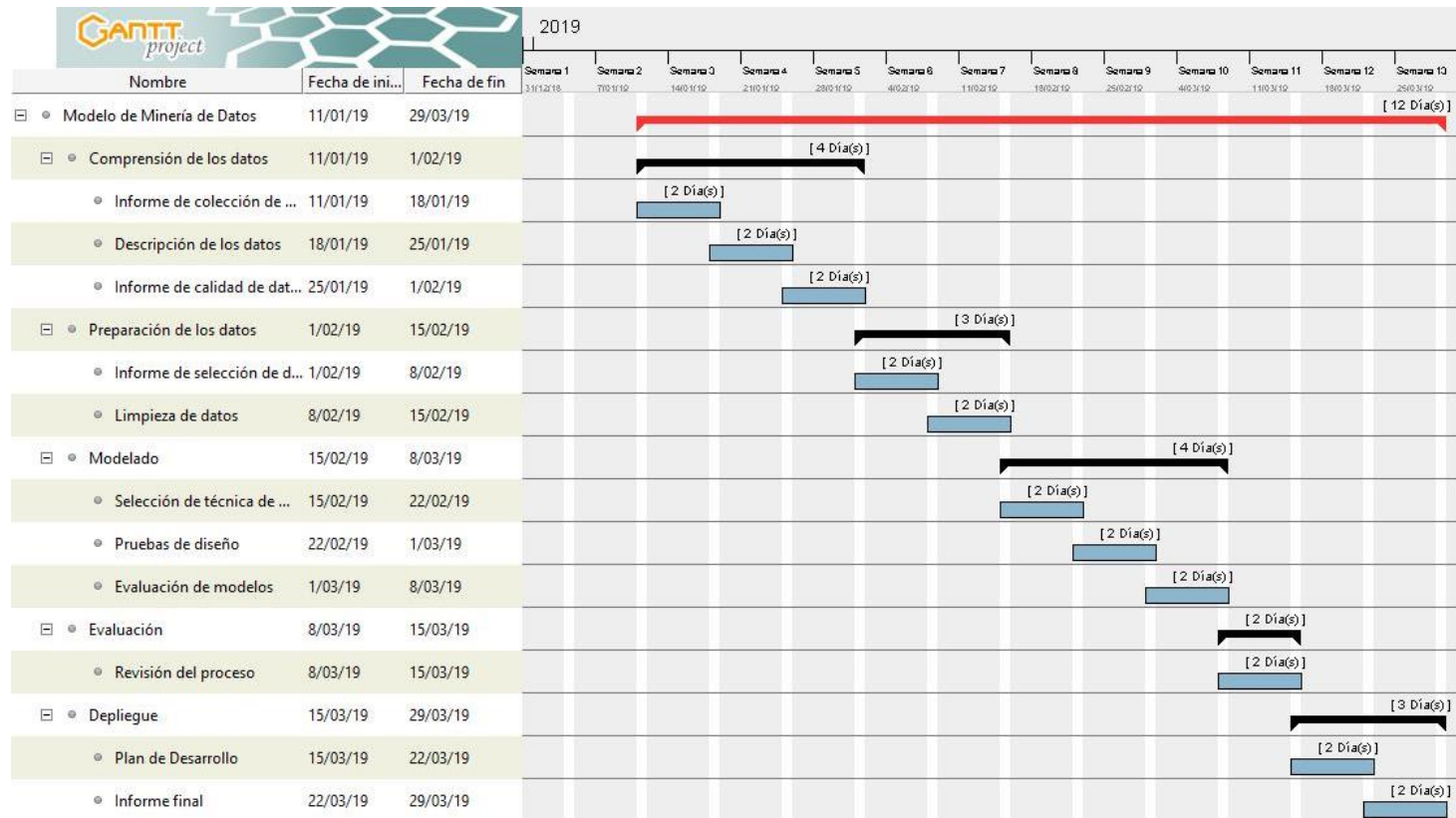
**Conclusiones:** De las 12 variables que se utilizaron para la construcción del modelo, la variable edad, es aquella que mayor incide en la deserción, junto a la variable promedio del primer y segundo semestre que resulta en promedio general con una intensidad similar, siendo así que mayor tendencia a desertar es para los que cuentan con edad más avanzada y el rendimiento académico determinado por los promedios, los estudiantes que se encuentren con la edad apropiada o menor edad muestran menos probabilidad y no son posibles desertores.

De los 20 datos registrados de los alumnos analizados la variable edad muestran mayor cantidad de datos atípicos conjuntamente con

la variable promedio que no contaba con datos en blanco y el modelo propuesto abre oportunidades, para la creación de nuevos, modelos de predicción, usando técnicas de clasificación, más complejas como redes neuronales y regresión, que permitan un análisis comparativo, de los factores que influyen en la deserción escolar.

### 5.3.2. Diagrama de Gantt

Gráfico Nro. 18: Diagrama de Gantt Implementación



Fuente: Elaboración Propia

5.3.3. Propuesta económica

Tabla Nro. 28: Presupuesto de la Implementación

DESCRIPCIÓN	N°	COSTO UNIT. S/.	SUB TOTAL S/.	TOTAL S/.
<b>1. BIENES DE INVERSIÓN</b>				
1.1. Equipo de Computo	1	S/1,000.00	S/1,000.00	S/2,550.00
1.2. Impresora	1	S/300.00	S/300.00	
1.3. Software Especializado	1	S/500.00	S/500.00	
1.4. Disco Duro Externo	1	S/250.00	S/250.00	
1.5. Material Bibliográfico	5	S/100.00	S/500.00	
<b>2. BIENES DE CONSUMO</b>				
2.1. Papel bond A4 80g. (Millar)	5	S/12.00	S/60.00	S/610.00
2.2. Material de escritorio	1	S/50.00	S/50.00	
2.3. Tóner para Impresora	2	S/250.00	S/500.00	
<b>3. SERVICIOS</b>				
3.1. Servicio de Internet	1	S/80.00	S/80.00	S/180.00
3.2. Servicio Energía Eléctrica	1	S/100.00	S/100.00	
<b>4. REMUNERACIONES</b>				
4.1. Especialista en Sistemas	1	S/9,000.00	S/9,000.00	S/13,500.00
4.2. Profesional en Estadístico	1	S/3,000.00	S/3,000.00	
4.3. Operador de Sistema	1	S/1,500.00	S/1,500.00	
<b>TOTAL</b>				<b>S/16,840.00</b>

Fuente: Elaboración Propia

## VI. CONCLUSIONES

Acorde a los resultados obtenidos y luego de ser analizados e interpretados, existe la necesidad de implementación de Minería de Datos en el Colegio Nacional de la Libertad, con la finalidad de mejorar el proceso actual de identificación de posibles casos de riesgo, debido al índice considerable de insatisfacción que existe frente a la aprobación en su totalidad de la necesidad de mejorar el proceso. Por ello, se concluye que la implementación de Minería de Datos mejoró el procedimiento para identificar patrones de comportamiento de la deserción escolar, ya que esta mejora permite que la información se encuentre organizada de manera digital y en tiempo real para facilitar la toma de decisiones respecto a la deserción escolar. Estos resultados concuerdan con la hipótesis general, concluyendo de esta manera que la hipótesis general queda aceptada. En referencia a lo especificado en las hipótesis específicas se tienen las siguientes conclusiones específicas:

1. La identificación de la situación actual permitió reconocer los problemas y necesidades que existen en el proceso actual para identificar casos de riesgo de deserción escolar en el Colegio Nacional de la Libertad, y con ello se logró identificar la necesidad de mejorar dicho proceso, de esta manera podemos decir, que la hipótesis específica queda aceptada.
2. La aplicación de la metodología Knowledge Discovery in Databases – KDD, permitió encontrar patrones de comportamiento relacionada a la deserción escolar en el Colegio Nacional de la Libertad, ya que se logró la implementación y mejora del proceso actual, de esta manera podemos decir que la hipótesis específica queda aceptada.
3. El diseño de la implementación de Minería de Datos permitió mejorar el proceso actual de identificación de patrones de comportamiento de la deserción escolar en el Colegio Nacional de la Libertad, ya que se logró

observar la mejorar en la toma de decisiones, de esta manera podemos decir que la hipótesis específica queda aceptada

Como aporte a la investigación tenemos que, la implementación de Minería de Datos el cual permite identificar patrones de comportamiento que influye en la deserción escolar mediante el comportamiento de la información almacenada de los alumnos de la Institución Educativa, permite que dichos patrones sean considerados como factores determinantes y constantes que se deben considerar como responsables directos de la deserción escolar y que al ser identificados de manera anticipada se puede prevenir los posibles casos de riesgos de abandono escolar.

Como valor agregado se considera que, la implementación de Minería de Datos que sea capaz de predecir de manera anticipada mediante el análisis del comportamiento de la información en los alumnos del Colegio Nacional de la Libertad de la ciudad de Huaraz, los posibles casos de riesgo de deserción escolar, permite dejar en evidencia que mediante el uso de las tecnología de la Información y comunicación (TIC) por medio de una de sus innumerables herramientas como es la inteligencia artificial utilizada en la presente tesis, se puede simplificar los diferentes procesos para el cumplimiento de las actividades y obligaciones de la Institución Educativa, quedando demostrado que en la actualidad la automatización de la información es fundamental para contar en tiempo real con datos preciso que permitan la mejora de manera significativa la toma de decisiones en la intervención oportuna en los diferentes caso que se presente.

## VII. RECOMENDACIONES

1. Que se realice el proceso de recolección y actualización de información referente a temas particulares relacionados directamente con los estudiantes los cuales deben ser almacenada en bases de datos dentro de un ordenador de desde el inicio al término de su vida estudiantil, además se debe implemente métodos de validación de dicha información recolectada, permitiendo así contar con datos de calidad que permitan alimentar futuras implementaciones de diversos sistemas de mejoramiento de los proceso realizado en la Institución educativa.
2. Que se capacite de forma constante al personal docente, auxiliar, administrativos y de apoyo sobre el uso de las tecnologías de la información y comunicación (TIC) teniendo en cuenta el cargo o responsabilidad en la Institución Educativa, permitiendo de esta manera su adaptación de manera progresiva a una nueva era que ya se encuentra presente en nuestra vida cotidiana y que aun esta en evolución logrando abarcar cada vez más los diferentes procesos tradicionales que se desarrolla sin el uso de la TIC y mejorando, mediante las actualizaciones, los procesos automatizados implementados en la Institución Educativa
3. Que dentro de los presupuestos otorgados por el gobierno para el gasto público se debe considerar la adquisición de nuevas herramientas de las Tecnologías de la Información y Comunicación con la capacidad de soportar la implementación de aplicaciones conocidos actualizados a sus últimas versiones, la implementación de nuevas aplicaciones desarrolladas que automatizan diferentes procesos que utilizando grandes recursos y que permitan el almacenamiento de gran cantidad información clasificada y de calidad dentro de un ordenador.

4. Que se implemente un área de soporte informático dentro de la Institución Educativa bajo la administración de personal capacitado que se dedique exclusivamente al mantenimiento tanto preventivo como correctivo de las diferentes herramientas de las tecnologías de información y comunicación (TIC), además contar con la capacidad de proporcionar asistencia a los usuarios ante posibles problemas que se puedan presentar al utilizar un producto o servicio, ya sea este el hardware o software de un ordenador, el servicio de Internet, configuración de los periféricos de entrada o salida, o de cualquier otro equipo o dispositivo utilizado en la Institución Educativa.



## VIII. REFERENCIAS BIBLIOGRÁFICAS

1. Hernández Orallo J, Ramírez Quintana J, Ferri Ramírez C. Introducción a la Minería de Datos. 20048420540919th ed. Madrid, España: Pearson Educación S. Æ; 2004.
2. Espinoza O, Castillo D, González LE, Loyola J. Web Site de Luis Eduardo González Fiegehen. [Online].; 2010 [cited 2020 08 02. Available from: <http://www.luisedogonzalez.cl/pdf/2010/2010-03.pdf>.
3. Alarcón Glasinovich W. CLACSO - Consejo Latinoamericano de Ciencias Sociales. [Online].; 2016 [cited 2020 08 07. Available from: [http://biblioteca.clacso.edu.ar/Peru/desco/20170223025151/pdf\\_1301.pdf](http://biblioteca.clacso.edu.ar/Peru/desco/20170223025151/pdf_1301.pdf).
4. Cueto Caballero S. ESCALE - Estadística de la Calidad Educativa. [Online].; 2004 [cited 2020 08 02. Available from: <http://repositorio.minedu.gob.pe/handle/123456789/150>.
5. Alcázar L. Asistencia y deserción en escuelas secundarias rurales del Perú. 2008th ed. Benavides M, editor. Lima: GRADE, Grupo de Análisis para el Desarrollo; 2008.
6. Oñate Bowen AA. Análisis de la Deserción y Permanencia Académica en la Educación Superior Aplicando Minería De Datos. Tesis o trabajo de investigación presentado como requisito parcial para optar al título de: MSc. en Ingeniería de Sistemas y Computación. Bogotá, Colombia: Universidad Nacional de Colombia; 2016.
7. Vásquez Verdugo J. Modelo predictivo para estimar la deserción de estudiantes en una institución de educación superior. Tesis para optar al grado de magíster en control de gestión. Santiago, Chile: Universidad de Chile; 2016.
8. Córdova J. Aplicación de técnicas de minería de datos para predecir la deserción de los estudiantes que pertenecen al Colegio Fisco-Misional “San Francisco” de la Ciudad de Ibarra. Tesis de grado previo a la obtención del Título de Ingeniera

en Sistemas e Informática. Ibarra, Ecuador: Universidad Regional Autónoma de los Andes; 2014.

9. Yamao E. Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la escuela profesional de Ingeniería de Computación y Sistemas, Universidad de San Martín de Porres, Lima-Perú. Tesis para optar el grado de maestro en ingeniería de computación y sistemas con mención en gestión de Tecnologías de Información. Lima – Perú: Universidad de San Martín de Porres; 2018.
10. Piscoya L. Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la Región de Lambayeque. Tesis para optar el Título Profesional de Ingeniero de Sistemas. Chiclayo, Lambayeque: Universidad Señor de Sipán; 2016.
11. Gálvez M, Flores K. Modelo predictivo de deserción universitaria de la carrera de Ingeniería Informática en la Universidad Ricardo Palma. Tesis para obtener el título profesional de Ingeniera Informática. Lima – Perú: Universidad Ricardo Palma; 2015.
12. Duran E. Data Mart como parte de una solución de inteligencia de negocios, para el soporte de la toma de decisiones de la gestión académica de la Facultad de Ciencias en la UNASAM. Tesis para optar el título de Ingeniero de Sistemas e Informática. Huaraz: Universidad Nacional Santiago Antúnez de Mayolo; 2017.
13. Torres L. Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz – 2016. Tesis para obtener el grado de doctor en administración de la educación. Caraz, Ancash: Universidad César Vallejo; 2017.
14. Díaz S. Sistema integral bajo el enfoque de minería de datos y redes neuronales para la predicción y control de la contaminación atmosférica por pm10 en la Ciudad de Chimbote. Tesis Doctoral. Chimbote, Santa, Ancash: Universidad Nacional del Santa; 2014.

15. Superintendencia Nacional de Administración Tributaria - SUNAT. Consulta RUC - SUNAT. [Online].; 1997 [cited 08 08 2020. Available from: <https://e-consultaruc.sunat.gob.pe/cl-ti-itmrconsruc/jcrS00Alias>.
16. Colegio Nacional De La Libertad. Sitio web del Colegio Nacional De La Libertad - Nosotros. [Online].; 2010 [cited 2020 08 08. Available from: <http://ciberinformatica.webcindario.com/nosotros.html>.
17. Colegio Nacional De La Libertad. Sitio web del Colegio Nacional De La Libertad - Ubicación Geográfica. [Online].; 2010 [cited 2020 15 08. Available from: <http://ciberinformatica.webcindario.com/ubicacion.html>.
18. Colegio Nacional De La Libertad. Sitio web del Colegio Nacional De La Libertad - Resumen Historico. [Online].; 2010 [cited 2020 08 08. Available from: <http://ciberinformatica.webcindario.com/imagenes/flash/Resumen%20hist%C3%B3rico.pdf>.
19. Perez M. SCielo Perú. [Online].; 2017 [cited 2020 08 08. Available from: [http://www.scielo.org.pe/scielo.php?script=sci\\_arttext&pid=S1018-130X2017000400008](http://www.scielo.org.pe/scielo.php?script=sci_arttext&pid=S1018-130X2017000400008).
20. Equipo de Expertos de la Universidad Internacional de Valencia. Universidad Internacional de Valencia. [Online].; 2018 [cited 2020 08 08. Available from: <https://www.universidadviu.com/la-evolucion-las-tics-desde-abaco-al-smartphone/>.
21. Marqués M. Meta biblioteca. [Online].; 2011 [cited 2020 08 29. Available from: <https://libros.metabiblioteca.org/bitstream/001/353/5/978-84-693-0146-3.pdf>.
22. López Abellán J. Economipedia. [Online].; 2020 [cited 2020 08 08. Available from: <https://economipedia.com/definiciones/mineria-de-datos.html>.
23. Cuji B, Gavilanes W, Sanchez R. Revista Espacios. [Online].; 2017 [cited 2020. Available from: <https://www.revistaespacios.com/a17v38n55/a17v38n55p17.pdf>.
24. Wikipedia. Wikipedia la enciclopedia libre. [Online].; 2020 [cited 2020 08 28. Available from: <https://es.wikipedia.org/wiki/OLAP>.

25. Ahumada Tello E, Perusquia Velasco JMA. Scielo. [Online].; 2016 [cited 2020 08 08. Available from: [http://www.scielo.org.mx/scielo.php?script=sci\\_arttext&pid=S0186-10422016000100127](http://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0186-10422016000100127).
26. Salesforce. Salesforce Blog. [Online].; 2017 [cited 2020 08 08. Available from: <https://www.salesforce.com/mx/blog/2017/6/Que-es-la-inteligencia-artificial.html>.
27. Tecnologias Informacion. Tecnologias Informacion Sitio Web. [Online].; 2018 [cited 2020 08 08. Available from: <https://www.tecnologias-informacion.com/sistemas-expertos.html>.
28. Uvidia Fassler M, Cisneros Barahona A, Méndez Naranjo P, Villa Yáñez H. Dialnet. [Online].; 2018 [cited 2020 08 29. Available from: <https://dialnet.unirioja.es/descarga/articulo/6836545.pdf>.
29. Conecta Software. Conecta Software. [Online].; 2020 [cited 2020. Available from: <https://conectasoftware.com/analytics/arboles-de-decisiones-en-la-mineria-de-datos/>.
30. Sucar LE. Instituto Nacional de Astrofísica, Óptica y Electrónica. [Online].; 2012 [cited 2020 09 30. Available from: <https://ccc.inaoep.mx/~esucar/Clases-mgp/caprb.pdf>.
31. Beltrán Martínez B. Benemérita Universidad Autónoma de Puebla. [Online].; 2018 [cited 2020 08 29. Available from: <http://bbeltran.cs.buap.mx/NotasMD.pdf>.
32. Wikipedia®. Wikipedia - Perceptrón multicapa. [Online].; 2020 [cited 2020 09 30. Available from: [https://es.wikipedia.org/wiki/Perceptr%C3%B3n\\_multicapa](https://es.wikipedia.org/wiki/Perceptr%C3%B3n_multicapa).
33. Oocities.org. Oocities.org - SISTEMAS Y HERRAMIENTAS DE MINERÍA DE DATOS. [Online].; 1994 [cited 2020 09 30. Available from: [http://www.oocities.org/es/mineria.datos/sistemas\\_herramientas\\_mineria\\_datos.pdf](http://www.oocities.org/es/mineria.datos/sistemas_herramientas_mineria_datos.pdf).

34. Botero M. SCIPEDIA - Metodologías para el Desarrollo de Proyectos de Minería de Datos. [Online].; 2020 [cited 2019 09 30. Available from: [https://www.scipedia.com/public/Botero\\_Villada\\_2019a](https://www.scipedia.com/public/Botero_Villada_2019a).
35. Hernández Sampieri R, Fernández Collado C, Baptista Lucio MdP. Metodología de la investigación. Quinta edición ed. Mares Chacón J, editor. Mexico: MCGRAW-HILL / INTERAMERICANA EDITORES, S.A. DE C.V.; 2010.
36. Pérez Porto J, Merino M. Definicion.de - Definición de Deserción Escolar. [Online].; 2008 [cited 2020 09 30. Available from: <https://definicion.de/desercion-escolar/>.
37. Comité Institucional de Ética en Investigación. Universidad Católica Los Ángeles de Chimbote. [Online].; 2019 [cited 2020 08 15. Available from: <https://www.uladech.edu.pe/images/stories/universidad/documentos/2019/codigo-de-etica-para-la-investigacion-v002.pdf>.
38. UIAF - Unidad de Información y Análisis Financiero. Universidad del Rosario de Colombia. [Online].; 2014 [cited 2020 09 30. Available from: [https://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos\\_Lavados/Tecnicas-de-mineria-de-datos-para-la-prevencion-de.pdf](https://www.urosario.edu.co/observatorio-de-lavado-de-activos/Archivos_Lavados/Tecnicas-de-mineria-de-datos-para-la-prevencion-de.pdf).

# ANEXOS

### ANEXO NRO. 1: CRONOGRAMA DE ACTIVIDADES

N°	Actividades	Año 2020 - Semestre I							
		1	2	3	4	5	6	7	8
1	Elaboración del Proyecto	x							
2	Revisión del proyecto por el jurado de investigación		x						
3	Aprobación del proyecto por el Jurado de Investigación			x					
4	Exposición del proyecto al Jurado de Investigación				x				
5	Mejora del marco teórico y metodológico					x			
6	Elaboración y validación del instrumento de recolección de Información						x		
7	Elaboración del consentimiento informado							x	

Fuente: Elaboración Propia

## ANEXO NRO. 2: PRESUPUESTO

TITULO: PROPUESTA DE IMPLEMENTACIÓN DE MINERÍA DE DATOS  
PARA EL COLEGIO NACIONAL DE LA LIBERTAD – HUARAZ; 2020.

ESTUDIANTE: RONDO ROPA ELÍAS MANUEL

INVERSIÓN: S/. 2,384.00

FINANCIAMIENTO: RECURSOS PROPIOS

DESCRIPCIÓN	CANT.	COSTO UNITARIO S/.	TOTAL PARCIAL S/.	TOTAL S/.
<b>1. BIENES DE INVERSIÓN</b>				
1.1. Equipo de Computo	1	S/1,000.00	S/1,000.00	S/1,650.00
1.2. Impresora	1	S/300.00	S/300.00	
1.3. Software Especializado	1	S/350.00	S/350.00	
<b>2. BIENES DE CONSUMO</b>				
2.1. Papel bond A4 80g. (Millar)	2	S/12.00	S/24.00	S/554.00
2.2. Material de escritorio	1	S/30.00	S/30.00	
2.3. Tóner para Impresora	2	S/250.00	S/500.00	
<b>3. SERVICIOS</b>				
3.1. Servicio de Internet	1	S/80.00	S/80.00	S/180.00
3.2. Servicio Energía Eléctrica	1	S/100.00	S/100.00	
<b>TOTAL</b>				<b>S/2,384.00</b>

Fuente: Elaboración Propia



### ANEXO NRO. 3: CUESTIONARIO DE ALUMNOS

TITULO: PROPUESTA DE IMPLEMENTACIÓN DE MINERÍA DE DATOS PARA EL COLEGIO NACIONAL DE LA LIBERTAD – HUARAZ; 2020.

TESISTA: RONDO ROPA ELÍAS MANUEL

PRESENTACIÓN: El presente instrumento forma parte de la investigación antes mencionada, se solicita su participación respondiendo cada pregunta de manera objetiva y veraz. La información a proporcionar es de carácter confidencial y reservado y los resultados de la misma serán utilizados solo para efectos académicos y de investigación científica.

INSTRUCCIONES: A continuación, se le presenta una lista de preguntas, agrupadas por dimensión, las cuales se solicita se responda, marcando una sola alternativa en el recuadro correspondiente (SI o NO) según considere.

DIMENSIÓN 1: NIVEL DE SATISFACCIÓN SOBRE MÉTODO DE DETECCIÓN DE POSIBLES CASOS DE ABANDONO ESCOLAR.			
NRO.	PREGUNTA	SI	NO
1	¿Conoces algún compañero que tenga la intención de dejar sus estudios escolares y este siendo apoyado por el colegio para cambiar de opinión?		
2	¿Tus docentes te repiten constantemente sobre la importancia de terminar tus estudios escolares dentro de la edad apropiada?		
3	¿En algún momento tu colegio averiguo si tú o tus compañeros tienen familiares que no hayan culminado sus estudios escolares?		

4	¿Tus docentes han llegado a conocer en algún momento a tus amistades del colegio y de tu entorno social fuera del colegio?		
5	¿El colegio sabes si tú o quizás tus compañeros estudian y trabajan para aportar con ingresos económicos para los gastos tu hogar?		
6	¿Tus docentes conocen con quien vives, si es solo con tu papá, solo con tu mamá, con tu papá y tu mamá o con algún otro familiar que no sean tus padres?		
7	¿En la actualidad el colegio brinda información preventiva ante el incremento de la delincuencia alrededor de la Institución educativa o en la ciudad?		
8	¿Has recibido en tu colegio alguna charla u orientación sobre la prevención de embarazos en la etapa escolar?		
9	¿Has recibido en tu colegio alguna charla u orientación relacionada a la prevención de acoso sexual dentro o fuera de la institución?		
10	¿Conoces algún caso de maltrato físico o psicológico constante a uno o más compañeros, donde el colegio haya intervenido para solucionarlo?		

Fuente: Elaboración Propia

DIMENSIÓN 2: NECESIDAD DE IMPLEMENTACIÓN DE MINERÍA DE DATOS PARA LA DETERMINACIÓN DE PATRONES DE COMPORTAMIENTO.			
NRO.	PREGUNTA	SI	NO
1	¿Te gustaría que tu colegio cuenta con nuevas tecnologías que permitan mejorar cada día en la atención a los alumnos?		

2	¿Estarías de acuerdo con implementar de nuevas aplicaciones tecnológicas inteligentes que apoyen a tu colegio?		
3	¿Te gustaría que exista un sistema informático que sea capaz de apoyar a tus docentes, detectando de manera anticipada algún posible caso de abandono de estudios por parte de tus compañeros?		
4	¿Tu colegio almacena información en computadoras de datos personales, registros de calificaciones u otra información de los alumnos o docentes?		
5	¿Estarías de acuerdo de que tu colegio cuente con la posibilidad de reaccionar inmediatamente ante algún indicio detectado en tu persona?		
6	¿Crees que es importante que tu colegio cuente con información automatizada que le permita tomar mejores decisiones en un corto plazo?		
7	¿Estarías de acuerdo que tu colegio se anticipe detectando y apoyando a algún posible caso que se presente de abandono de estudios por parte de los alumnos?		
8	¿Consideras que sería bueno que tu colegio tenga la posibilidad de contar con información inmediata sin mucha demora para tomar alguna decisión importante?		
9	¿Para ti es importante que tu colegio cuente con información confiable y de calidad que se encuentre clasificada y almacenada en una computadora?		
10	¿Es importante que tu colegio cuente con un sistema que le permita analizar comportamientos de información ingresada, para la intervención oportuna de posibles casos de abandono de estudios?		

Fuente: Elaboración Propia

## ANEXO NRO. 4: CONSENTIMIENTO INFORMADO

### PROTOCOLO DE CONSENTIMIENTO INFORMADO PARA ENCUESTAS

La finalidad de este protocolo en Ingeniería y tecnología es informarle sobre el trabajo de investigación y solicitarle su consentimiento. De aceptar, el investigador y usted se quedarán con una copia.

La presente investigación se titula “Propuesta de Implementación de Minería de Datos para el Colegio Nacional de la Libertad – Huaraz; 2020” y es dirigido por Elías Manuel Rondo Ropa, investigador de la Universidad Católica Los Ángeles de Chimbote.

El propósito de la investigación es: Ejecutar la propuesta de implementación de minería de datos para el Colegio Nacional de la Libertad en la ciudad de Huaraz, con la finalidad de determinar patrones de comportamiento de la deserción escolar.

Para ello, se le invita a participar en una encuesta que le tomará 10 minutos de su tiempo. Su participación en la investigación es completamente voluntaria y anónima. Usted puede decidir interrumpirla en cualquier momento, sin que ello le genere ningún perjuicio. Si tuviera alguna inquietud y/o duda sobre la investigación, puede formularla cuando crea conveniente.

Al concluir la investigación, usted será informado de los resultados a través del repositorio de la universidad. Si desea, también podrá escribir al correo [rr.eliasm@hotmail.com](mailto:rr.eliasm@hotmail.com) para recibir mayor información. Asimismo, para consultas sobre aspectos éticos, puede comunicarse con el Comité de Ética de la Investigación de la universidad Católica los Ángeles de Chimbote.

Si está de acuerdo con los puntos anteriores, complete sus datos a continuación:

Nombre : \_\_\_\_\_

Fecha : \_\_\_/\_\_\_/2020.

Correo electrónico : \_\_\_\_\_

Firma del participante: \_\_\_\_\_

Firma del investigador (o encargado de recoger la información) : \_\_\_\_\_