UNIVERSIDAD NACIONAL DE UCAYALI

FACULTAD DE INGENIERÍA DE SISTEMAS E INGENIERÍA CIVIL

CARRERA PROFESIONAL DE INGENIERÍA DE SISTEMAS

"MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS PARA LA IRREGULARIDAD ESTUDIANTIL EN LA ESCUELA ACADÉMICA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNU, 2018"

TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE

INGENIERO DE SISTEMAS

ENRRIQUE IGLESIAS PORTUGUEZ DURAND

PUCALLPA – PERÚ

2022

UNIVERSIDAD NACIONAL DE UCAYALI

FACULTAD DE INGENIERÌA DE SISTEMAS E INGENIERÌA CIVIL
ESCUELA PROFESIONAL DE INGENIERÌA DE SISTEMAS
COMISIÒN DE GRADOS Y TITULOS

ACTA DE EVALUACIÒN DE SUSTENTACIÒN DE TESIS N° 009-2022

GRADUANDO : **Bach. PORTUGUEZ DURAND, ENRRIQUE IGLESIAS**

TEMA : "MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS PARA LA IRREGULARIDAD ESTUDIANTIL EN LA ESCUELA ACADÉMICA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNU, 2018".

ASESOR : Msc. PANDURO PADILLA, EUCLIDES.

CRITERIOS	PUNTAJE			
	3	2	1	0
I- PRESENTACION				
* MOTIVACION		X		
* TONO DE VOZ		X		
* CALIDAD DE MATERIAL AUDIOVISUAL		X		
II- DESARROLLO DE CONTENIDO				
* SECUENCIA		X		
* DOMINIO DE VOZ	X			
* USO ADECUADO DEL MATERIAL		X		
* VALIDEZ DE LOS HALLAZGOS			X	
III- ABSOLUCION DE PREGUNTAS				
* RESPONDE A LAS PREGUNTAS FORMULADAS POR EL JURADO			X	
IV- APORTE CIENTIFICO Y TECNOLOGICO				
* AL DESARROLLO DE LA COMUNIDAD		X		
* A LA PROFESION		X		

EVALUACION

- EXCELENCIA 26 A 30 PUNTOS
- UNANIMIDAD 21 A 25 PUNTOS
- MAYORIA 16 A 20 PUNTOS
- DESAPROBADO 15 A MENOS

CALIFICACION FINAL:

PUNTAJE : 19/ 30

CONCLUSIONES:

APROBADO POR : MAYORIA
DESAPROBADO POR :

JURADO EVALUADOR

- DR. JORGE LUIS HILARIO RIVAS presidente
- Ing. Mg. ARTURO YUPANQUI VILLANUEVA Miembro
- Ing. MSC. FREDDY ELAR FERRARI FERNANDEZ Miembro

SECRETARIA ACADÉMICA: ING. MSC. CLOTILDE RÍOS HIDALGO DE CERNA

PUCALLPA, 30 De junio Del 2022

"MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS PARA LA IRREGULARIDAD ESTUDIANTIL EN LA ESCUELA ACADÉMICA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNU, 2018"

Elaborado por:

Bach. Enrrique Iglesias Portuguez Durand

Aprobado por:

Dr. Jorge Luis Hilario Rivas
Presidente

Ing. Mg. Arturo Yupanqui Villanueva
Miembro

Ing. Mg. Freddy Elar Ferrari Fernandez
Miembro

Asesor:

Ing. Msc. Euclides Panduro Padilla
Miembro

UNIVERSIDAD NACIONAL DE UCAYALI
VICERRECTORADO DE INVESTIGACION
DIRECCION GENERAL DE PRODUCCION INTELECTUAL

CONSTANCIA

ORIGINALIDAD DE TRABAJO DE INVESTIGACION

SISTEMA ANTIPLAGIO URKUND

N°0025-2022

La **Dirección de Producción Intelectual**, hace constar por la presente, que el **Informe final de Tesis, titulado:**

"MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS PARA LA IRREGULARIDAD ESTUDIANTIL EN LA ESCUELA ACADÉMICA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNU, 2018"

Cuyo(s) autor (es) : **PORTUGUEZ DURAND, ENRRIQUE IGLESIAS.**

Facultad : **INGENIERÍA DE SISTEMAS E INGENIERÍA CIVIL.**
Escuela Profesional : **ING. SISTEMAS.**
Asesor(a) : **MSC. PANDURO PADILLA, EUCLIDES.**

Después de realizado el análisis correspondiente en el Sistema Antiplagio URKUND, dicho documento presenta un **porcentaje de similitud de 0%.**
En tal sentido, de acuerdo a los criterios de porcentaje establecidos en el artículo 9 de la DIRECTIVA DE USO DEL SISTEMA ANTIPLAGIO URKUND, el cual indica que no se debe superar el 10%. Se declara, que el trabajo de investigación: SI Contiene un porcentaje aceptable de similitud, por lo que SI se aprueba su originalidad.

En señal de conformidad y verificación se FIRMA Y CODIFICA la presente constancia

FECHA 24/01/2022

Dr. ABRAHAM ERMITANIO HUAMAN ALMIRON
Dirección de Producción Intelectual

AUTORIZACIÓN DE PUBLICACIÓN DE TESIS

REPOSITORIO DE LA UNIVERSIDAD NACIONAL DE UCAYALI

Yo, Enrrique Iglesias Portuguez Durand.
Autor de la TESIS titulada:
"Modelo predictivo aplicando Técnicas de minería de datos para la irregularidad estudiantil en la escuela académica profesional de Ingeniería de Sistemas de la UNU, 2018"

Sustentada el año: 2022
Con la asesoría de: Msc. Euclides Panduro Padilla
En la Facultad de: Ingeniería de Sistemas e Ingeniería Civil
Carrera Profesional de: Ingeniería de Sistemas

Autorizo la publicación:

PARCIAL [ ] Significa que se publicará en el repositorio institucional solo La caratula, la dedicatoria y el resumen de la tesis. Esta opción solo es válida marcar **si su tesis o documento presenta material patentable**, para ello deberá presentar el trámite de CATI y/o INDECOPI cuando se lo solicite la DGPI UNU.

TOTAL [X] Significa que todo el contenido de la tesis y/o documento será publicada en el repositorio institucional.

De mi trabajo de investigación en el Repositorio Institucional de la Universidad Nacional de Ucayali (www.repositorio.unu.edu.pe), bajo los siguientes términos:

Primero: Otorgo a la Universidad Nacional de Ucayali **licencia no exclusiva** para reproducir, distribuir, comunicar, transformar (únicamente mediante su traducción a otros idiomas) y poner a disposición del público en general mi tesis (incluido el resumen) a través del Repositorio Institucional de la UNU, en formato digital sin modificar su contenido, en el Perú y en el extranjero; por el tiempo y las veces que considere necesario y libre de remuneraciones.

Segundo: Declaro que la **tesis es una creación de mi autoría** y exclusiva titularidad, por tanto, me encuentro facultado a conceder la presente autorización, garantizando que la tesis no infringe derechos de autor de terceras personas, caso contrario, me hago único(a) responsable de investigaciones y observaciones futuras, de acuerdo a lo establecido en el estatuto de la Universidad Nacional de Ucayali y del Ministerio de Educación.

En señal de conformidad firmo la presente autorización.

Fecha: 30 / 06 / 2022

Email: portuguezei@gmail.com
Firma:
Teléfono: 921423211
DNI: 73372296

DEDICATORIA

A Dios, por su inmenso amor y su cálida compañía en cada momento de alegría y tristeza.

A Xavier y Rosa, mis amados padres, por su incondicional apoyo, quienes con arduo esfuerzo supieron orientarme con paciencia y amor.

AGRADECIMIENTO

A los docentes de la carrera profesional de ingeniería de sistemas de la Universidad Nacional de Ucayali, por sus orientaciones, consejos y motivación para continuar creciendo profesionalmente.

A la PhD. Masun Nabhan Homsi, por su paciencia, dedicación, orientación y enseñanza.

ÍNDICE DE CONTENIDO

Pág.

DEDICATORIA vi

AGRADECIMIENTO vii

ÍNDICE DE CONTENIDO viii

ÍNDICE DE CUADROS xi

ÍNDICE DE IMÁGENES xiii

ÍNDICE DE FIGURAS xiv

ÍNDICE DE ANEXOS xv

RESUMEN xvi

ABSTRACT xvii

INTRODUCCIÓN xviii

CAPÍTULO I. PLANTEAMIENTO DEL PROBLEMA 1

1.1. ANTECEDENTES DEL PROBLEMA 1

1.2. FORMULACIÓN DEL PROBLEMA 3

1.2.1. Problema General 3

1.2.2. Problemas Específicos 3

1.3. OBJETIVOS DE LA INVESTIGACIÓN 4

1.3.1. Objetivo General 4

1.3.2. Objetivos Específicos 4

1.4. JUSTIFICACIÓN E IMPORTANCIA 5

1.4.1. Justificación 5

1.4.2. Importancia 6

1.5. HIPÓTESIS Y/O SISTEMA DE HIPÓTESIS 6

1.5.1. Hipótesis General 6

1.5.2. Hipótesis Específicas 6

1.6. VARIABLES (DEFINICIÓN CONCEPTUAL Y OPERACIONAL) 8
CAPÍTULO II. MARCO TEÓRICO..... 9
2.1. ANTECEDENTES DE LA INVESTIGACIÓN 9
2.1.1. Tesis Internacionales..... 9
2.1.2. Tesis Nacionales 11
2.2. DEFINICIÓN DE TÉRMINOS BÁSICOS 12
CAPÍTULO III. METODOLOGÍA..... 18
3.1. MÉTODO DE INVESTIGACIÓN..... 18
3.1.1. Tipo de la Investigación..... 18
3.1.2. Nivel de la Investigación..... 18
3.1.3. Diseño de la Investigación..... 18
3.2. POBLACIÓN Y MUESTRA..... 19
3.2.1. Población..... 19
3.2.2. Muestra..... 19
3.3. PROCEDIMIENTO DE RECOLECCIÓN DE DATOS..... 20
3.4. TÉCNICAS E INSTRUMENTOS PARA RECOLECCIÓN DE DATOS 20
3.4.1. Técnica..... 20
3.4.2. Instrumento..... 20
3.5. PROCESAMIENTO DE RECOLECCIÓN DE DATOS..... 22
3.6. TRATAMIENTO DE DATOS..... 23
CAPÍTULO IV. RESULTADOS..... 25
4.1. SELECCIÓN DE DATOS..... 25
4.2. PRE – PROCESAMIENTO DE DATOS..... 52
4.3. TRANSFORMACIÓN DE DATOS 58
4.4. MINERÍA DE DATOS 62

4.5. INTERPRETACIÓN 73

CAPÍTULO V. DISCUSIÓN 85

CONCLUSIONES 87

RECOMENDACIONES 89

REFERENCIA BIBLIOGRÁFICA 90

ANEXO 95

ÍNDICE DE CUADROS

Pág.

Cuadro 1: Definición conceptual y operacional de las variables. 8

Cuadro 2: Entorno físico de la vivienda..... 28

Cuadro 3: Lugar de vivienda 28

Cuadro 4: Zona de vivienda 29

Cuadro 5: Material de vivienda..... 29

Cuadro 6: Tiempo que radica en Pucallpa 30

Cuadro 7: Número de hijos 31

Cuadro 8: Edad de hijos..... 32

Cuadro 9: Estado civil 32

Cuadro 10: Condición laboral..... 33

Cuadro 11: Dependencia económica 33

Cuadro 12: Percibe económicamente – mensual..... 34

Cuadro 13: Nivel educativo del padre 35

Cuadro 14: Nivel educativo de la madre 36

Cuadro 15: Trabajo del padre 37

Cuadro 16: Trabajo de la madre 39

Cuadro 17: Edad..... 40

Cuadro 18: Sexo 41

Cuadro 19: Horas que dedica a estudiar al día..... 41

Cuadro 20: Días que estudia fuera de la universidad 42

Cuadro 21: Transporte..... 45

Cuadro 22: Deporte..... 45

Cuadro 23: Vocación..... 46

Cuadro 24: Año que terminó la secundaria 46

Cuadro 25: Tipo de colegio 47

Cuadro 26: Colegio 48

Cuadro 27: Preparación de ingreso a la universidad 49

Cuadro 28: Año de ingreso a la universidad 50

Cuadro 29: Método de ingreso a la universidad 51

Cuadro 30: Traslado de la universidad 51

Cuadro 31: Universidad de procedencia 52

Cuadro 32: Tabla de resumen de algoritmos procesados 73

ÍNDICE DE IMÁGENES

Pág.

Imagen 1: Diagrama de dispersión situación social 55

Imagen 2: Diagrama de dispersión situación económica 56

Imagen 3: Diagrama de dispersión situación educacional 57

Imagen 4: Diagrama de dispersión situación personal 58

Imagen 5: Software weka 59

Imagen 6: Promedio discretizado 59

Imagen 7: Situación social vs promedio 60

Imagen 8: Situación económica vs promedio 61

Imagen 9: Situación educacional vs promedio 61

Imagen 10: Situación personal vs promedio 62

Imagen 11: Algoritmo RandomTree 63

Imagen 12: Árbol de decisión de la situación social 74

Imagen 13: Árbol de decisión de la dimensión económica 75

Imagen 14: Árbol de decisión de la situación educacional 76

Imagen 15: Árbol de decisión de la dimensión personal 77

ÍNDICE DE FIGURAS

Pág.

Figura 1: Esquema de diseño de la investigación 18

Figura 2: Libro Data Mining - Weka..... 21

Figura 3: IDE Mysql Workbench..... 22

Figura 4: Recopilación de información en Microsoft Excel 23

Figura 5: Proceso de la metodología KDD 24

Figura 6: Modelo entidad – relación de la base de datos 26

Figura 7: Diagrama relacional de base de datos 27

ÍNDICE DE ANEXOS

Pág.

Anexo 1: Matriz de identificación del problema 96

Anexo 2: Hoja de ruta de las variables..... 97

Anexo 3: Matriz de operacionalización de las variables..... 98

Anexo 4: Matriz de consistencia sistémica..... 99

Anexo 5: Cuestionario de preguntas de recolección de datos 100

Anexo 6: Validación de los instrumentos por expertos.....103

Anexo 7: Base de datos recolectada.....105

Anexo 8: Documento de autorización de tratamiento de datos de la UNU ...106

Anexo 9: Constancia por parte de la Escuela Académica Profesional de Ingeniería de Sistemas.....107

RESUMEN

Esta investigación se llevó a cabo para desarrollar un modelo predictivo aplicando técnicas de minería de datos para predecir el desempeño académico de los alumnos cursantes de la Carrera Profesional de Ingeniería de Sistemas de la Universidad Nacional de Ucayali de Pucallpa (UNU), con información académica desde el año 2015 al 2017. Para la elaboración de este proyecto se optó por usar la metodología KDD que dispone el proceso de minería de datos en cinco fases consecutivas; y WEKA como herramienta software para el tratamiento de los datos y comparativa de resultados en el proceso de la metodología. Se evaluó con el modelo de árbol de decisión (RandomTree) con regresión logística, teniendo en cuenta cuatro dimensiones: situación social, situación económica, situación personal y situación académica del alumno. La veracidad de los modelos es medida a partir de la información que brindó la oficina de dirección de escuela de la Carrera Profesional de Ingeniería de Sistemas y de las encuestas brindadas a los alumnos matriculados en la carrera, en el año 2018.

Palabras claves: Desempeño estudiantil, minería de datos, modelo predictivo, árbol de decisión.

ABSTRACT

This research was carried out to develop a predictive model applying data mining techniques to predict the academic performance of students in the systems engineering career at the National University of Ucayali in Pucallpa (UNU), with academic information from 2015 to 2017. For the elaboration of this project, the KDD methodology was chosen, which provides the data mining process in five consecutive phases. And WEKA as a software tool for data processing and comparison of results in the process of the methodology. It was evaluated with the decision tree model (RandomTree) with logistic regression, taking into account four dimensions: social situation, economic situation, personal situation and academic situation of the student. The veracity of the models is measured based on the information provided by the school's management office of the systems engineering career and on the surveys provided to the students enrolled in the career in 2018.

Keywords: Student performance, data mining, predictive model, decision tree.

INTRODUCCIÓN

El rendimiento académico de la población universitaria forma un importante agente sobre el tema de la calidad educativa, técnicamente es un tema mundial, en todas las universidades existe este fenómeno, que a lo largo del tiempo se ha convertido en un tema considerable de estudio. Hipotéticamente existe relación entre los medios sociales, psicológicos, económicos con el bajo rendimiento estudiantil que a diario toca vivir a millones de estudiantes universitarios en todo el mundo, bajo esta premisa, en este estudio se busca explicar las causas de este fenómeno utilizando una rama de la ciencia de la computación.

Cada ciclo académico, la escuela académica profesional de ingeniería de sistemas de la Universidad Nacional de Ucayali enfrenta a un gran porcentaje de irregularidad estudiantil, esto debido a diversos factores psicopedagógicos, económicos y sociales a los que el alumno enfrenta.

Por lo tanto, en el siguiente proyecto de tesis se recolectó datos sociales, económicos, personales y académicos de los alumnos de la carrera profesional de Ingeniería de Sistemas, a través de una encuesta con 30 ítems (elaboración propia), para determinar el desempeño estudiantil mediante un modelo predictivo utilizando técnicas de minería de datos.

El siguiente plan de tesis se enfoca en cinco capítulos:

Capítulo I. Planteamiento del problema, en este capítulo se describe los antecedentes del problema, la formulación del problema, los objetivos de la investigación, justificación e importancia, la hipótesis y variables de la investigación, así mismo, las variables con su definición conceptual y operacional.

Capítulo II: Marco teórico, en este capítulo se señala los antecedentes la investigación tanto internacional como nacional y la definición de términos básicos.

Capítulo III: Metodología, se presenta en este capítulo el método de investigación, población y muestra, los procedimientos de recolección de datos, las técnicas e instrumentos para recolección de datos, procesamiento de recolección de datos y finalmente el tratamiento de datos.

Capítulo IV: Resultados en este capítulo se desarrolla la metodología KDD, en sus cinco fases, para finalmente interpretar la data procesada.

Capítulo V: Discusión, en este capítulo se muestra el impacto de otras investigaciones sobre los resultados obtenidos en el presente estudio.

Finalmente se presenta las conclusiones, recomendaciones, referencias bibliográficas y los anexos correspondientes.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1. ANTECEDENTES DEL PROBLEMA

Entre los problemas más dificultosos y común que enfrentan las instituciones de estudios superiores en el mundo, son: el bajo rendimiento y la irregularidad estudiantil como base en el rezago, lo que da lugar a la deserción universitaria, afectando tanto el nivel de calidad y eficiencia educativa. En una investigación de (Camborda, 2014), procedente de la Universidad Continental, menciona que un 27% de los alumnos tienen un bajo desempeño estudiantil por motivos económicos, un 32% fracasan académicamente por motivos personales, un 52% de los estudiantes tienen un bajo desempeño estudiantil debido al poco tiempo que dedican a estudiar, un 43% de los alumnos tienen padres que culminaron la educación secundaria, un 69% de los mencionados trabajan y estudian, un 96% de estudiantes estaban solteros y un 2% ya contraían matrimonio, un 70% de los alumnos eran de sexo masculino y un 29% de ellos eran féminas, y todos estos factores socioeconómicos, personales y pedagógicos conllevan al fracaso académico. Entonces podemos afirmar que existe una relación invulnerable entre el rendimiento académico y el concluir los estudios satisfactoriamente en la universidad, esto a través de un determinado tiempo, ya que los obstáculos sociales, económicos y psicológicos toman un punto importante en la relación ya mencionada.

Desde hace unas décadas atrás el concepto y su representación de irregularidad estudiantil ya no era novedad en las universidades, incluso se

creía que mientras más alumnos desaprobaban los cursos, estaba bien, puesto que era el efecto del buen nivel "educativo" o en otros términos exigencia por parte de la institución, sin comprender que en realidad era un problema que se debía batallar.

De este modo es necesario identificar a los alumnos irregulares, (Reglamento Académico UNU, 2017), son aquellos que están matriculados por condición, los que tienen desaprobados una o más materias, los observados, los que se matriculan con menos de doce créditos.

La irregularidad estudiantil es un problema complejo e incluye diversas causas (pedagógicos, psicopedagógicos, sociales y económicos). La presente intenta conocer y explicar las razones del fenómeno de la irregularidad estudiantil en la escuela académica profesional de ingeniería de sistemas de la Universidad Nacional de Ucayali. De tal manera que a través de un modelo predictivo utilizando técnicas de minería de datos se pueda predecir la irregularidad estudiantil universitaria.

EN LA CARRERA PROFESIONAL DE INGENIERÍA DE SISTEMAS

- Un gran porcentaje del alumnado de la carrera profesional en mención, trabajan para su solvencia económica, siendo este un factor de desequilibrio en el rendimiento académico, puesto que físicamente se sobrecarga, con las tareas a su cargo.

- El bajo nivel de enseñanza que recibe el alumnado en los colegios, es una de las principales causas de la irregularidad estudiantil, puesto que los niveles de enseñanzas no tienen paridad en un primer

momento.

- En muchos casos ingresan a la carrera mencionada por segunda opción de admisión, o por presión de los padres a estudiar una carrera, o quizá porque un amigo ingresó a esa carrera y le animó a seguir esta, en todos los casos mencionados y los existentes en la realidad, hace que no se tenga vocación en la carrera, y tras ello la falta de interés en su formación como profesional de calidad, dando pie a la irregularidad estudiantil.

1.2. FORMULACIÓN DEL PROBLEMA

1.2.1. Problema General

- ¿De qué manera un modelo predictivo aplicando técnicas de minería de datos determina la irregularidad estudiantil en la Escuela Académica Profesional de Ingeniería de Sistemas de la UNU?

1.2.2. Problemas Específicos

- ¿Cuál es la estimación de la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación social del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU?
- ¿Cuál es la estimación de la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación económica del alumno de la E.A.P. de

Ingeniería de Sistemas de la UNU?

- ¿Cuál es la estimación de la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación personal del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU?

- ¿Cuál es la estimación de la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación educacional del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU?

1.3. OBJETIVOS DE LA INVESTIGACIÓN

1.3.1. Objetivo General

- Establecer un modelo predictivo aplicando técnicas de minería de datos para determinar la irregularidad estudiantil en la escuela académica profesional de Ingeniería de Sistemas de la UNU.

1.3.2. Objetivos Específicos

- Estimar la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación social del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

- Definir la irregularidad estudiantil empleando modelo predictivo aplicando técnicas de minería de datos considerando la situación económica del alumno de la E.A.P. de Ingeniería de Sistemas de la

UNU.

- Precisar la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación personal del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

- Especificar la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación educacional del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

1.4. JUSTIFICACIÓN E IMPORTANCIA

1.4.1. Justificación

Es desbordante los índices de irregularidad estudiantil en la carrera profesional de ingeniería de sistemas de la UNU, casi no nos encontramos con alumnos que no hayan desaprobado un curso por lo menos en un ciclo de estudios, esto motivó y dio pie para empezar este estudio e investigar las causas de este fenómeno, que poco a poco va b sumando alumnado ciclo tras ciclo, no sé sabe a ciencia cierta los motivos por el cual ocurre este problema. Por consiguiente, este estudio ayudará a comprender las causas de este fenómeno.

1.4.2. Importancia

La minería de datos se adapta a la necesidad de encontrar la causa probable de la irregularidad estudiantil universitaria, a través de patrones ocultos por los grandes volúmenes de datos, que podemos obtener de los alumnos, tanto en su situación pedagógica, psicopedagógica, sociales y económicos. Esto dará lugar a cuantificar y medir el problema ya mencionado.

1.5. HIPÓTESIS Y/O SISTEMA DE HIPÓTESIS

1.5.1. Hipótesis General

- El modelo predictivo aplicando técnicas de minería de datos determina la irregularidad estudiantil en la escuela académica profesional de Ingeniería de Sistemas de la UNU.

1.5.2. Hipótesis Específicas

- La estimación de la irregularidad estudiantil es significativa empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación social del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.
- La estimación de la irregularidad estudiantil es significativa empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación económica del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

- La estimación de la irregularidad estudiantil es significativa empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación personal del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

- La estimación de la irregularidad estudiantil es significativa empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación educacional del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

1.6. VARIABLES (DEFINICIÓN CONCEPTUAL Y OPERACIONAL)

Cuadro 1: Definición conceptual y operacional de las variables.

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES
MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS	Arimetrics (2019). Es el agrupamiento de funciones matemáticas y de la estadística inferencial para predecir resultados.	El modelo predictivo para este estudio comprende estas cinco dimensiones: selección de datos, pre-procesamiento, transformación de datos, minería de datos e interpretación.	Selección de datos	• Analizar los datos obtenidos.
			Pre - procesamiento	• Preparación y limpieza de datos.
			Transformación de datos	• Tratamiento dimensional de los datos.
			Minería de datos	• Extracción de patrones de datos.
			Interpretación	• Interpretación de los datos.
IRREGULARIDAD ESTUDIANTIL	Reglamento Académico UNU (2017). Referente a los que están matriculados por condición, los que tienen desaprobados una o más materias, los observados, los que se matriculan con menos de doce créditos.	La irregularidad estudiantil para este estudio comprende de cuatro dimensiones: situación social, situación económica, situación personal y situación educacional del alumno.	Situación social	• Entorno físico vivienda • Lugar vivienda • Zona vivienda • Material de vivienda • Tiempo que radica en Pucallpa • Número de hijos • Edad de hijos • Estado civil
			Situación económica	• Condición laboral • Dependencia económica • Cuánto dinero percibe mensual • Nivel educativo padre y madre • Trabajo padre y madre
			Situación personal	• Edad • Sexo • Horas que dedica a estudiar • Días que estudia fuera de la universidad • Transporte • Deporte que practica • Vocación
			Situación educacional	• Año que termino la secundaria • Tipo de colegio • Nombre de colegio • Preparación para ingreso a la universidad • Año de ingreso a la universidad • Método de ingreso a la universidad • Realizó traslado de otra universidad • Universidad de procedencia

CAPÍTULO II

MARCO TEÓRICO

2.1. ANTECEDENTES DE LA INVESTIGACIÓN

2.1.1. Tesis Internacionales

Mohammad (2018), “Measuring Academic Performance of Students in Higher Education Using Data Mining Techniques”. Tesis de posgrado, Mohammad plantea la EDM (Educational Data Mining) para explicar en medida el rendimiento académico con diferentes factores en estudio, principalmente la asistencia y la nota promedio del semestre del alumno, para ello se entrevistaron a 230000 alumnos. Se obtuvo comparativas con los algoritmos KNN y Artificial Neural Network, obteniendo mejores resultados con el algoritmo de KNN.

Whitley (2018), “Educational data mining and its uses to predict the most prosperous learning environment”. Tesis de posgrado, en esta investigación realizada en EE.UU. Whitley afirma que el estudio en un área determina conformada por centros de instituciones académicas, pueden explicar mejor el rendimiento académico, puesto que nos da una mayor visión de enfoque para la predicción de rendimientos académicos, y se logró a través de algoritmos de clasificación, como el de árboles de decisiones (C4.5) con un 70% de exactitud en los algoritmos.

Bugueño (2017). “Modelo predictivo para la selección de postulantes destacados a una institución de educación superior”. Tesis de posgrado, Bugueño en su proyecto de tesis indica la necesidad de identificar a

alumnos con capacidades potenciales para que en un futuro no ingresen al porcentaje de deserción estudiantil en su país. Para ello toma como metodología en minería de datos el proceso KDD, creando así un modelo predictivo con un porcentaje de 89.4% de exactitud en el algoritmo. Se tomó en comparativa los algoritmos de Bagging y árboles de decisiones. Obteniendo mejores resultados con el algoritmo de árbol de decisión.

Vásquez (2016), “Modelo predictivo para estimar la deserción de estudiantes en una institución de educación superior”. Tesis de posgrado, en la investigación mencionada se sustenta la obtención de factores influyentes en la deserción estudiantil en la facultad de economía y finanzas de la Universidad de Chile, las variables de estudio fueron: ambientales, académicas y antecedentes. Siendo un atributo influyente en la predicción el régimen del colegio (femenino, masculino o mixto). Para ello se creó seis modelos predictivos para el semestre académico, obteniendo mejores resultados con el algoritmo de árbol de decisión.

Hernández (2015). “Modelo de minería de datos para identificación de patrones que influyen en el aprovechamiento académico”. Tesis de posgrado, el tesista Hernández, argumenta que cuatro dimensiones (individuales, institucionales, académicos y socioeconómicos) pueden explicar variables de aprovechamiento estudiantil, con información desde el año 2008 hasta el 2013.

CRISP-DM fue la metodología encargada para ayudar a comprender el modelo del negocio de patrones de comportamiento, siendo el de mejor resultado para predecir el algoritmo árbol de decisión, comparado con otros como redes neuronales artificiales.

2.1.2. Tesis Nacionales

Candia (2019). **“**Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático”. Tesis de posgrado, esta investigación refiere algoritmos de machine learning, entre ellos J-48, RandomForest y KNN, obteniendo mejores resultados de modelo de predicción el algoritmo RandomForest con un 69.35%, además menciona que las variables con más efecto en el modelo predictivo son la nota de ingreso a la universidad, la modalidad de ingreso y el número de cursos en los que se matriculó.

Alania (2018). “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la facultad de ingeniería de la universidad nacional Daniel Alcides Carrión”. Tesis de posgrado, su investigación revela que tiene como objetivo predecir el número de estudiantes propensos al abandono estudiantil (deserción), además para su veracidad en predicción, utilizó el algoritmo Random Forest para su comparativa con el algoritmo J-48 (árbol de decisión).

Eiriku (2018). “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 de postgrado, Eiriku argumenta su trabajó con 1304 alumnos de primer ciclo, recopilando datos desde dimensiones como: académicos, económicos y sociales. Los algoritmos que utilizó para la medición de la exactitud del modelo predictivo, fueron support vector

machine, árbol de decisión y regresión lineal, obteniendo mejores resultados con el 82.87% de exactitud con el algoritmo de árbol de decisión.

Piscoya (2016). “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la región de Lambayeque”. Tesis de pregrado, en su investigación, Piscoya propone algoritmos de ETS y redes neuronales, y con metodología CRISP-DM. Para el tratamiento de sus datos implementó las series de tiempos para poder mostrar su evolución desde el año 2006 al 2015. Los mejores resultados para la predicción fueron para las redes neuronales autoregreseiva con un 96%.

2.2. DEFINICIÓN DE TERMINOS BÁSICOS

2.2.1. Minería de datos

Minería de datos es el proceso de extraer conocimiento importante y útil desde volúmenes grandes de bases de datos. (García y Molina, 2012)

2.2.2. Irregularidad estudiantil

Referente a los que están matriculados por condición, los que tienen desaprobados una o más materias, los observados, los que se matriculan con menos de doce créditos. (Reglamento Académico UNU, 2017)

2.2.3. Árbol de decisión

Es una forma de graficada en forma de árbol, por las ramas, de exhibir sucesos que nacen a partir de una decisión en un determinado tiempo. (gestiondeoperaciones, 2019)

2.2.4. Datos

Los datos de investigación son aquella materia recaudada en el periodo de tiempo de la realización de un estudio o investigación. (Universidad Politécnica de Cartagena, 2018)

2.2.5. Información selectiva

Según Idalberto Chiavenato, información selectiva es "el material que podemos escoger y el cual nos conviene para satisfacer nuestros propósitos de estudio". (Chavienato, 2006, p.110)

2.2.6. Prospectiva

La prospectiva contribuye herramientas útiles para la construcción de modelos predictivos a través de algoritmos matemáticos de un determinado fenómeno. (Biblioguías - Bibliotecas CEPAL, 2018)

2.2.7. Descubrimiento

En muchas ocasiones son producto de un estudio, el revelar una hipótesis que se asumió cierta, de otro modo es base para generar nuevo conocimiento. (Diccionario de Leyes, 2018)

2.2.8. Base de datos

Es el contexto que puede ser físico o lógico donde se puede almacenar información que puede útil para para nuestra conveniencia, la base de datos es fundamental hoy en día, en grandes empresas para

gestionar su información, tanto es así que se crearon empresas dedicadas a este rubro como Oracle. (Platzi, 2018)

2.2.9. Análisis

El análisis de datos es la ciencia que examina datos para un solo propósito que es demostrar conclusiones sobre la data en la que se está trabajando. El análisis de datos es usado generalmente por estudios, ejemplo de ello son las empresas que necesitan saber aún más sobre la información que manejan. Los diagnósticos de enfermedades y resultados son en base a esta ciencia. (Searchdatacenter, 2018)

2.2.10. Metodología

Es una serie y conjunto de pasos que se rigen para obtener un resultado, o un descubrimiento. Las metodologías pueden ser usadas a nivel de investigación como de producción. (Carvajal, 2018)

2.2.11. Investigación

Es una actividad metódica que tiene por finalidad generar nuevo conocimiento, expuesto claramente para solucionar problemas en la mayoría de casos, este se lleva a cabo siguiendo una metodología. (EcuRed, 2018)

2.2.12. Computación

Se refiere a la agrupación de conocimientos que hace posible el

tratamiento y gestión de la información, con el uso de tecnologías como las computadoras u ordenadores. (la revista informática, 2018)

2.2.13. Preparación de datos

Refiere a uno de los pasos de la metodología KDD en el procesamiento de la información de minería de datos, precisamente en este paso es donde se designa la calidad en datos para su posterior transformación, donde se puede escoger y liberar información ambigua. (Textos sobre la Lupa, 2018)

2.2.14. Transformación de datos

Es en esta fase donde se puede convertir los datos de un formato a otro, el proceso lo realiza un programa especial que permite su cambio para su próximo destino. (Tecnologías Información, 2018)

2.2.15. Recolección de datos

Es juntar la información que será útil para una determinada investigación. Donde se puede usar diversas técnicas y herramientas, entre ellas las más conocidas, las encuestas, entrevistas, etc. (Baustista, 2018)

2.2.16. Confiabilidad de datos

Esta menciona que tan certero son los resultados de una investigación, que tanto concuerda con la realidad, en muchos casos existen algoritmos que permiten verificar la veracidad y confiabilidad de los

datos. (Corral, 2009, p.11)

2.2.17. Agregación de datos

Es un proceso ejecutado en la minería de datos, donde se puede realizar a través de un software, cabe resaltar que también compone soluciones de inteligencia de negocios, este último está sumamente ligado a la minería de datos. (Tecnologías Información, 2018)

2.2.18. Interpretación de datos

Refiérase a deducir y aclarar los datos que finalmente nos muestra todo el proceso de minería de datos. Este por lo general es parte final de investigaciones. (Leaf group, 2018)

2.2.19. Alumnos

Es toda persona que recibe enseñanza por parte de una docente, por lo general estos pertenecen a una institución que brinda estos servicios de enseñanza. (Real Academia Española, 2018)

2.2.20. Weka

Herramienta java de aprendizaje automático creada en la Universidad de Waikato, procedente de Nueva Zelanda. En ella se puede resolver un sinfín de tareas de clasificación, clustering y regresión, y en detalle es de código abierto con licencia GNU. (Autentia, 2019)

2.2.21. Dataset

Dataset o conjunto de datos es información recabada, reunida y asociada que permite al usuario utilizarla a su conveniencia, para su tratamiento o estudio. (Microsoft, 2019)

CAPÍTULO III

METODOLOGÍA

3.1. MÉTODO DE INVESTIGACIÓN

3.1.1. Tipo de la Investigación

La investigación en mención reunió los requisitos de una investigación aplicada, para ello fue necesario los conocimientos sobre minería de datos.

3.1.2. Nivel de la Investigación

Esta investigación cumple con características de un estudio exploratorio. Puesto que buscó predecir la irregularidad estudiantil a través de cuatro dimensiones (social, económica, personal y educacional).

3.1.3. Diseño de la Investigación

Diseño no experimental: transeccional exploratorio

Se utilizó este modelo del diseño no experimental por las características de describir la variable independiente sobre la dependiente.

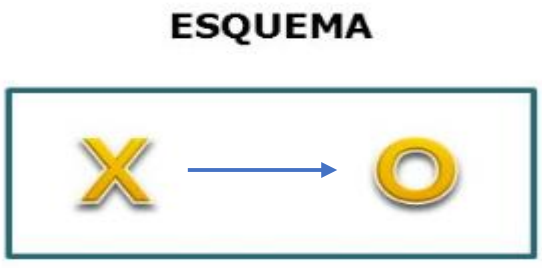


Figura 1: Esquema de diseño de la investigación

Donde:

X: Modelo predictivo aplicando técnicas de minería de datos.

O: Irregularidad estudiantil en la E.A.P. de ingeniería de sistemas de la UNU.

3.2. POBLACIÓN Y MUESTRA

3.2.1. Población

La población de estudio en esta investigación estuvo constituida por los registros académicos de los alumnos de la escuela académica profesional de Ingeniería de Sistemas de la UNU, entre los periodos 2015 y 2018.

3.2.2. Muestra

La muestra de la investigación en curso fue no probabilística-intencionada o muestreo de conveniencia, conformada por los registros académicos de 596 estudiantes de la Escuela Académica Profesional de Ingeniería de Sistemas de la UNU registrados en los periodos 2015 y 2018. Siendo esta investigación de nivel exploratorio, se consideró tomar la población y muestra en equidad de valor, referenciándose esta decisión con el apoyo del marco teórico nacional e internacional de la presente. (Eiriku, 2018)

3.3. PROCEDIMIENTO DE RECOLECCIÓN DE DATOS

Se procedió a recolectar información de los alumnos a través de una encuesta que en su contenido se extiende con preguntas cerradas y abiertas. También se contó con el apoyo de la Escuela Académica Profesional de Ingeniería de Sistemas, específicamente a la oficina de Dirección de la Escuela de Ingeniería de Sistemas, que pudo brindarme información académica de los alumnos, siendo este las notas estudiantiles. (Etzel, Walker y Stanton, 2004)

3.4. TÉCNICAS E INSTRUMENTOS PARA RECOLECCIÓN DE DATOS

3.4.1. Técnica

Encuesta: QuestionPro (2018)**.** Es una táctica de recopilación de datos para obtener información deseada, en un determinado momento y lugar**.**

3.4.2. Instrumento

Cuestionario: Meneses y Rodríguez (2019)**.** Instrumento estándar que permitió recolectar información en el trabajo de campo en un determinado estudio o investigación.

Software Weka: Weka es un software de aprendizaje automático, desarrollado por el departamento de Ciencias de la computación de la Universidad de Waikato en Nueva Zelanda, este software creado en el lenguaje Java tiene la capacidad de realizar tareas de minería de datos con

apoyo de herramientas para agrupación de clústeres, preparación de datos, predicción, regresión, clasificación, visualización y reglas de asociación.

Además, Weka es de código abierto bajo licencia GNU, cuenta con cursos en línea gratuito desde la plataforma de YouTube, en el canal con nombre **WekaMOOC**, para los interesados.

En cuanto a su interfaz intuitiva sobresale con otros proyectos de software de su misma categoría libre como R. Por este motivo Weka postula con gran capacidad para ser una de las mejores herramientas de trabajo cuando se trata de minería de datos, analizando a grandes escalas las bases de datos.

Los autores de este interesante software, también publicaron el libro “DATA MINING – Practical Machine Learning tools and techniques”, que a la actualidad está a disposición del público en su cuarta edición, en los principales sitios web de ventas confiables, como Amazon. (Machine Learning Software in Java, 2019)

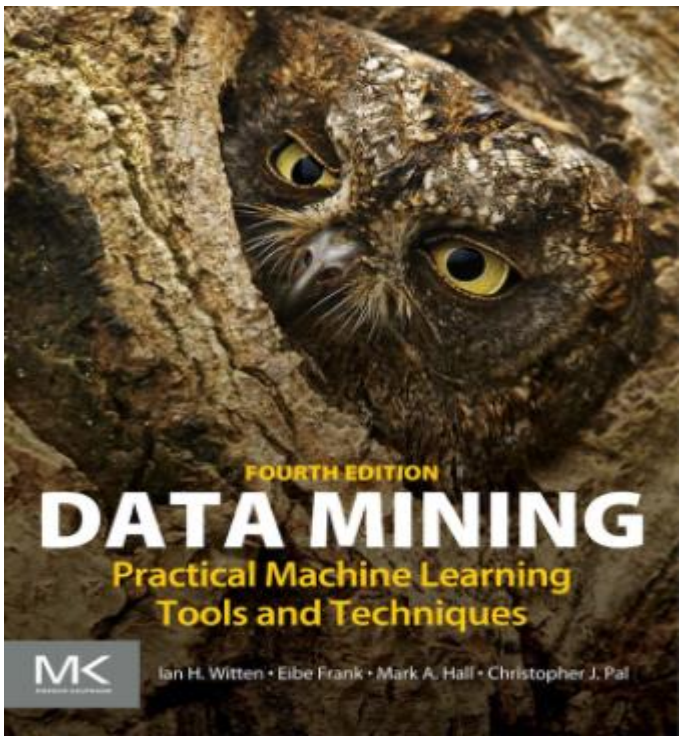


Figura 2: Libro Data Mining - Weka

3.5. PROCESAMIENTO DE RECOLECCIÓN DE DATOS

La información obtenida con la ayuda de las técnicas e instrumentos ya citados anteriormente; se integró a softwares, para su almacenamiento y su posterior manipulación. Estos fueron:

- **Microsoft Office-Excel:** Sistema informático deseable para trabajos contables, financieros y estadísticos. (conceptodefinicion, 2019)

- **Mysql Workbench:** Herramienta visual informática que permite diseñar, modelar y generar bases de datos. (UBUNLOG, 2019)

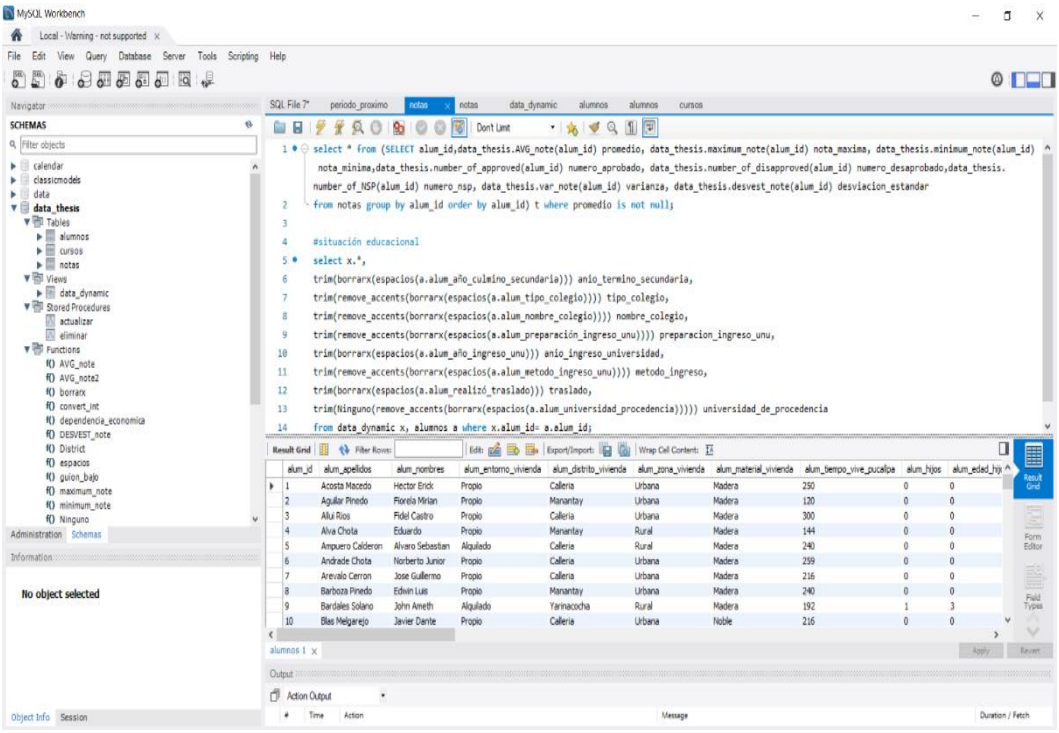


Figura 3: IDE Mysql Workbench

- **Weka:** Software que permite utilizar algoritmos de aprendizaje automático, su uso más común es para minería de datos. (Córdova, 2019)

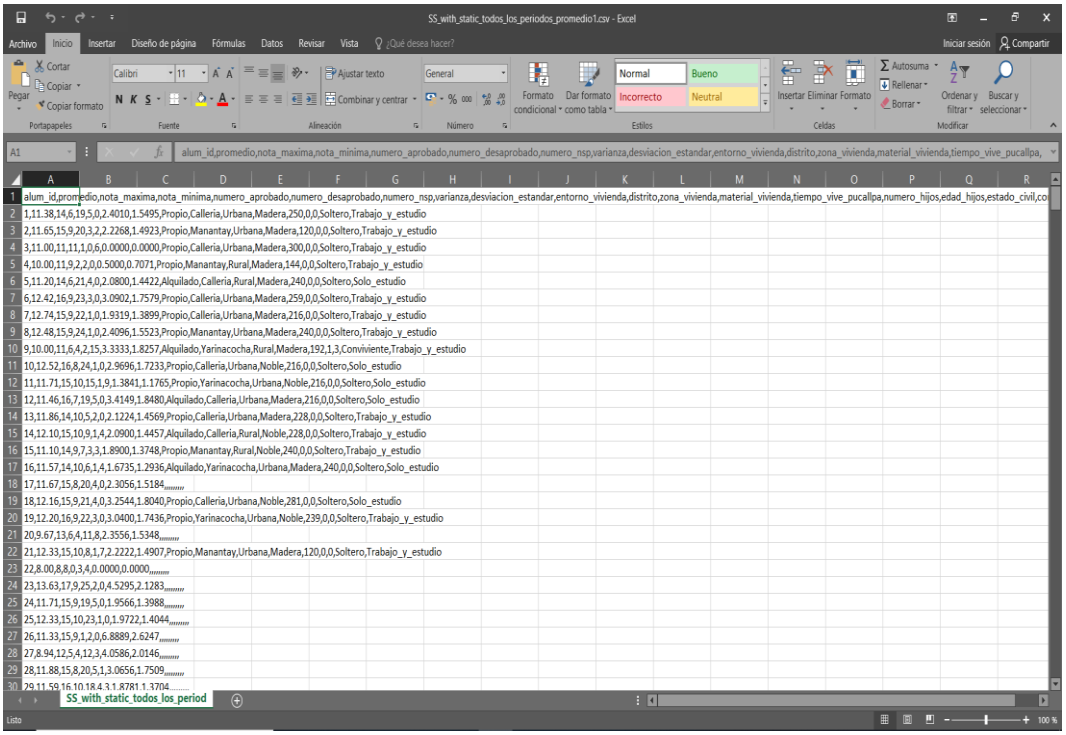


Figura 4: Recopilación de información en Microsoft Excel

3.6. TRATAMIENTO DE DATOS

Para el tratamiento de los datos se hizo uso de la metodología KDD, siendo esta una de las mejores metodologías para juntar los datos de manera entendible para la comprensión del usuario. Dividiéndose en 5 fases o etapas: (KDNUGGETS, 2019)

- **Selección de datos,** a los datos que se le manipulará.
- **Pre – procesamiento,** tratamiento de datos no correctos y ausentes.
- **Transformación de datos,** reducimos la dimensionalidad.
- **Minería de datos,** obtenemos los patrones de interés.
- **Interpretación,** del conocimiento generado.

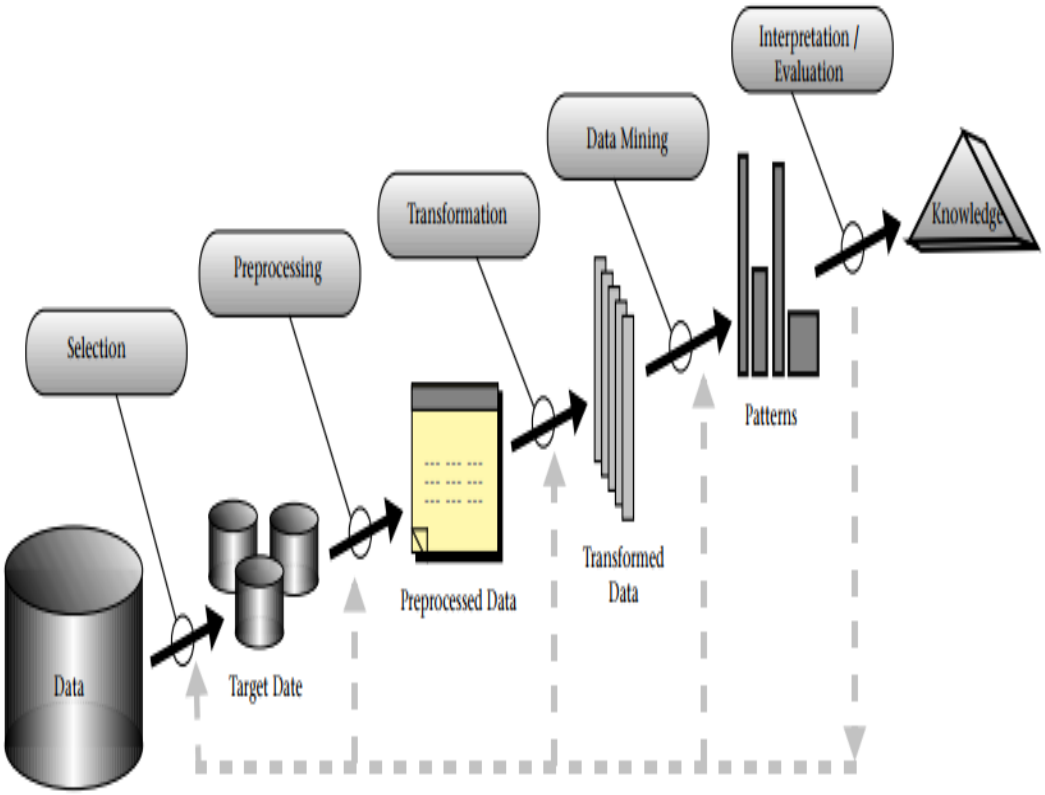


Figura 5: Proceso de la metodología KDD

CAPÍTULO IV

RESULTADOS

4.1. SELECCIÓN DE DATOS

Para la recolección o selección de datos, se obtuvo de dos fuentes, la primera fue la que fue brindada por la oficina de dirección de escuela de la Carrera Profesional de Ingeniería de Sistemas (datos académicos).

La segunda fuente fue la del instrumento del cuestionario (Ver anexo 5), el cual fue aplicado a los estudiantes de la población ya mencionada con anterioridad.

En este punto se creó una base de datos de tipo relacional, con tres tablas, con las siguientes entidades:

- Alumnos (33 campos).
- Notas (3 campos).
- Cursos (3 campos).

Se introdujo la información recolectada a la base de datos, para posteriormente exportarla a un archivo CSV, puesto que el software WEKA necesita leer archivos con este formato.

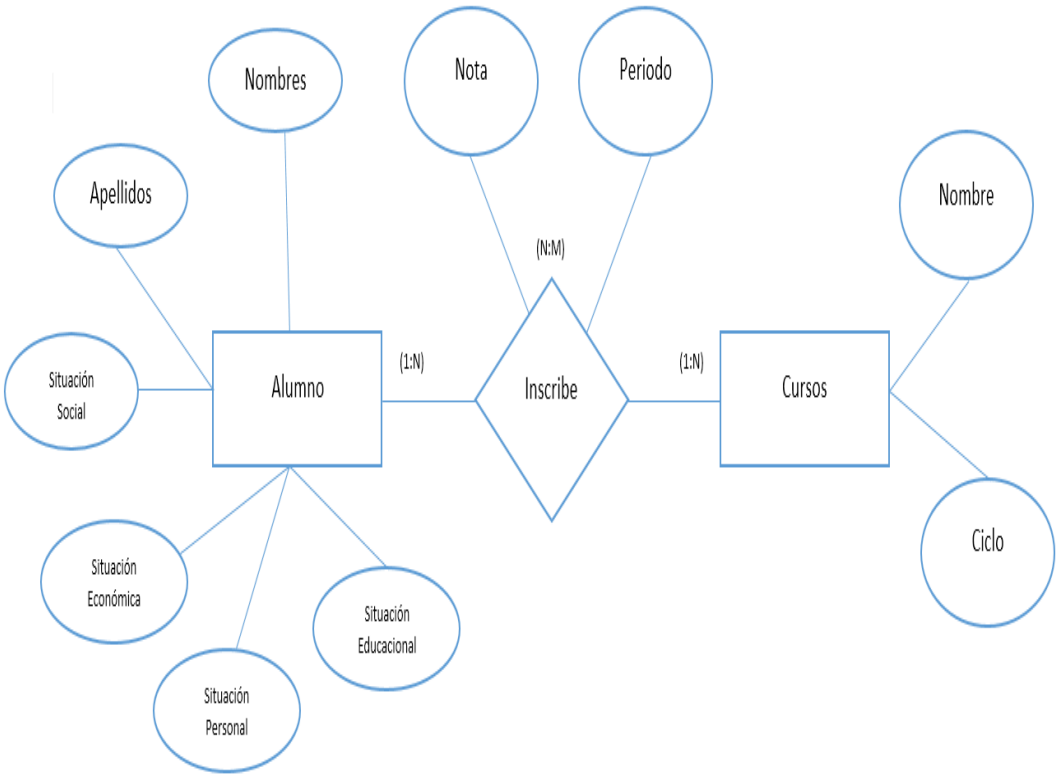


Figura 6: Modelo entidad – relación de la base de datos

Alumno (alum_id, nombres, apellidos, situación social, situación económica, situación personal, situación educacional)
Cp: alum_id

Curso (curs_id, nombre, ciclo)
Cp: curs_id

Notas (nota_id, alum_id, curs_id, nota_periodo, estado, periodo)
Cp: nota_id
Cs: alum_id, curs_id

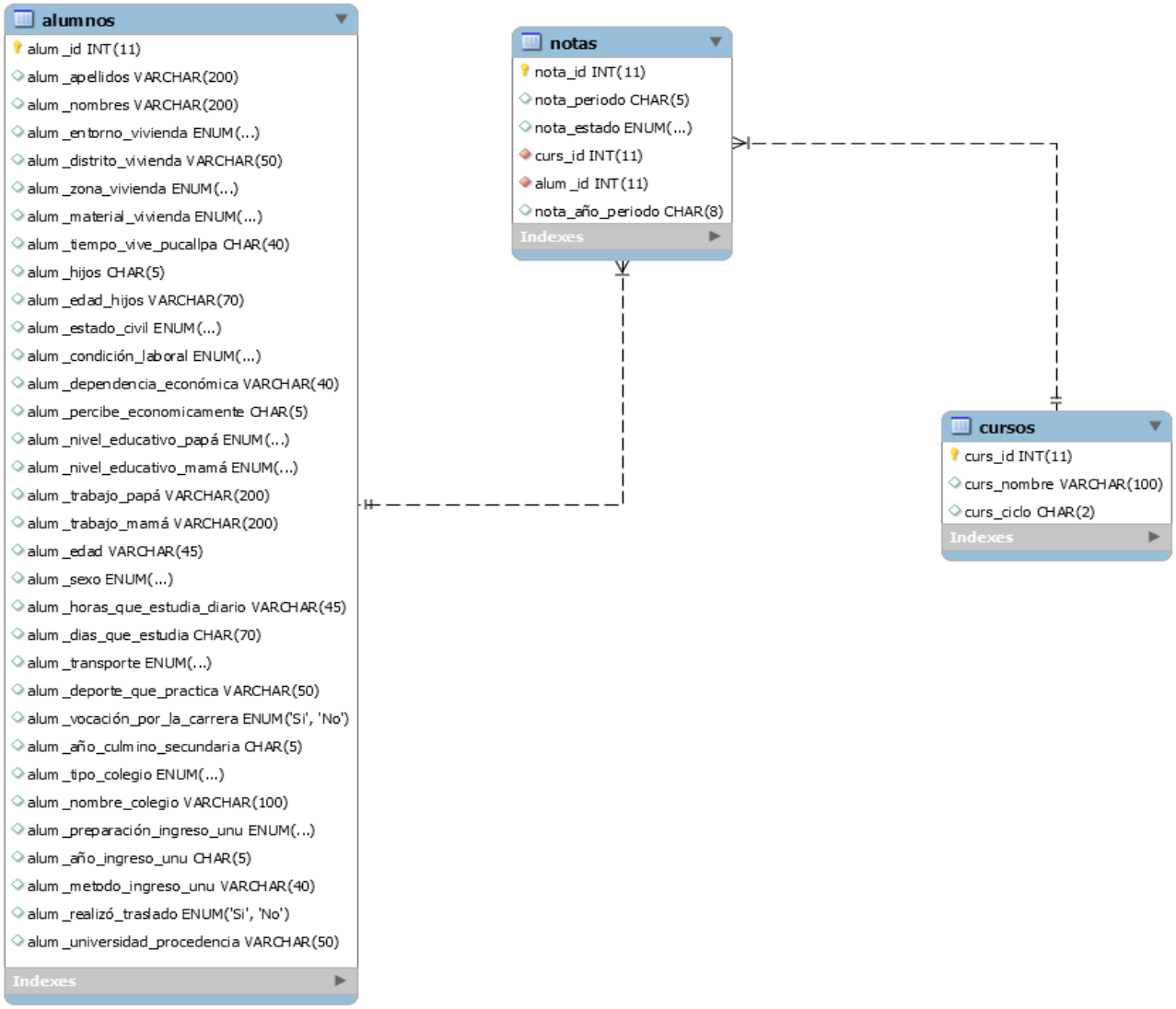


Figura 7: Diagrama relacional de base de datos

4.1.1. Interpretación de los datos de la Encuesta

1. Entorno físico de la vivienda

Cuadro 2: Entorno físico de la vivienda

Entorno	Instancias	Porcentaje
Propio	6381	95.67%
Alquilado	256	3.07%
Cedido	29	0.35%

Interpretación: La mayoría de alumnos tienen casa propia (95.67%), solo el 3.07% vive en una casa alquilada, y el 0.35% vive en una casa cedida.

2. Lugar de vivienda

Cuadro 3: Lugar de vivienda

Distrito	Instancias	Porcentaje
Yarinacocha	304	3.65%
Callería	7625	91.50%
Manantay	404	4.85%

Interpretación: Un gran porcentaje de alumnado vive en el distrito de Callería (91.50%), el 4.85% vive en Manantay y el 3.65% vive en Yarinacocha.

3. Zona de vivienda

Cuadro 4: Zona de vivienda

Zona	Instancias	Porcentaje
Urbana	7645	91.74%
Rural	397	4.76%
Céntrica	291	3.49%

Interpretación: El 91.74% de los alumnos viven en una zona urbana de la ciudad de Pucallpa, mientras que el 4.76% vive en un zona rural, y tan solo el 3.49% vive en zona céntrica.

4. Material de vivienda

Cuadro 5: Material de vivienda

Material	Instancias	Porcentaje
Madera	7408	88.89%
Noble	794	9.53%
Seminoble	111	1.33%
Estera	20	0.24%

Interpretación: El 88.89% de los alumnos viven en una casa de madera siendo este el más alto porcentaje. Un 9.53% del alumnado vive en una casa de material noble, el 1.33% vive en una casa seminoble, y por último un 0.24% de los alumnos viven en una casa esteras.

5. Tiempo que radica en Pucallpa

Cuadro 6: Tiempo que radica en Pucallpa

Tiempo	Instancias	Porcentaje
0 años	6606	79.28%
1 año	18	0.22%
2 años	19	0.23%
4 años	19	0.23%
5 años	106	1.27%
6 años	11	0.13%
7 años	19	0.23%
8 años	50	0.60%
9 años	76	0.91%
10 años	43	0.52%
11 años	30	0.36%
12 años	39	0.47%
14 años	84	1.01%
15 años	57	0.68%
16 años	18	0.22%
17 años	153	1.84%
18 años	21	0.25%
19 años	218	2.62%
20 años	191	2.29%
21 años	211	2.53%
22 años	238	2.85%
23 años	55	0.66%
24 años	27	0.32%
26 años	16	0.19%
29 años	8	0.09%

Interpretación: El 79.28% de los alumnos en mención viven menos de un año en Pucallpa, el 0.22% vive un año en Pucallpa, el 0.23% vive dos años en Pucallpa, el 0.23% vive cuatro años en Pucallpa, el 1.27% vive cinco años en Pucallpa, el 0.13% vive seis años en Pucallpa, el 0.23% vive el siete años en Pucallpa, el 0.60% vive ocho años en Pucallpa, el 0.91% vive nueve años en Pucallpa, el 0.52% vive diez años en Pucallpa, el 0.36% vive once años en Pucallpa, el 0.47% vive doce años en Pucallpa, el 1.01% vive catorce años en Pucallpa, el 0.68% vive quince años en Pucallpa, el 0.22% vive diecisiseis años en Pucallpa, el 1.84% vive diecisiete años en Pucallpa, el 0.25% vive dieciocho años en Pucallpa, el 2.62% vive diecinueve años en Pucallpa, el 2.29% vive veinte años en Pucallpa, el 2.53% vive veintiun años en Pucallpa, el 2.85% vive veintidos años en Pucallpa, el 0.66% vive veintitres años en Pucallpa, el 0.32% vive veinticuatro años en Pucallpa, el 0.19% vive veintiseis años en Pucallpa y por último el 0.09% vive veintinueve años en Pucallpa.

6. Número de hijos

Cuadro 7: Número de hijos

Número de hijos	Instancias	Porcentaje
0	8226	98.72%
1	87	1.04%
2	20	0.24%

Interpretación: El 98.72% de los estudiantes no tienen hijos, mientras que tan solo el 1.04% tiene un hijo y el 0.24% tiene dos hijos.

7. Edad de Hijos

Cuadro 8: Edad de hijos

Edad de hijos	Instancias	Porcentaje
0	8243	98.92%
3	21	0.25%
4	39	0.47%
2-4	30	0.36%

Interpretación: El 98.92% de los alumnos marcaron cero, puesto que no tienen hijos a quien categorizar sus edades, el 0.25% de alumnos marcaron tiene un hijo(a) de tres años de edad, el 0.47% tiene un hijo(a) de cuatro años, y solo el 0.36% tiene dos hijos cuyas edades son de dos y cuatro años consecutivamente.

8. Estado civil

Cuadro 9: Estado civil

Estado Civil	Instancias	Porcentaje
Soltero	8257	99.08%
Conviviente	51	0.61%
Casado	25	0.30%

Interpretación: El 99.08% de los estudiantes son solteros, el 0.61% de ellos viven con una pareja (conviviente) y el 0.30% son casados.

9. Condición laboral

Cuadro 10: Condición laboral

Condición laboral	Instancias	Porcentaje
Trabajo y estudio	7548	90.58%
Solo estudio	785	9.42%

Interpretación: El 90.58% de los alumnos trabajan y estudian, mientras que el 9.42% de ellos solo estudia.

10. Dependencia económica

Cuadro 11: Dependencia económica

Dependencia económica	Instancias	Porcentaje
Padres	7482	89.79%
Conviviente	17	0.20%
Trabajo	575	6.90%
Padres y trabajo	259	3.11%

Interpretación: El 89.79% de los alumnos dependen económicamente de sus padres, el 6.90% dependen de un trabajo, el 3.11% depende de sus padres y un trabajo y solo el 0.20% depende de una pareja (conviviente).

11. Percibe económicamente – mensual

Cuadro 12: Percibe económicamente - mensual

Soles	Instancias	Porcentaje
S/. 0	7144	85.73 %
S/. 10	20	0.24 %
S/. 150	19	0.23 %
S/. 200	55	0.66 %
S/. 250	45	0.54 %
S/. 300	47	0.56 %
S/. 320	19	0.23 %
S/. 380	26	0.31 %
S/. 400	50	0.60 %
S/. 450	11	0.13 %
S/. 500	263	3.16 %
S/. 550	26	0.31 %
S/. 600	105	1.26 %
S/. 700	62	0.74 %
S/. 750	45	0.54%
S/. 800	37	0.44%
S/. 900	62	0.74%
S/. 950	26	0.31%
S/. 1000	48	0.58%
S/. 1200	31	0.37%
S/. 1300	27	0.32%
S/. 1500	26	0.31%
S/. 1800	50	0.60%
S/. 2000	28	0.34%
S/. 3500	25	0.30%
S/. 5000	17	0.20%
S/. 7500	19	0.23%

Interpretación: El 85.73% de los alumnos no percibe un salario mensual, el 0.24% percibe 10 soles, el 0.23% percibe cientocincuenta soles, el 0.66% percibe docientos soles, el 0.54% percibe docientoscincuenta soles, el 0.56% percibe trecientos soles, el 0.23% percibe trecientos veinte soles, el 0.31% percibe trecientos ochenta soles, el 0.60% percibe cuatrocientos soles, el 0.13% percibe cuatrocientos cincuenta soles, el 3.16% percibe quiniento soles, el 0.31% percibe quinientos cincuenta soles, el 1.26% percibe seicientos soles, el 0.74% percibe setecientos soles, el 0.54% percibe setecientos cincuenta soles, el 0.44% percibe ochocientos soles, el 0.74% percibe novecientos soles, el 0.31% percibe novecientos cincuenta soles, el 0.58% percibe mil soles, el 0.37% percibe mil docientos soles, el 0.32% percibe mil trecientos soles, el 0.31% percibe mil quinientos soles, el 0.60% percibe mil ochocientos soles, el 0.34% percibe dos mil soles, el 0.30% percibe tres mil soles, el 0.20% percibe cinco mil soles, el 0.23% percibe siete mil quinientos soles.

12. Nivel educativo del padre

Cuadro 13: Nivel educativo del padre

Nivel educativo	Instancias	Porcentaje
Educación primaria	78	0.94%
Educación secundaria	7522	90.27%
Educación superior (Universidad)	377	4.52%
Educación superior (Tecnológico)	356	4.27%

Interpretación: El 90.27% de los papás de los alumnos encuestados estudiaron o culminaron sus estudios secundarios, el 4.52% estudió o culminó sus estudios en una universidad, el 4.27% estudió o culminó sus estudios en un tecnológico y el 0.94% solo estudió primaria.

13. Nivel educativo de la madre

Cuadro 14: Nivel educativo de la madre

Nivel educativo	Instancias	Porcentaje
Educación primaria	106	1.27%
Educación secundaria	7504	90.05%
Educación superior (Universidad)	363	4.36%
Educación superior (Tecnológico)	360	4.32%

Interpretación: El 90.05% de las mamás de los encuestados estudió o culminó la educación secundaria, el 4.36% estudió o culminó sus estudios en una universidad, el 4.32% estudió o culminó sus estudios en un tecnológico y el 1.27% solo estudió primaria.

14. Trabajo del padre

Cuadro 15: Trabajo del padre

Trabajo	Instancias	Porcentaje
Soldador	49	0.58%
Agricultor	52	0.62%
Mototaxista	125	1.50%
Administrador	25	0.30%
Psicólogo	24	0.28%
Empresario	89	1.07%
Operador de maquinaria pesada	26	0.31%
Ingeniero industrial	26	0.31%
Director I.E.	23	0.28%
Fontanero	25	0.30%
Asesor de crédito	28	0.34%
Médico	11	0.13%
Cerrajero	19	0.23%
Supervisor de flota	21	0.25%
Ingeniero automotriz	21	0.25%
Contador	25	0.30%
Jubilado	18	0.22%
Taxista	30	0.36%
Embotellador en ajeper	21	0.25%
Ingeniero civil	18	0.22%
Abogado	17	0.20%
Palmicultor	18	0.22%
Veterinario	20	0.24%
Militar	19	0.23%
Docente	94	1.13%
Comerciante	7090	85.08%
Independiente	71	0.85%
Maderero	80	0.96%
Mecánico	51	0.61%
Electricista	60	0.72%
Albañil	64	0.77%
Carpintero	28	0.34%
Empleado público	45	0.54%

Interpretación: El 0.58% del trabajo de los papás de los encuestados concierne a soldador, el 0.62% son agricultor, el 1.50% son mototaxistas, el 030% son administradores, el 0.28% son psicólogos, el 1.07% son empresarios, el 0.31 son operadores de maquinaria pesada, el 0.31% son ingenieros industriales, el 0.28% son directores de una institución,el 0.30% son fontaneros, el 0.34% son asesores de créditos, el 0.13% son médicos, el 0.23% son cerrajeros, el 0.25% son supervisores de flotas, el 0.25% son ingenieros automotriz, el 0.30% son contadores, el 0.22% son jubilados, el 0.36% son taxistas, el 0.25% son embotelladores en Ajeper, el 0.22% son ingenieros civiles, el 0.20% son abogados, el 0.22% son palmicultores, el 0.24% veterinarios, el 0.23% son militares, el 1.13% son docentes, 85.08% son comerciantes con el índice más alto, el 0.85% trabajan independientemente (Trabajo no estable), el 0.96% son madereros, el 0.61% son mecánicos, el 0.72% son electricistas, el 0.77% son albañiles, el 0.34% son carpinteros, el 0.54% son empleados públicos.

15. Trabajo de la madre

Cuadro 16: Trabajo de la madre

Trabajo	Instancias	Porcentaje
Camarera	4	0.05%
Promotora en educación inicial	26	0.31%
Psicóloga	25	0.30%
Sastre	25	0.30%
Ayudante en distribuidora	26	0.31%
Venta de comida	27	0.32%
Secretaria	54	0.65%
Administrativo	18	0.22%
Administradora	18	0.22%
Dentista	17	0.20%
Palmicultor	18	0.22%
Docente	327	3.92%
Independiente	56	0.67%
Empresaria	43	0.52%
Adiministradora de hospedaje	43	0.52%
Ama de casa	7284	87.41%
Comerciante	218	2.61%
Enfermera	104	1.25%

Interpretación: El 87.41% de las mamás de los alumnos son ama de casa, mientras que el 0.05% son camareras, el 0.31% son promotoras de educación inicial, el 0.30% son psicólogas, el 0.30% son sastres, el 0.31% son ayudantes en distribuidoras, el 0.32% se dedican a la venta de comida, el 0.65% son secretarias, el 0.22% son administrativas, el 0.22% son administradoras, el 0.20% son dentistas, el 0.22% se dedican a la

palmicultura, el 3.92% son docentes, el 0.67% trabajan independientemente (Trabajo no estable), el 0.52% son empresarias, el 0.52% son administradoras de hospedajes, el 2.61% son comerciantes y el 1.25% son enfermeras.

16. Edad

Cuadro 17: Edad

Edad	Instancias	Porcentaje
18 años	113	1.36%
19 años	95	1.14%
20 años	5573	66.88%
21 años	255	3.06%
22 años	227	2.72%
23 años	262	3.14%
24 años	35	0.42%
25 años	63	0.75%
26 años	26	0.31%
27 años	11	0.13%
34 años	6	0.07%

Interpretación: El mayor porcentaje en el atributo edad es de 66.88% con veinte años, en promedio hay más encuestados de veinte años para este estudio, el 1.36% tienen diesciocho años, el 1.14% tienen diecinueve años, el 3.06% tiene veintiun años, el 2.72% tienen veintidos años, el 3.14% tienen veintitres años, el 0.42% tienen veinticuatro años, el 0.75% tienen veinticinco años, el 0.31% tienen veintiseis años, el 0.13% tienen veintisiete años, el 0.07% tienen treinta y cuatro años.

17. Sexo

Cuadro 18: Sexo

Sexo	Instancias	Porcentaje
Masculino	7287	87.45%
Femenino	1046	12.55%

Interpretación: El 87.45% son de sexo masculino en los encuestados, y el 12.55% son de sexo femenino, hay más varones que mujeres en la carrera de ingeniería de sistemas.

18. Horas que dedica a estudiar al día

Cuadro 19: Horas que dedica a estudiar al día

Horas	Instancias	Porcentaje
0	29	0.35%
1	181	2.17%
2	6966	83.59%
3	325	3.90%
4	341	4.09%
5	128	1.54%
6	191	2.29%
7	51	0.61%
8	81	0.97%
9	17	0.20%
12	32	0.38%

Interpretación: El 83.59% estudia dos horas al dia, siendo el más alto índice, luego el 0.35% no estudia nada en el día, el 2.17% estudia tan solo una hora, el 3.90% estudia tres horas, el 4.09% estudia cuatro horas, el 1.54% estudia cinco horas, el 2.29% seis horas, el 0.61% estudia siete horas, el 0.97% estudia ocho horas, el 0.20% estudia nueve horas y el 0.38% estudia doce horas al día.

19. Días que estudia fuera de la universidad

Cuadro 20: Días que estudia fuera de la universidad

Días	Instancias	Porcentaje
Miércoles - Viernes - Domingo	41	0.49%
Martes - Jueves	34	0.41%
Viernes – Sábado - Domingo	77	0.93%
Lunes – Miércoles – Viernes	47	0.56%
Lunes – Viernes	25	0.30%
Miércoles - Sábado	21	0.25%
Jueves – Viernes	25	0.30%
Lunes – Viernes – Domingo	25	0.30%
Martes – Miércoles – Jueves – Viernes	24	0.28%
Jueves	13	0.16%
Miércoles – Viernes	11	0.13%
Lunes – Martes – Miércoles – Jueves – Viernes – Sábado	46	0.55%
Lunes – Martes – Miércoles – Domingo	25	0.30%
Sábado	77	0.93%
Lunes – Martes – Miércoles – Jueves – Viernes	150	1.80%
Lunes – Martes – Miércoles – Jueves – Domingo	26	0.31%
Jueves – Viernes – Sábado – Domingo	25	0.30%
Lunes – Martes – Viernes – Sábado	23	0.28%

Martes – Miércoles – Domingo	25	0.30%
Martes – Viernes	28	0.34%
Domingo	183	2.20%
Viernes - Sábado	43	0.56%
Martes – Sábado – Domingo	23	0.28%
Lunes – Domingo	17	0.20%
Ninguno	23	0.28%
Martes – Jueves - Sábado	38	0.46%
Lunes – Martes – Miércoles – Viernes	37	0.44%
Lunes	5	0.06%
Lunes – Miércoles – Viernes – Sábado - Domingo	44	0.53%
Lunes – Miércoles	16	0.19%
Lunes – Miércoles – Sábado	17	0.20%
Lunes – Miércoles – Viernes – Sábado	18	0.22%
Lunes – Miércoles – Domingo	44	0.53%
Sábado – Domingo	159	1.91%
Todos los días	6898	82.78%

Interpretación: El 82.78% de los alumnos estudian todos los días, siendo este el índice más alto, el 0.49% solo estudian los días miércoles - viernes – domingo, el 0.41% solo estudian los días martes – jueves, el 0.93% solo estudian los días viernes – sábado – domingo, el 0.56% solo estudian los días lunes – miércoles – viernes, el 0.30% solo estudian los días lunes – viernes, el 0.25% solo estudian los días miércoles – sábado, el 0.30% solo estudian los días jueves – viernes, el 0.30% solo estudian los días lunes – viernes – domingo, el 0.28% solo estudian los días martes – miércoles – jueves – viernes, el 0.16% solo estudia el día jueves, el 0.13% solo estudia los días miércoles – viernes, el 0.55% solo estudia los días lunes – martes – miércoles – jueves – viernes – sábado, el 0.30% solo estudia los días

lunes – martes – miércoles – domingo, el 0.93% solo estudia el día sábado, el 1.80% solo estudia los días lunes – martes – miércoles – jueves – viernes, el 0.31% solo estudia los días Lunes – martes – miércoles – jueves – domingo, el 0.30% solo estudia los días jueves – viernes – sábado – domingo, el 0.28% solo estudia los días lunes – martes – viernes – sábado, el 0.30% solo estudia los días martes – miércoles – domingo, el 0.34% solo estudia los días martes – viernes, el 2.20% solo estudia el día domingo, el 0.56% estudia los días viernes – sábado, el 0.28% solo estudia los días martes – sábado – domingo, el 0.20% solo estudia el lunes – domingo, el 0.28% no estudia ningún día, el 0.46% solo estudia los días martes – jueves – sábado, el 0.44% solo estudia los días lunes – martes – miércoles – viernes, el 0.06% solo estudia el día lunes, el 0.53% solo estudia los días lunes – miércoles – viernes – sábado – domingo, el 0.19% solo estudia los días lunes – miércoles, el 0.20% solo estudia los días lunes – miércoles – sábado, el 0.22% solo estudia los lunes – miércoles – viernes – sábado, el 0.53% solo estudia los días lunes – miércoles – domingo, el 1.91% solo estudia los días sábado – domingo.

20. Transporte

Cuadro 21: Transporte

Transporte	Instancias	Porcentaje
Propio	7921	95.06%
Servicio público	412	4.94%

Interpretación: El 95.06% de los estudiantes encuestados tienen movilidad propia, el 4.49% se moviliza en transporte público.

21. Deporte

Cuadro 22: Deporte

Deporte	Instancias	Porcentaje
Voley	120	1.44%
Fútbol	7271	87.26%
Basquetbol	165	1.98%
Basquetbol - Ajedrez	26	0.31%
Ajedrez	250	3.00%
Voley – Basquetbol	25	0.30%
Fútbol – Voley – Ajedrez	7	0.08%
Fútbol – Voley	17	0.20%
Fútbol – Basquetbol - Ajedrez	21	0.25%
Fútbol – Basquetol – Taekwondo	13	0.16%
Taekwondo	136	1.63%
Ninguno	282	3.38%

Interpretación: El deporte más practicado por los alumnos encuestados es el fútbol (87.26%), el 1.44% solo juega voley, el 1.98% juega solo basquetbol, el 0.31% juega basquetbol y ajedrez, el 3% juega ajedrez, el

0.30% juega voley y basquetbol a su vez, el 0.08% juega fútbol, voley y ajedrez, el 0.20% juega fútbol y voley, el 0.25% juega fútbol, basquetbol y ajedrez, el 0.16% juega fútbol, basquetbol y taekwondo, solo el 1.63% practica taekwondo y el 3.38% no practica ningun deporte.

22. Vocación

Cuadro 23: Vocación

Vocación	Instancias	Porcentaje
Si	8255	99.06%
No	78	0.94%

Interpretación: El 99.06% de los alumnos tienen vocación por la carrera profesional de ingenieria de sistemas, y solo un 0.94% de ellos no lo tiene.

23. Año que terminó la secundaria

Cuadro 24: Año que terminó la secundaria

Año	Instancias	Porcentaje
2001	8	0.10%
2007	27	0.32%
2008	16	0.19%
2009	21	0.25%
2010	112	1.34%
2011	104	1.25%
2012	7146	85.76%
2013	375	4.50%
2014	280	3.36%
2015	59	0.71%
2016	185	2.22%

Interpretación: El 85.76% de los encuestados ingresaron en el año 2012 a la universidad, siendo este el índice más alto, el 0.10% de los alumnos ingresaron en el año 2001, el 0.32% de los alumnos ingresaron en el año 2007, el 0.19% ingresaron en el año 2008, el 0.19% de los alumnos ingresaron en el año 2008, el 0.25% de los alumnos ingresaron en el año 2009, el 1.34% de los alumnos ingresaron en el año 2010, el 1.25% de los alumnos ingresaron en el año 2011, el 4.50% de los alumnos ingresaron en el año 2013, el 3.36% de los alumnos ingresaron en el año 2014, el 0.71% de los alumnos ingresaron en el año 2015, el 2.22% de los alumnos ingresaron en el año 2016.

24. Tipo de colegio

Cuadro 25: Tipo de colegio

Tipo de colegio	Instancias	Porcentaje
Estatal	8077	96.93%
Particular	256	3.07%

Interpretación: El 96.93% de los alumnos estudió en un colegio estatal, mientras que el 3.07% de ellos estudio en un colegio particular.

25. Colegio

Cuadro 26: Colegio

Colegio	Instancias	Porcentaje
I.E. Oswaldo	7	0.08%
I.E. Jorge Chavez	64	0.77%
I.E. Coronel Pedro Portillo	275	3.30%
I.E. Victor Manuel Begazo	25	0.30%
I.E. Ingenieros UNI	26	0.31%
I.E. Miguel Grau Seminario	25	0.30%
I.E. Trebol	72	0.86%
I.E. Santiago Antunes de Mayolo	11	0.13%
I.E. Victor Andrés Belaunde	17	0.20%
I.E. José Carlos Mariátegui	20	0.24%
I.E. Albert Einstein	21	0.25%
I.E. Margarita Aurora Arista	14	0.17%
I.E. Colegio Nacional de Iquitos	19	0.23%
I.E. Hossana	21	0.25%
I.E. Ricardo Bentín Grande	16	0.19%
I.E. Ciencios	17	0.20%
Colegio Particular Peruano Americano	24	0.28%
I.E. La Florida	30	0.36%
I.E. Nuestra Sra. De las Mercedes	16	0.19%
I.E. Monte Alegre Neshuya	18	0.22%
I.E. Faustino Maldonado	19	0.23%
I.E. Comercio	175	2.10%
I.E. Inmaculada	6966	83.60%
I.E. Don Bosco	55	0.66%
I.E. Horacio Zevallos Gamez	62	0.74%
I.E. Marko Jara	81	0.97%
I.E. Nuestra sra. De Guadalupe	145	1.74%
I.E. Sollertia	53	0.64%
I.E. Colegio Particular Adventista Ucayali	39	0.49%

Interpretación: El 83.60% de los alumnos encuestados estudiaron en el colegio estatal Inmaculada, siendo este el ítem con mayor porcentaje, luego el 0.08% estudió en el colegio Oswaldo, el 0.77% estudió en el colegio Jorge Chavez, el 3.30% en el colegio Pedro Portillo, el 0.30% en el colegio Victor Manuel Begazo, el 0.31% en el colegio Ingenieros UNI, el 0.30% en el colegio Miguel Grau Seminario, en el 0.86% en el colegio Trebol, el 0.13% en el colegio Santiago Antunes de Mayolo, el 0.20% en el colegio Victor Andrés Belaunde, el 0.24% en el colegio José Carlos Mariátegui, el 0.25% en el colegio Albert Einstein, el 0.17% en el colegio Margarita Aurora Arista, el 0.23% en el colegio Colegio Nacional de Iquitos, el 0.25% en el colegio Hossana, el 0.19% en el colegio Ricardo Bentín Grande, el 0.20% en el colegio Ciencios, el 0.28% en el colegio Colegio Particular Peruano Americano, el 0.36% en el colegio la Florida, el 0.19% en el colegio Nuestra Sra. de las Mercedes, el 0.22% en el colegio Monte Alegre Neshuya, el 0.23% en el colegio Faustino Maldonado, el 2.10% en el colegio Comercio, el 0.66% en el colegio Don Bosco, el 0.74% en el colegio Horacio Zevallos Gamez, el 0.97% en el colegio Marko Jara, el 1.74% en el colegio Nuestra Sra. de Guadalupe, el 0.64% en el colegio Sollertia y el 0.49% en el Colegio Particular Adventista Ucayali.

26. Preparación de ingreso a la universidad

Cuadro 27: Preparación de ingreso a la universidad

Preparación	Instancias	Porcentaje
Cepreunu	7625	91.50%
Autoestudio	708	8.50%

Interpretación: El 91.50% de los encuestados se prepararon en la Cepreunu para ingreso a la UNU, el 8.50% de ellos estudió por su cuenta.

27. Año de ingreso a la universidad

Cuadro 28: Año de ingreso a la universidad

Año	Instancias	Porcentaje
2004	8	0.10%
2010	11	0.13%
2011	63	0.76%
2012	60	0.72%
2013	7083	84.99%
2014	396	4.75%
2015	399	4.78%
2016	160	1.92%
2017	153	1.84%

Interpretación: La mayoría de estudiantes encuestados ingresaron en el año 2013 (84.99%), el 0.10% ingresó en el año 2004, el 0.13% ingresó en el año 2010, el 0.76% ingresó en el año 2011, el 0.72% inrgesó en el año 2012, el 4.75% ingresó en el año 2014, el 4.78% ingresó en el año 2015, el 1.92% ingresó en el año 2016, el 1.84% ingresó en el año 2017.

28. Método ingreso a la universidad

Cuadro 29: Método de ingreso a la universidad

Método Ingreso	Instancias	Porcentaje
Deportista Calificado	24	0.29%
Cepreunu	7471	89.65%
Examen de Admisión	767	9.20%
Exoneración	71	0.85%

Interpretación: La gran mayoria de alumnos ingresó a la universidad por la Cepreunu (89.65%), el 9.20% de alumnos ingresó por examen de admisión, el 0.85% de alumnos ingresó por exoneración y el 0.29% de alumnos ingresó por ser deportista calificado.

29. Traslado de la universidad

Cuadro 30: Traslado de la universidad

Traslado	Instancias	Porcentaje
Si	8314	99.77%
No	19	0.23%

Interpretación: El 99.77% de los alumnos no hicieron traslado de otra universidad a la UNU, mientras que el 0.23% si realizó traslado de otra universidad.

30. Universidad de procedencia

Cuadro 31: Universidad de procedencia

Universidad	Instancias	Porcentaje
Universidad Nacional de la Amazonía Peruana	19	0.23%
Ninguno	8314	99.77%

Interpretación: El mayor porcentaje para este ítem es Ninguno con 99.77%, puesto que no hay otra universidad de procedencia más que la Universidad Nacional de la Amazonía Peruana (0.23%)

4.2. PRE – PROCESAMIENTO DE DATOS

Se crea una instancia de serie de tiempo por cada alumno en la data obtenida desde los años 2015, 2016 y 2017, repartiéndose en dos partes la data estática y la data dinámica.

Para la data dinámica se tomó en cuenta la nota más alta, la nota más baja, el número de NSP, el promedio, la varianza y la desviación estándar. Para la data estática se dividió en 4 partes o dimensiones siendo estos los siguientes: (Ver anexo 5).

1. Situación social.

- Entorno de vivienda.
- Distrito.
- Zona de vivienda.
- Tiempo que radica en Pucallpa.
- Número de hijos.

- Edad de hijos.
- Estado civil.

2. Situación económica

- Condición laboral.
- Dependencia económica.
- Dinero que percibe mensual.
- Nivel educativo del padre.
- Nivel educativo de la madre.
- Trabajo del padre.
- Trabajo de la madre.

3. Situación personal

- Edad.
- Sexo.
- Horas de estudio.
- Días de estudio.
- Transporte.
- Deporte.
- Vocación.

4. Situación educacional.

- Año que culminó la secundaria.
- Tipo de colegio.
- Nombre de colegio.
- Preparación para ingreso a la UNU.

- Año de ingreso a la UNU.
- Modalidad de ingreso a la UNU.
- Traslado universitario.
- Universidad de procedencia.

La creación de las consultas en la base de datos, era conveniente en este paso de la metodología, de este modo se creó las funciones a nivel de base de datos para ayudar al pre-procesado de datos, con la finalidad de quitar los espacios entre cada contenido de los atributos, categorizar la información, puesto que había datos con similitud de escritura, pero no iguales, obtener la nota más alta, la nota más baja, el promedio, el número de NSP, la variación y la desviación estándar.

Se tomó estas medidas, puesto que sin estos pre requisitos mencionados el software WEKA no reconoce correctamente el archivo de formato CSV que se genera a partir del IDE workbench MYSQL.

Weka, nos permite visualizar el análisis de la data a través de la pestaña visualice con diagramas de dispersión, esto antes de que la data sea tratada o manipulada por los algoritmos de aprendizaje automático. A continuación, un versus entre el atributo más destacado de cada dimensión con el promedio de notas, siendo este último atributo binario, aprobado o desaprobado.

4.2.1. Mejor atributo de la Situación Social Vs Promedio (DIAGRAMA DE DISPERSIÓN)

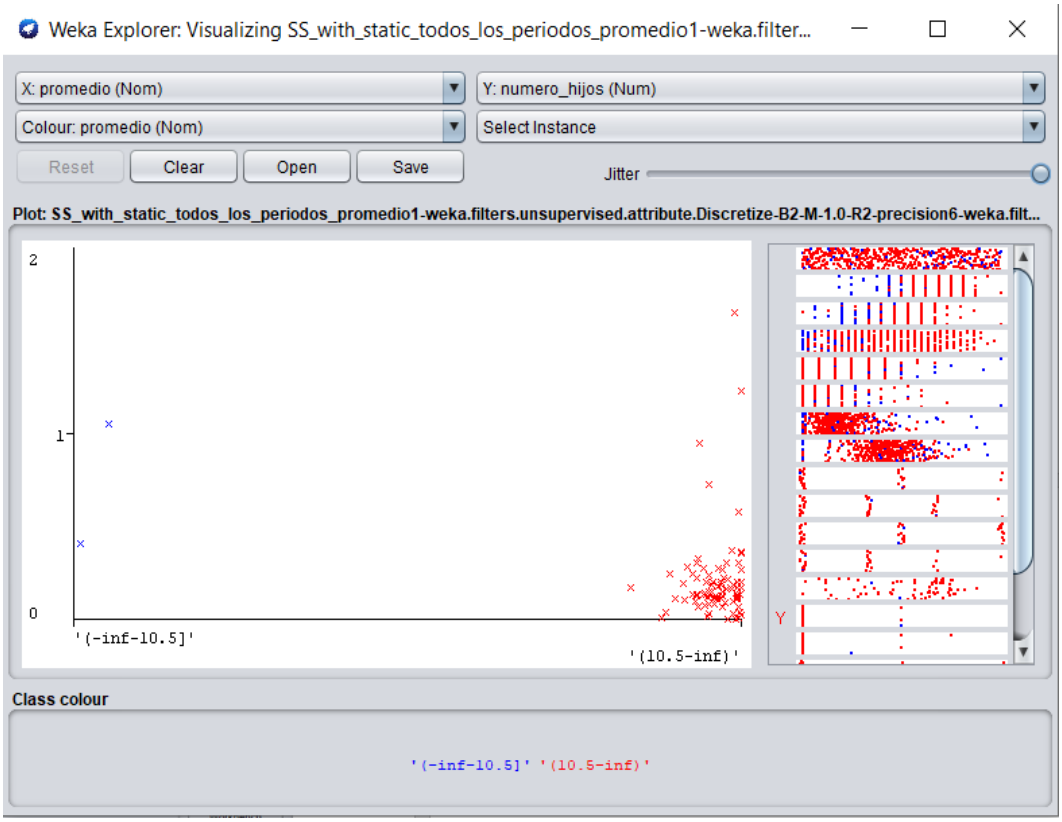


Imagen 1: Diagrama de dispersión situación social

Interpretación: Se observa que la variable número de hijos tiene alta relación con la variable promedio, aludiendo que un gran porcentaje de los alumnos que no tienen hijos tienden a aprobar el semestre.

4.2.2. Mejor atributo de la Situación Económica Vs Promedio (DIAGRAMA DE DISPERSIÓN)

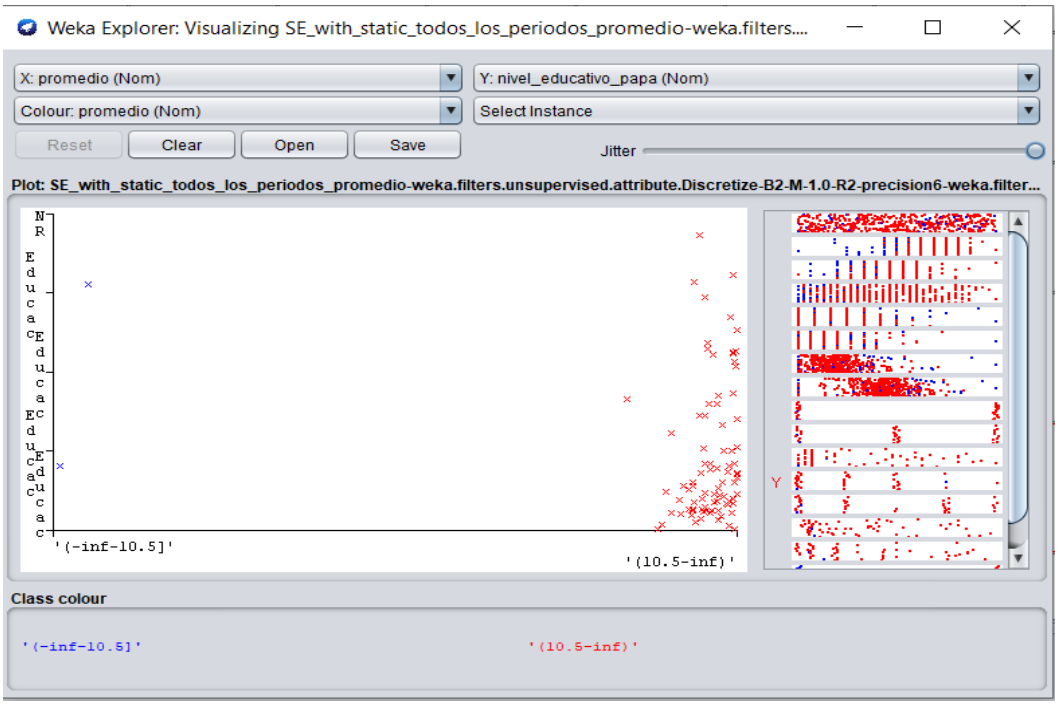


Imagen 2: Diagrama de dispersión situación económica

Interpretación: Se observa que la variable nivel educativo del padre tiene alta relación con la variable promedio, si el nivel educativo del padre es de nivel secundario, el alumno tiende a aprobar el semestre.

4.2.3. Mejor atributo de la Situación Educacional Vs Promedio (DIAGRAMA DE DISPERSIÓN)

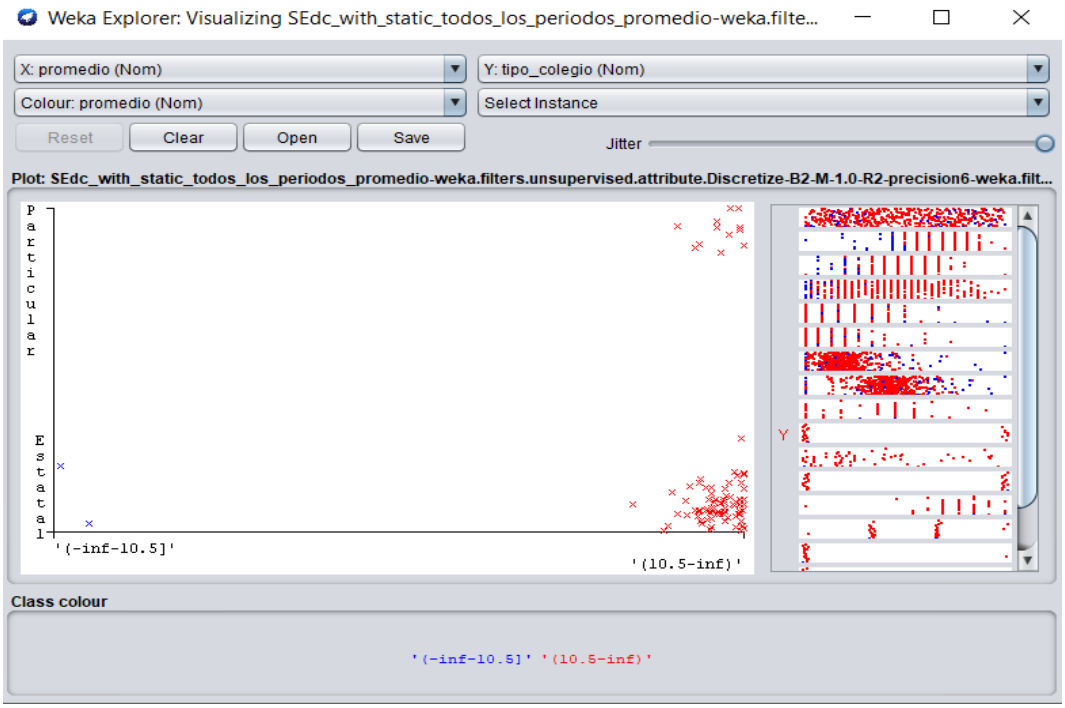


Imagen 3: Diagrama de dispersión situación educacional

Interpretación: Se puede observar que los alumnos que estudiaron en los colegios de tipo estatal, tienen un alto porcentaje de aprobar el semestre.

4.2.4. Mejor atributo de la Situación Personal Vs Promedio (DIAGRAMA DE DISPERSIÓN)

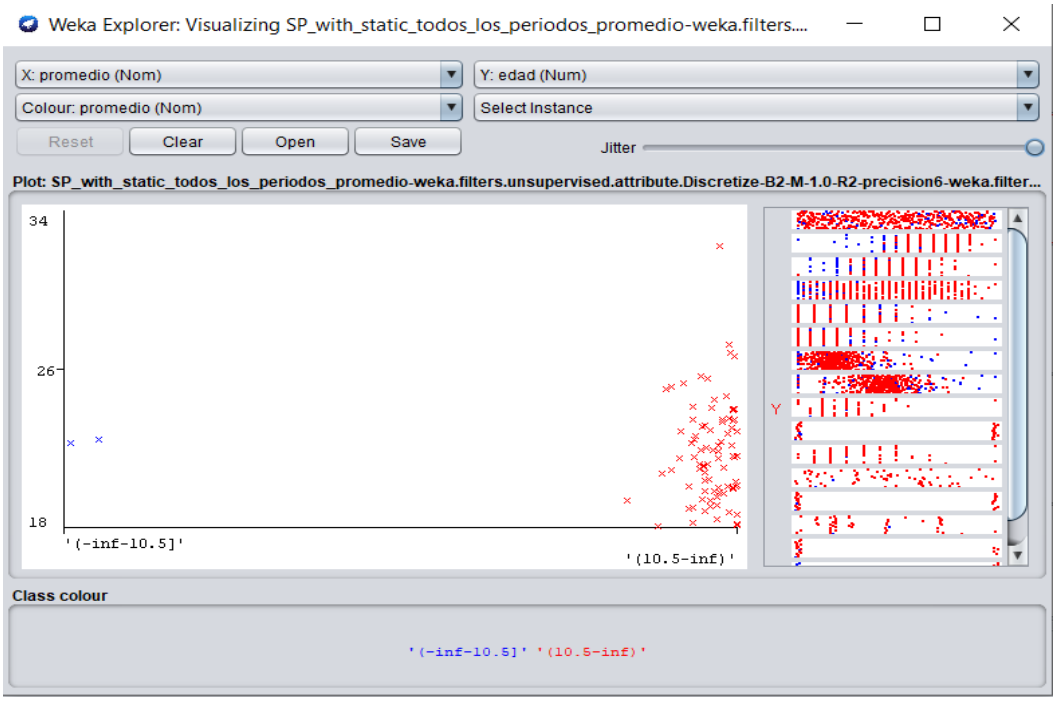


Imagen 4: Diagrama de dispersión situación personal

Interpretación: los alumnos de tienen de 18 a 26 años cumplidos, son los que tienen un alto porcentaje de aprobar el semestre.

4.3. TRANSFORMACIÓN DE DATOS

En esta etapa nos encontramos con la interacción del software Weka, en el ítem Explorer. Se procede a cargar el archivo CSV previamente preparado, para luego transformarlos en archivos de tipo ARFF, que es la extensión con el que se trabaja en Weka, ya en este formato mencionado nos muestra meta información de los atributos en el archivo.

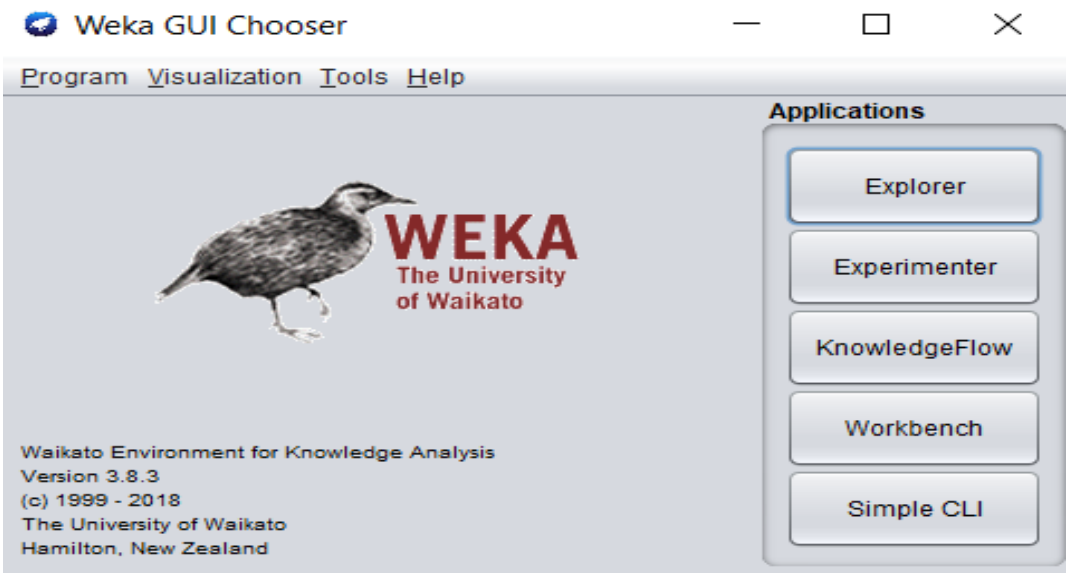


Imagen 5: Software weka

PARA CADA DIMENSIÓN

Se procede a llenar la información faltante, de los atributos de la data estática, con ayuda del algoritmo no supervisado que nos brinda weka, de nombre replacemissingvalue, y a continuación se discretiza el atributo llamado **promedio** en dos partes abarcando este desde (-inf-10.5] y (10.5-inf), siendo estos valores que definan si el alumno está aprobado o desaprobado.

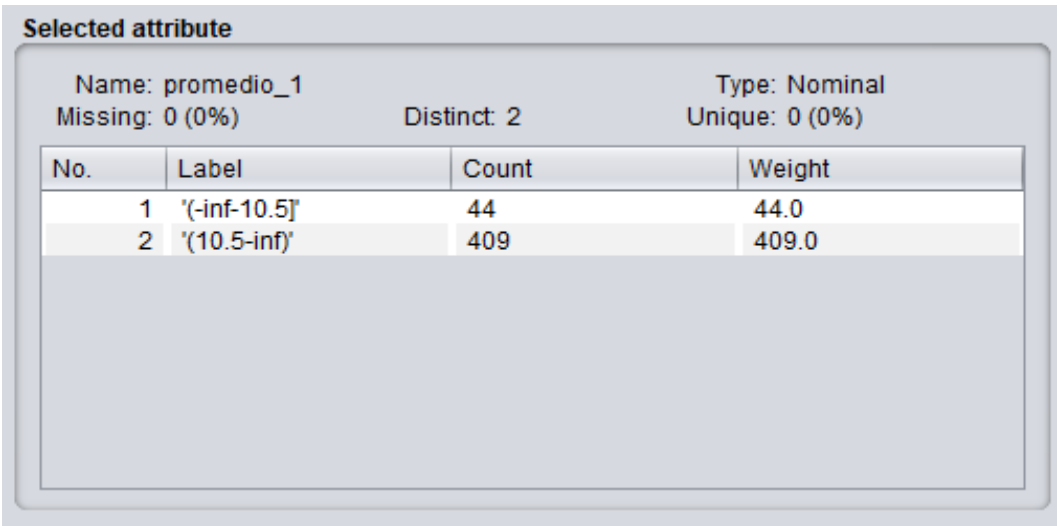


Imagen 6: Promedio discretizado

Luego definimos al atributo **promedio** que es de tipo numérico para la clase principal, que posteriormente se hará un versus entre el promedio y la data estática y dinámica de cada dimensión.

Para finalizar la transformación de datos. Se prosigue a particionar esta información transformada en dos dataset, llamados datos de entrenamiento y datos de prueba que serán valorizados en el 80% y 20% respectivamente.

4.3.1. Situación Social Vs Promedio

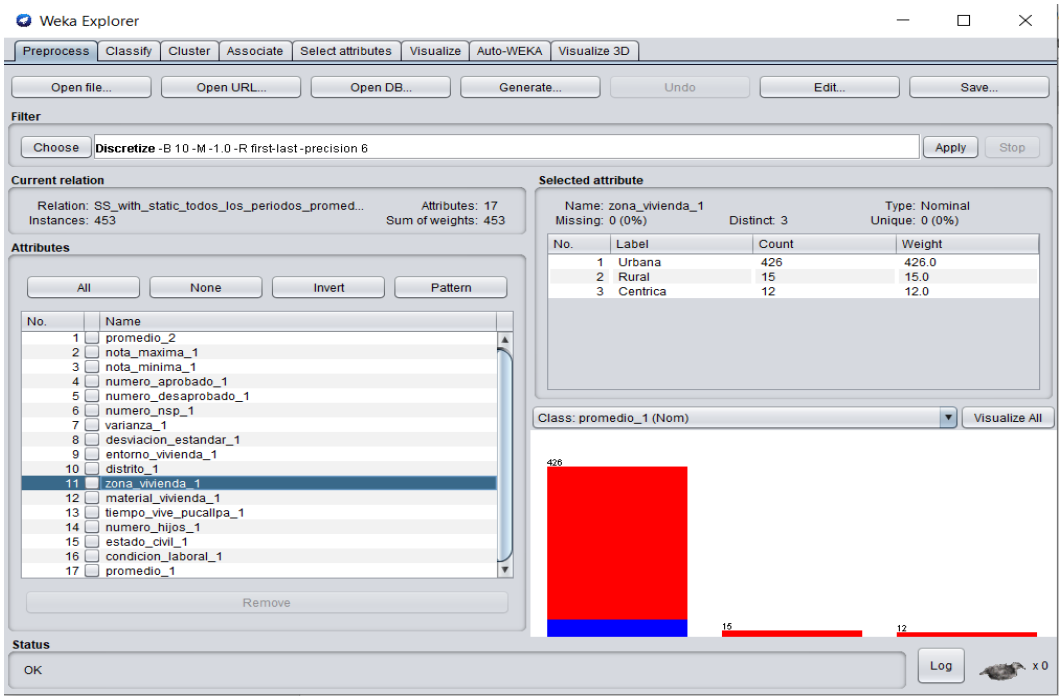


Imagen 7: Situación social vs promedio

4.3.2. Situación Económica Vs Promedio

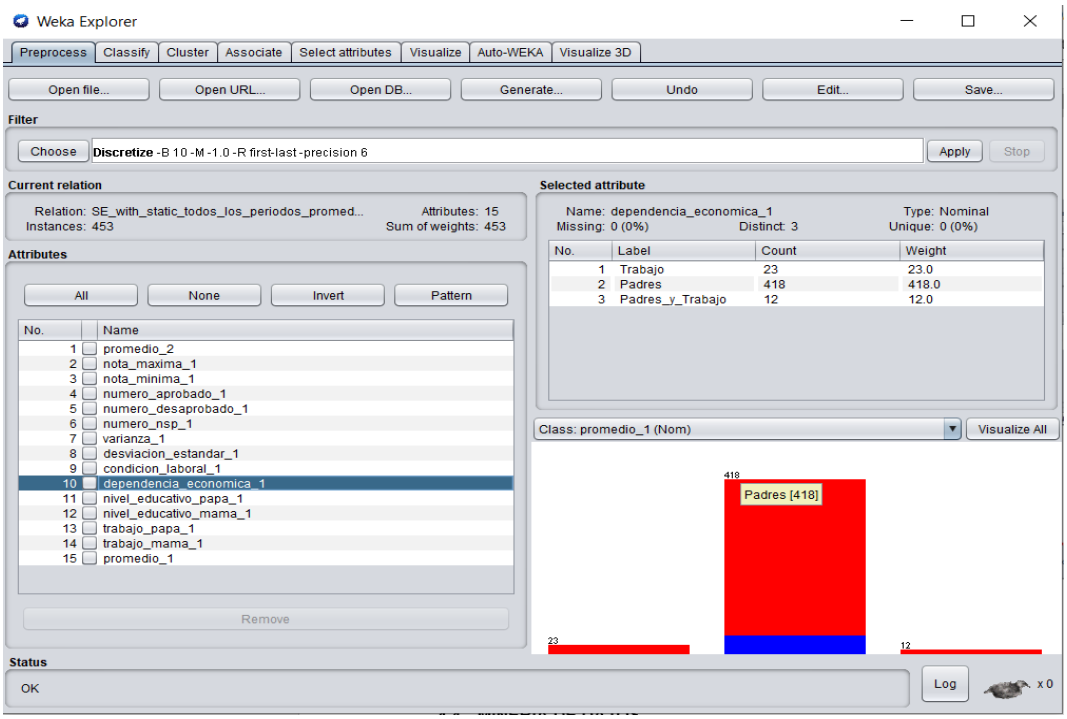


Imagen 8: Situación económica vs promedio

4.3.3. Situación Educacional Vs Promedio

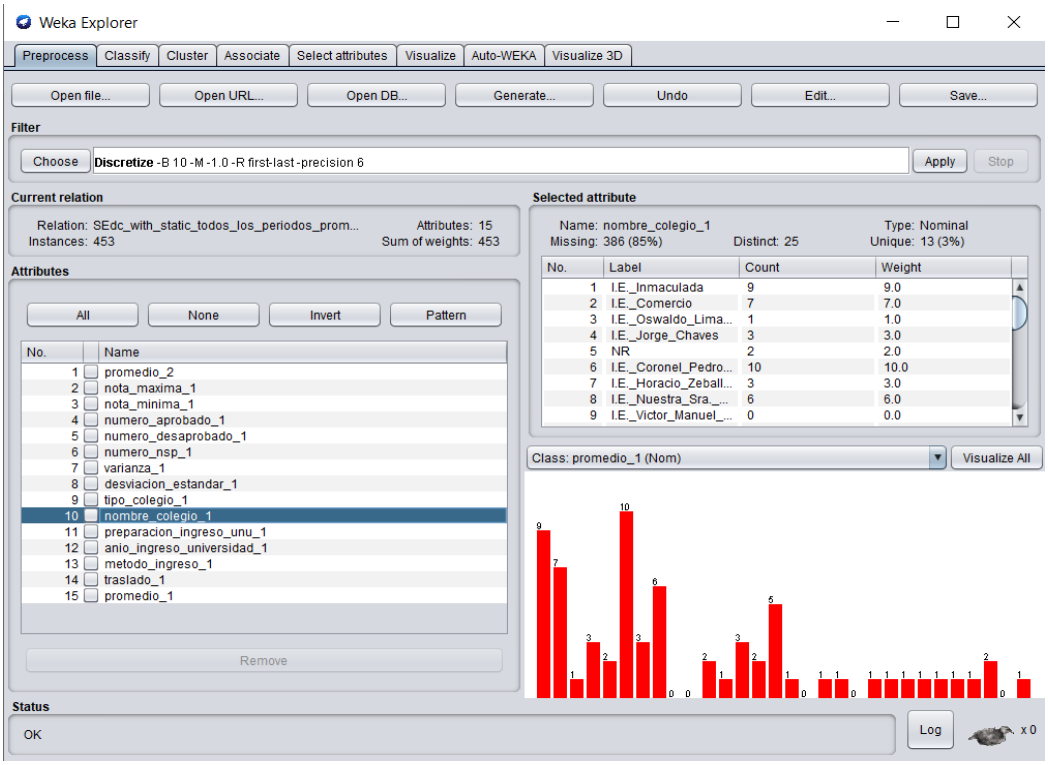


Imagen 9: Situación educacional vs promedio

4.3.4. Situación Personal Vs Promedio

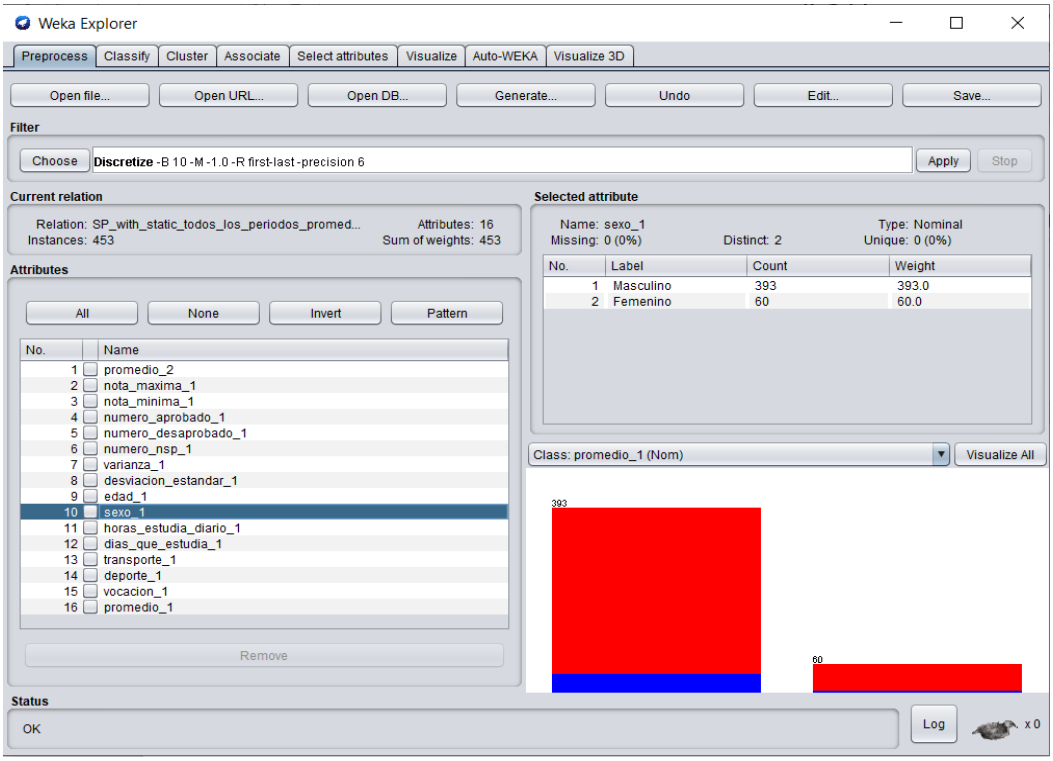


Imagen 10: Situación personal vs promedio

4.4. MINERÍA DE DATOS

Se escoge el algoritmo para entrenar los dataset generados previamente, encontrándose este en la pestaña classify de weka, seguidamente se modifica los parámetros de Randomtree, con maxDepth igual a 10, siendo este el número de ramas que generará el algoritmo en el árbol de decisión, y por última configuración la validación cruzada igual a 10, todo ello para su posterior comprensión e interpretación correcta.

De este modo se da paso al entrenamiento de la data con el algoritmo mencionado, y luego se ejecuta la data de prueba sobre la data entrenada para que weka nos muestre el patrón de datos que se busca.

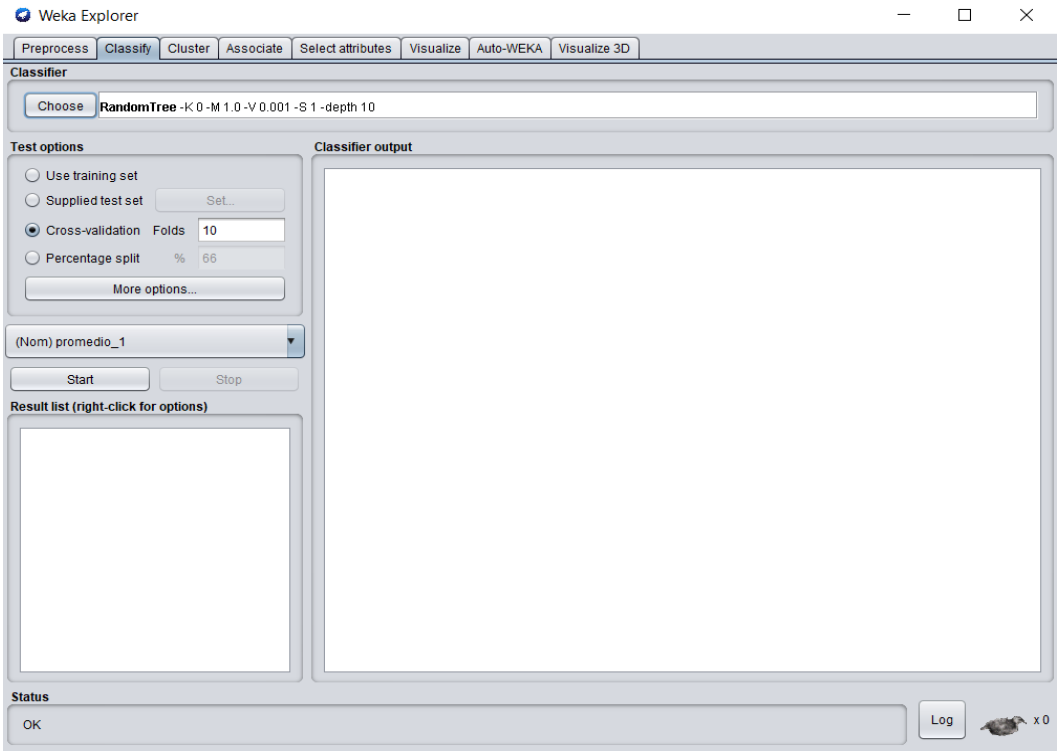


Imagen 11: Algoritmo RandomTree

4.4.1. Resultados del empleo de la técnica de árbol de decisión en la dimensión situación social

=== Run information ===

Scheme: weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -depth 10
Relation: SS_with_static_todos_los_periodos_promedio1-weka.filters.unsupervised.attribute.Remove-R1,16-weka.filters.unsupervised.attribute.Reorder-R1,first-last-weka.filters.unsupervised.attribute.Discretize-B2-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,1-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.supervised.instance.Resample-B0.0-S1-Z80.0-no-replacement-weka.filters.unsupervised.attribute.Remove-R17-

weka.filters.unsupervised.attribute.Reorder-
R2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,1
Instances: 453
Attributes: 16
nota_maxima_1
nota_minima_1
numero_aprobado_1
numero_desaprobado_1
numero_nsp_1
varianza_1
desviacion_estandar_1
entorno_vivienda_1
distrito_1
zona_vivienda_1
material_vivienda_1
tiempo_vive_pucallpa_1
numero_hijos_1
estado_civil_1
condicion_laboral_1
promedio_2
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

RandomTree
==========

Size of the tree: 355
Max depth of tree: 10

Time taken to build model: 0.02 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient	0.8797
Mean absolute error	0.428
Root mean squared error	0.5906
Relative absolute error	48.4915 %
Root relative squared error	48.5739 %
Total Number of Instances	453

=== Re-evaluation on test set ===

User supplied test set
Relation: SS_with_static_todos_los_periodos_promedio1-weka.filters.unsupervised.attribute.Remove-R1,16-weka.filters.unsupervised.attribute.Reorder-R1,first-last-weka.filters.unsupervised.attribute.Discretize-B2-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,1-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.supervised.instance.Resample-B0.0-S1-Z80.0-no-replacement-V
Instances: unknown (yet). Reading incrementally
Attributes: 17

=== Summary ===

Correlation coefficient	0.8974
Mean absolute error	0.4279
Root mean squared error	0.5898
Total Number of Instances	114

4.4.2. Resultados del empleo de la técnica de árbol de decisión en la dimensión situación económica

=== Run information ===

Scheme: weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -depth 10
Relation: SE_with_static_todos_los_periodos_promedio-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.Remove-R11-weka.filters.unsupervised.attribute.Reorder-R1,first-last-weka.filters.unsupervised.attribute.Discretize-B2-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,15,1-weka.filters.supervised.instance.Resample-B0.0-S1-Z80.0-no-replacement-weka.filters.unsupervised.attribute.Remove-R15-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,1
Instances: 453
Attributes: 14

- nota_maxima_1
- nota_minima_1
- numero_aprobado_1
- numero_desaprobado_1
- numero_nsp_1
- varianza_1
- desviacion_estandar_1
- condicion_laboral_1
- dependencia_economica_1
- nivel_educativo_papa_1
- nivel_educativo_mama_1
- trabajo_papa_1
- trabajo_mama_1

promedio_2
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===
RandomTree
==========
Size of the tree: 508
Max depth of tree: 10

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient	0.9084
Mean absolute error	0.3828
Root mean squared error	0.5121
Relative absolute error	43.3689 %
Root relative squared error	42.1216 %
Total Number of Instances	453

=== Re-evaluation on test set ===

User supplied test set
Relation: SE_with_static_todos_los_periodos_promedio-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.Remove-R11-weka.filters.unsupervised.attribute.Reorder-R1,first-last-weka.filters.unsupervised.attribute.Discretize-B2-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,15,1-weka.filters.supervised.instance.Resample-B0.0-S1-Z80.0-no-replacement-V

Instances: unknown (yet). Reading incrementally
Attributes: 15

=== Summary ===

Correlation coefficient	0.8905
Mean absolute error	0.413
Root mean squared error	0.6109
Total Number of Instances	114

4.4.3. Resultados del empleo de la técnica de árbol de decisión en la dimensión situación educacional

=== Run information ===

Scheme: weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -depth 10
Relation: SEdc_with_static_todos_los_periodos_promedio-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16-weka.filters.unsupervised.attribute.Remove-R9-weka.filters.unsupervised.attribute.Reorder-R1,first-last-weka.filters.unsupervised.attribute.Discretize-B2-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,15,1-weka.filters.supervised.instance.Resample-B0.0-S1-Z80.0-no-replacement-weka.filters.unsupervised.attribute.Remove-R15-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,1
Instances: 453
Attributes: 14
nota_maxima_1
nota_minima_1
numero_aprobado_1

numero_desaprobado_1
numero_nsp_1
varianza_1
desviacion_estandar_1
tipo_colegio_1
nombre_colegio_1
preparacion_ingreso_unu_1
anio_ingreso_universidad_1
metodo_ingreso_1
traslado_1
promedio_2

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

RandomTree
==========
Size of the tree : 523
Max depth of tree: 10

Time taken to build model: 0.01 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient	0.8956
Mean absolute error	0.3957
Root mean squared error	0.5436
Relative absolute error	44.8321 %
Root relative squared error	44.71 %
Total Number of Instances	453

=== Re-evaluation on test set ===

User supplied test set
Relation: SEdc_with_static_todos_los_periodos_promedio-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R16-weka.filters.unsupervised.attribute.Remove-R9-weka.filters.unsupervised.attribute.Reorder-R1,first-last-weka.filters.unsupervised.attribute.Discretize-B2-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,15,1-weka.filters.supervised.instance.Resample-B0.0-S1-Z80.0-no-replacement-V
Instances: unknown (yet). Reading incrementally
Attributes: 15

=== Summary ===

Correlation coefficient	0.8747
Mean absolute error	0.4641
Root mean squared error	0.6482
Total Number of Instances	114

4.4.4. Resultados del empleo de la técnica de árbol de decisión en la dimensión situación personal

=== Run information ===

Scheme: weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -depth 10
Relation: SP_with_static_todos_los_periodos_promedio-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.ReplaceMissingValues-

weka.filters.unsupervised.attribute.Reorder-R1,first-last-weka.filters.unsupervised.attribute.Discretize-B2-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,1-weka.filters.supervised.instance.Resample-B0.0-S1-Z80.0-no-replacement-weka.filters.unsupervised.attribute.Remove-R16-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,15,1-weka.filters.unsupervised.attribute.Remove-R11
Instances: 453
Attributes: 14
nota_maxima_1
nota_minima_1
numero_aprobado_1
numero_desaprobado_1
numero_nsp_1
varianza_1
desviacion_estandar_1
edad_1
sexo_1
horas_estudia_diario_1
transporte_1
deporte_1
vocacion_1
promedio_2
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===
RandomTree
==========
Size of the tree : 483
Max depth of tree: 10

Time taken to build model: 0.01 seconds

=== Cross-validation ===

=== Summary ===

Correlation coefficient	0.8893
Mean absolute error	0.4146
Root mean squared error	0.5655
Relative absolute error	46.9784 %
Root relative squared error	46.5079 %
Total Number of Instances	453

=== Re-evaluation on test set ===

User supplied test set

Relation: SP_with_static_todos_los_periodos_promedio-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.ReplaceMissingValues-weka.filters.unsupervised.attribute.Reorder-R1,first-last-weka.filters.unsupervised.attribute.Discretize-B2-M-1.0-R1-precision6-weka.filters.unsupervised.attribute.Reorder-R2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,1-weka.filters.supervised.instance.Resample-B0.0-S1-Z80.0-no-replacement-V

Instances: unknown (yet). Reading incrementally

Attributes: 16

=== Summary ===

Correlation coefficient	0.8576
Mean absolute error	0.5097
Root mean squared error	0.6937
Total Number of Instances	114

TABLA DE RESUMEN

Cuadro 32: Tabla de resumen de algoritmos procesados

ÍTEM	ERROR ABSOLUTO MEDIO	DIMENSIÓN
ENTRENAMIENTO	0.5906	SITUACIÓN SOCIAL
PRUEBA	0.4279	
ENTRENAMIENTO	0.3828	SITUACIÓN ECONÓMICA
PRUEBA	0.4130	
ENTRENAMIENTO	0.3957	SITUACIÓN EDUCACIONAL
PRUEBA	0.4641	
ENTRENAMIENTO	0.4146	SITUACIÓN PERSONAL
PRUEBA	0.5097	

4.5. INTERPRETACIÓN

A continuación, se muestra los árboles de decisiones generados por el algoritmo RandomTree, se configuró al algoritmo para que cada árbol de decisión tenga un máximo 10 niveles, puesto que sin esta configuración se podía expandir más, pero a su vez, complicaría su revisión al momento de la interpretación.

Se aprecia gráficamente la relación que hay entre la data estática y la data dinámica, todo esto a raíz de los atributos de las dimensiones, permitiéndose dar explicación de un patrón de datos que se muestra gracias a la técnica de árbol de decisión.

4.5.1. Árbol de decisión de la dimensión situación social

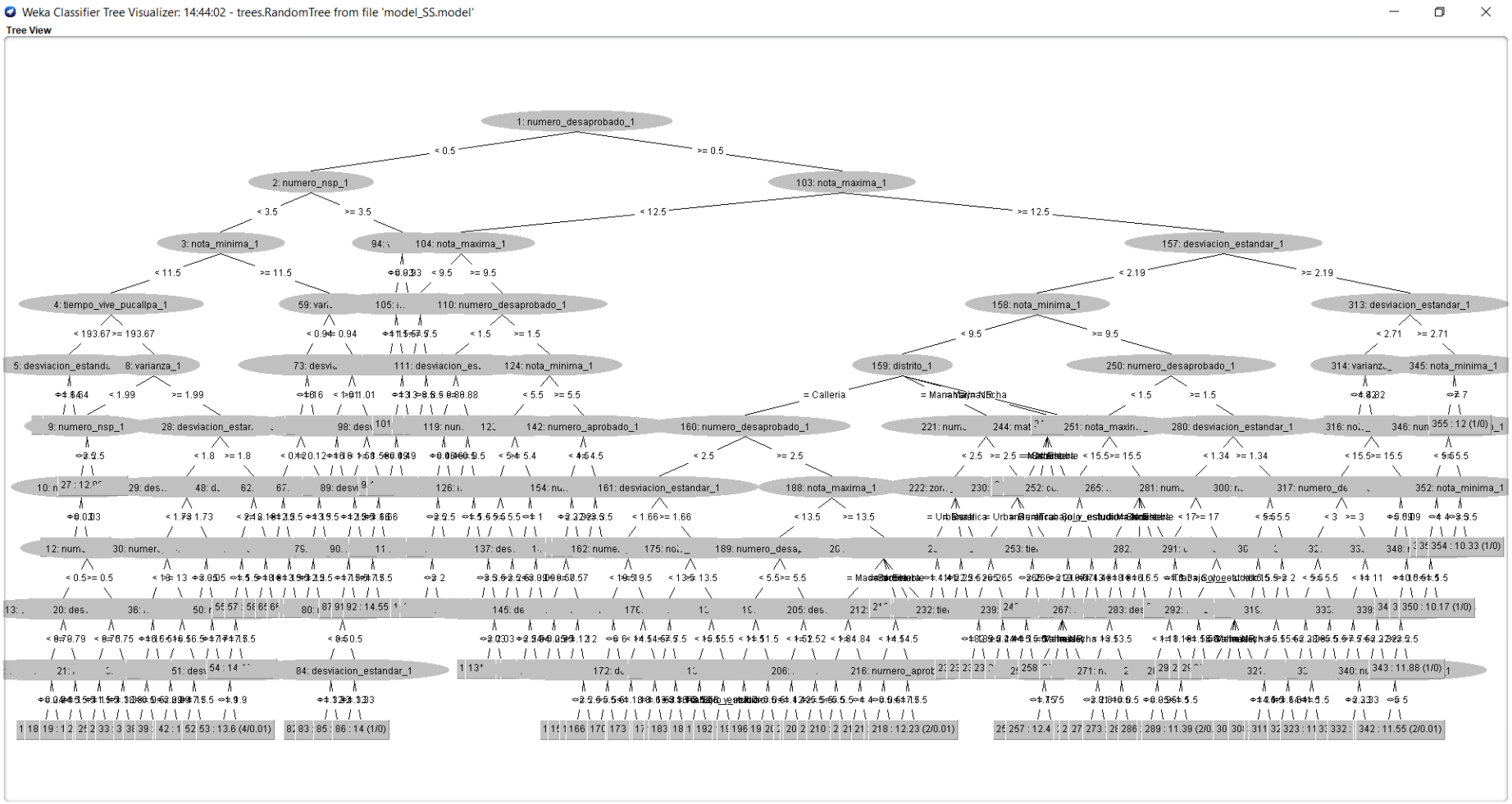


Imagen 12: Árbol de decisión de la situación social

4.5.2. Árbol de decisión de la dimensión situación económica

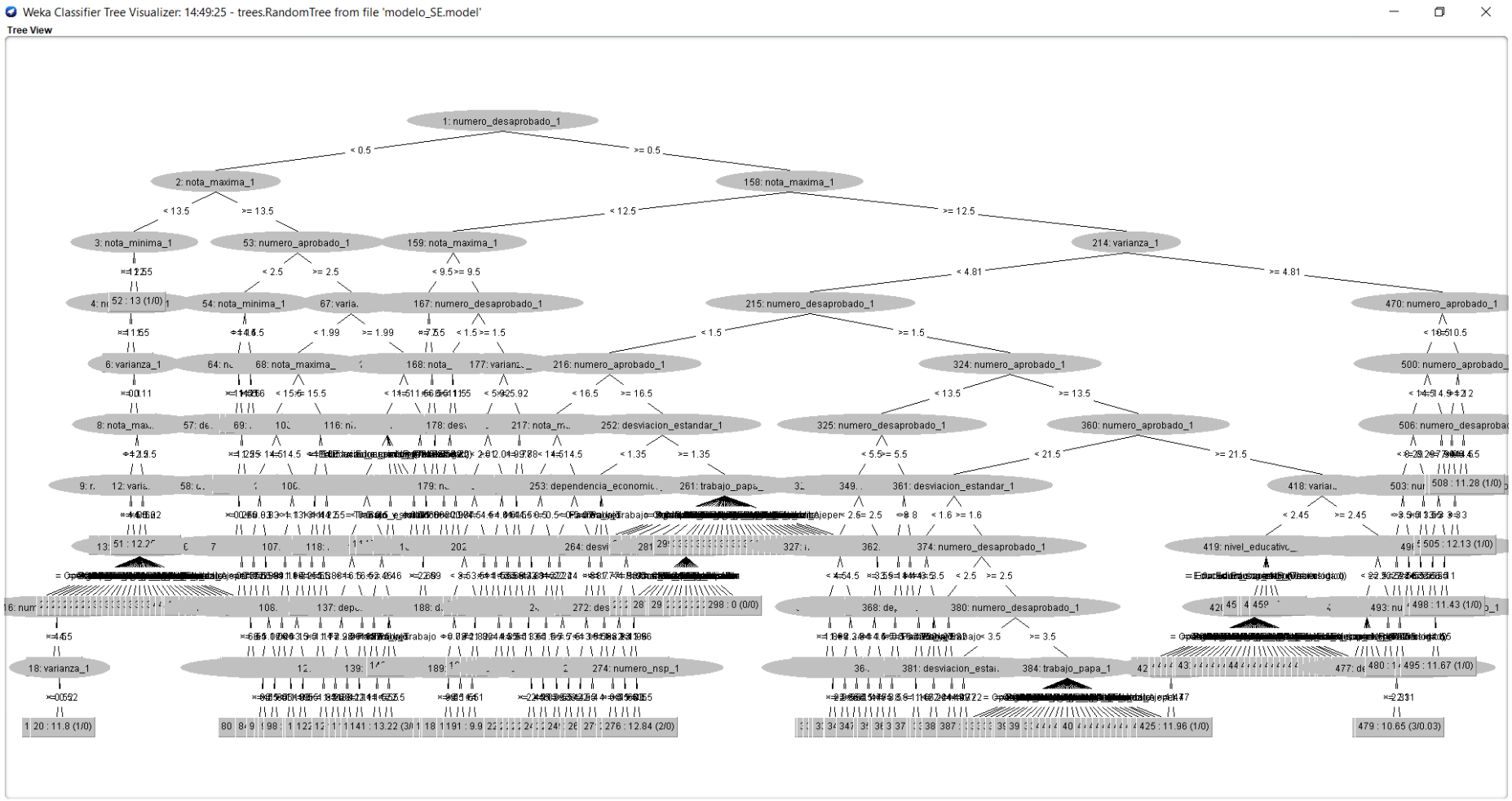


Imagen 13: Árbol de decisión de la dimensión económica

4.5.3. Árbol de decisión de la dimensión situación educacional

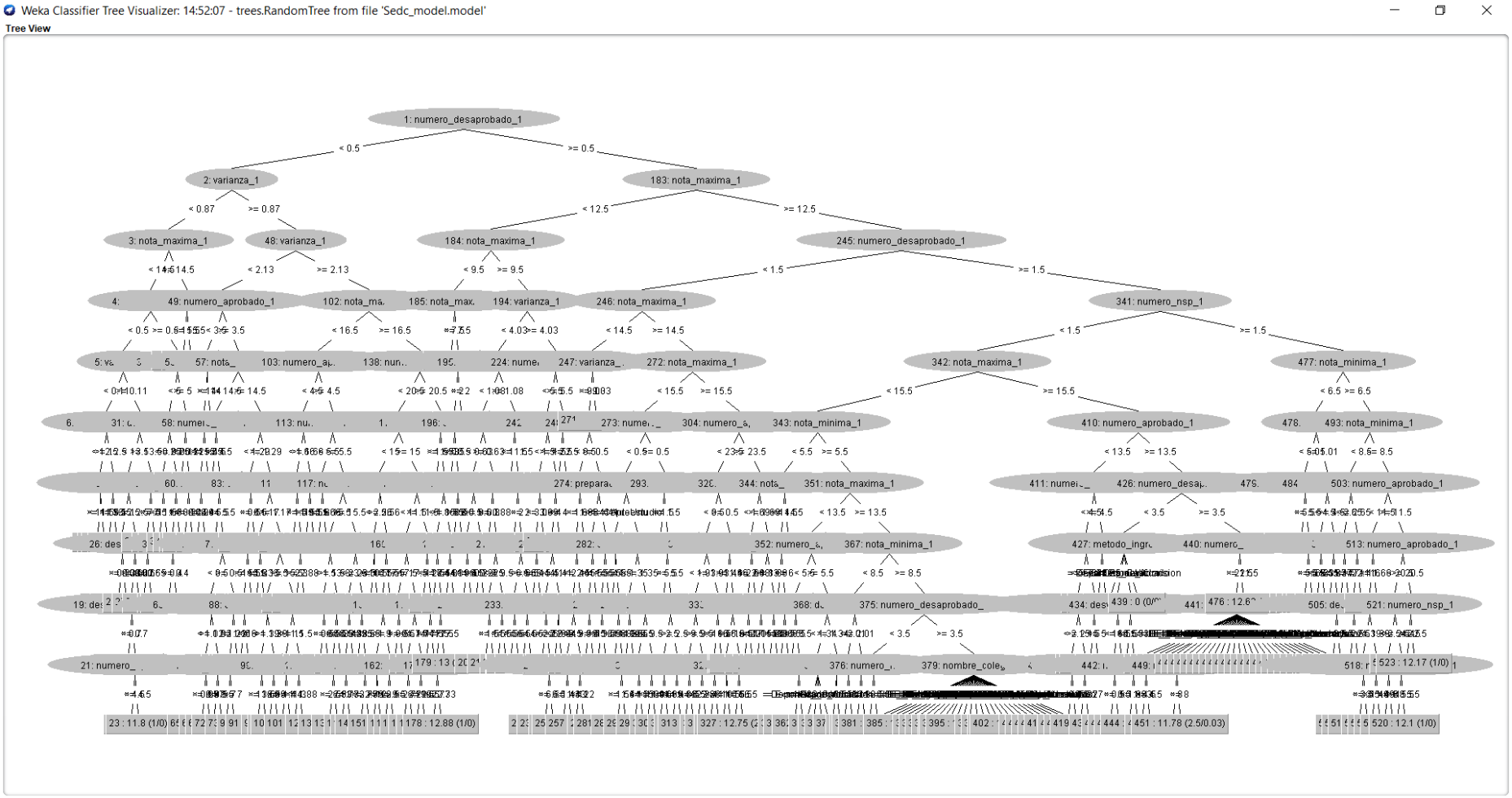


Imagen 14: Árbol de decisión de la situación educacional

4.5.4. Árbol de decisión de la dimensión situación personal

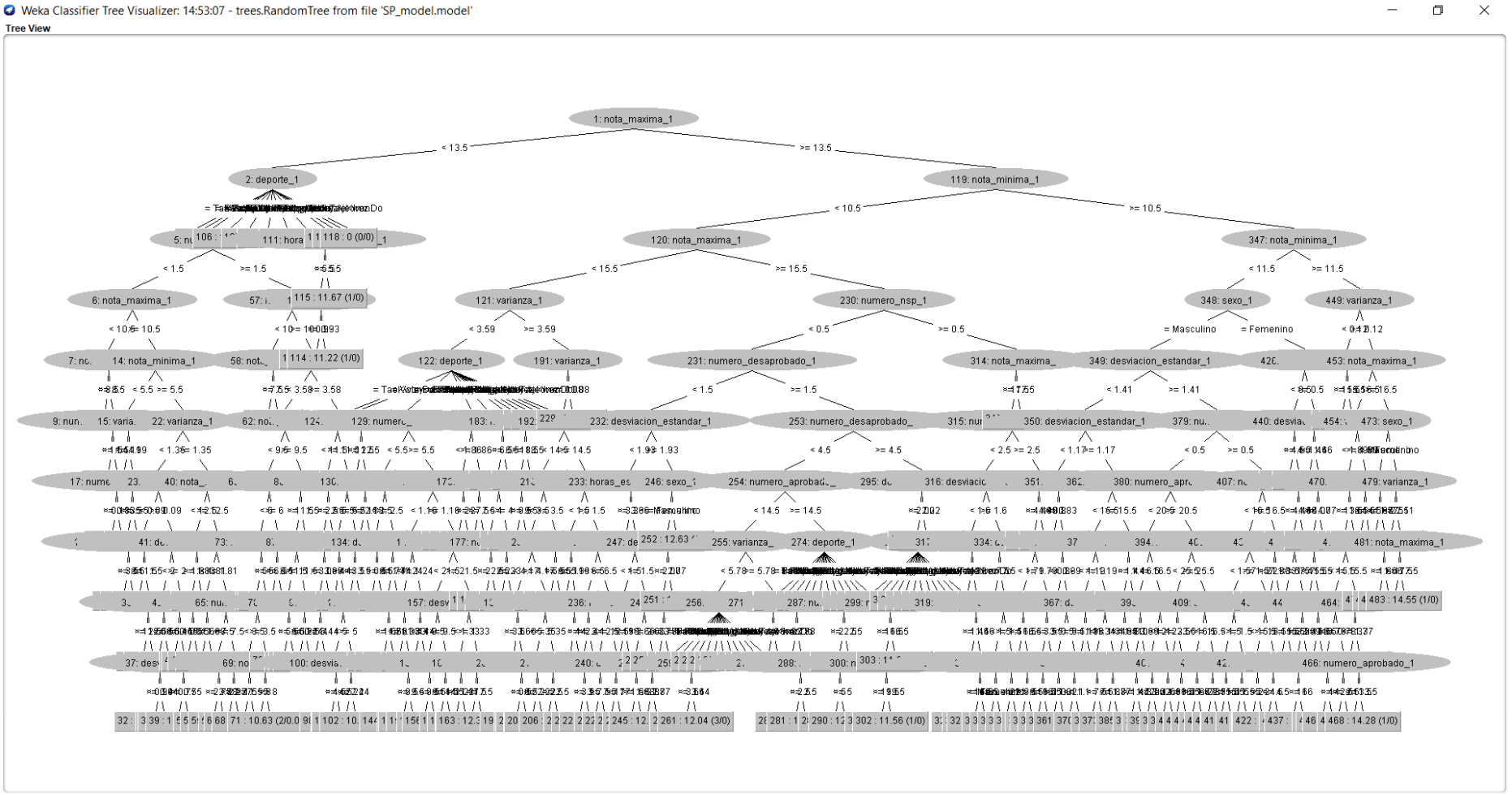


Imagen 15: Árbol de decisión de la dimensión personal

4.5.5. Interpretación del árbol de decisión de la dimensión situación social

Conforme a los resultados generados por el algoritmo RandomTree, se puede afirmar lo siguiente:

- Los alumnos cuyo número de cursos desaprobados sea mayor o igual a 1, su máxima nota sea mayor a 12.5, con desviación estándar menor a 2.19, con nota mínima menor a 9.5, que viven en el distrito de Callería, y el material de su vivienda es seminoble, se predice que pueden rendir académicamente con un promedio de 11.57.

- Los alumnos cuyo número de cursos desaprobados sea mayor o igual a 1, su máxima nota sea mayor a 12.5, con desviación estándar menor a 2.19, con nota mínima menor a 9.5, que viven en el distrito de Yarinacocha, y el material de su vivienda es de madera, se predice que pueden rendir académicamente con un promedio de 11.95.

- Los alumnos cuyo número de cursos desaprobados sea mayor o igual a 1, su máxima nota sea mayor a 12.5, con desviación estándar menor a 2.19, con nota mínima menor a 9.5, que viven en el distrito de Yarinacocha, y el material de su vivienda es noble, se predice que pueden rendir académicamente con un promedio de 12.18.

- Los alumnos cuyo número de cursos desaprobados sea mayor o igual a 1, su máxima nota sea mayor a 12.5, con desviación estándar menor a 2.19, con nota mínima mayor o igual a 9.5, con el número de cursos aprobados mayor o igual a 17, que viven en el distrito de Callería, se predice que pueden rendir académicamente con un promedio de 11.71.

- Los alumnos cuyo número de cursos desaprobados sea mayor o igual a 1, su máxima nota sea mayor a 12.5, con desviación estándar menor a 2.19, con nota mínima mayor o igual a 9.5, con el número de cursos aprobados mayor o igual a 17, que viven en el distrito de Manantay, se predice que pueden rendir académicamente con un promedio de 11.62.

De los resultados obtenidos, se puede afirmar que las variables distrito y material de vivienda, son de mucha importancia cuando se refiere al desempeño académico.

4.5.6. Interpretación del árbol de decisión de la dimensión situación económica

Conforme a los resultados generados por el algoritmo RandomTree, se puede afirmar lo siguiente:

- Los alumnos cuyo número de cursos desaprobados sean menores a 1, con una nota máxima mayor o igual a 13.5, con número de cursos aprobados que sean mayor o igual a 3, con una varianza mayor igual a 1.99, con una nota mínima menor a 11.5, y el nivel educativo del padre sea universitario, y su condición laboral sea que solo estudia, se predice que pueden rendir académicamente con un promedio de 13.39.

- Los alumnos cuyo número de cursos desaprobados sean mayor o igual a 1, con una nota máxima mayor o igual a 12.5, con una varianza menor a 4.81, con el número de cursos desaprobados menores a 2, número de cursos aprobados mayor o igual a 17, con una desviación estándar mayor o igual a 1.35, y si el trabajo de su padre es ser maderero, se predice que pueden rendir académicamente con un promedio de 12.74.

- Los alumnos cuyo número de cursos desaprobados sean mayor o igual a 1, con una nota máxima mayor o igual a 12.5, con una varianza menor a 4.81, con el número de cursos desaprobados menores a 2, número de cursos aprobados mayor o igual a 17, con una desviación estándar mayor o igual a 1.35, y si el trabajo de su padre es ser mototaxista, se predice que pueden rendir académicamente con un promedio de 13.

- Los alumnos cuyo número de cursos desaprobados sean mayor o igual a 1, con una nota máxima mayor o igual a 12.5, con una varianza menor a 4.81, con el número de cursos desaprobados menores a 2, número de cursos aprobados mayor o igual a 17, con una desviación estándar mayor o igual a 1.35, siendo el trabajo de su padre mecánico y el trabajo de la madre enfermera, se predice que pueden rendir académicamente con un promedio de 12.48.

- Los alumnos cuyo número de cursos desaprobados sean mayor o igual a 1, con una nota máxima mayor o igual a 12.5, con una varianza menor a 4.81, con el número de cursos desaprobados menores a 2, número de cursos aprobados mayor o igual a 17, con una desviación estándar mayor o igual a 1.35, siendo el trabajo de su padre mecánico y el trabajo de la madre es ser empresaria, se predice que pueden rendir académicamente con un promedio de 12.73.

De los resultados obtenidos se puede afirmar que las variables condición laboral, trabajo padre y trabajo de mamá, son de mucha importancia cuando se refiere al desempeño académico.

4.5.7. Interpretación del árbol de decisión de la dimensión situación educacional

Conforme a los resultados generados por el algoritmo RandomTree, se puede afirmar lo siguiente:

- Los alumnos cuyo número de cursos desaprobados sean mayor o igual a 1, con una nota máxima mayor o igual a 12.5, con NSP menor a 2, con una nota mínima mayor o igual a 6, con al menos 6 cursos aprobados, con una desviación estándar menor a 1.34 y siendo su método de ingreso a la universidad por Cepreunu, se predice que pueden rendir académicamente con un promedio de 11.35.

- Los alumnos cuyo número de cursos desaprobados sean mayor o igual a 1, con una nota máxima mayor o igual a 12.5, con NSP menor a 2, con una nota mínima mayor o igual a 6, con al menos 6 cursos aprobados, con una desviación estándar menor a 1.34 y siendo su método de ingreso a la universidad por examen de admisión, se predice que pueden rendir académicamente con un promedio de 11.27.

- Los alumnos cuyo número de cursos desaprobados sean mayor o igual a 1, con una nota máxima mayor o igual a 12.5, con NSP menor a 2, con una nota mínima mayor o igual a 6, y habiendo estudiado en el colegio Marko Jara, se predice que pueden rendir académicamente con un promedio de 11.37.

- Los alumnos cuyo número de cursos desaprobados sean mayor o igual a 1, con una nota máxima mayor o igual a 12.5, con NSP menor a 2, con una nota mínima mayor o igual a 6, y habiendo estudiado en el colegio Inmaculada, se predice que pueden rendir académicamente con un promedio de 11.63.

- Los alumnos cuyo número de cursos desaprobados sean mayor o igual a 1, con una nota máxima mayor o igual a 12.5, con NSP menor a 2, con una nota mínima mayor o igual a 6, y habiendo estudiado en el colegio Comercio, se predice que pueden rendir académicamente con un promedio de 11.43.

De los resultados obtenidos se puede afirmar que las variables método de ingreso y colegio, son de mucha importancia cuando se refiere al desempeño académico.

4.5.8. Interpretación del árbol de decisión de la dimensión situación personal

Conforme a los resultados generados por el algoritmo RandomTree, se puede afirmar lo siguiente**:**

- Los alumnos cuya nota máxima sea mayor o igual a 13.5, su nota mínima sea menor a 10.5, con NSP menor a 1, con el número de desaprobados mayor igual a 2, su desviación estándar menor a 2.2, y habiendo practicado el deporte de taekwondo, se predice que pueden rendir académicamente con un promedio de 11.46.

- Los alumnos cuya nota máxima sea mayor o igual a 13.5, su nota mínima sea menor a 10.5, con NSP menor a 1, con el número de desaprobados mayor igual a 2, su desviación estándar menor a 2.2, y habiendo practicado el deporte de voley, se predice que pueden rendir académicamente con un promedio de 11.89.

- Los alumnos cuya máxima nota sea mayor o igual a 13.5, su nota mínima sea menor a 10.5, con NSP menor a 1, con el número de cursos desaprobados menor a 1.5, su desviación estándar menor a 1.93, y estudiando menos de 3 horas diarias, se predice que pueden rendir académicamente con un promedio de 13.

- Los alumnos cuya máxima nota sea mayor o igual a 13.5, su nota mínima sea menor a 10.5, con NSP menor a 1, con el número de cursos desaprobados menor a 1.5, su desviación estándar mayor o igual a 1.93, y siendo de sexo femenino, se predice que pueden rendir académicamente con un promedio de 12.63.

- Los alumnos cuya nota máxima sea mayor o igual a 13.5, con una nota mínima menor a 10.5, con una varianza menor a 3.59, y que practican el deporte de basquetbol, y estudiando al menos 2 horas al día, se predice que pueden rendir académicamente con un promedio de 11.94.

De los resultados obtenidos se puede afirmar que las variables deporte, horas de estudio diario y sexo, son de mucha importancia cuando se refiere al desempeño académico.

CAPÍTULO V

DISCUSIÓN

La presente investigación logró predecir el promedio del alumno a través de cuatro dimensiones (situación social, situación económica, situación personal y situación educacional), habiéndose estudiado con algoritmos de aprendizaje automático y un estudio de serie de tiempo, valiéndose de los resultados con apoyo de la regresión logística.

Sin embargo, en la investigación “PREDICCIÓN DEL RENDIMIENTO ACADÉMICO DE LOS ESTUDIANTES DE LA UNSAAC A PARTIR DE SUS DATOS DE INGRESO UTILIZANDO ALGORITMOS DE APRENDIZAJE AUTOMÁTICO” de (Oviedo, 2019). Demuestra la construcción de un modelo predictivo con base en el entrenamiento de 10790 registros, y posteriormente tomando para pruebas 1819 de estos, obteniendo resultados con el mejor desempeño en el algoritmo Random Forest, no dejando de lado la esencia de estos buenos resultados que son los datos de ingreso del estudiando como nota de ingreso, escuela profesional, semestre y modalidad de ingreso.

A diferencia de la investigación “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” de (Eiriku, 2018) que al utilizar clasificadores bayesianos y el algoritmo KNN, obtiene entre 70% y 74% de exactitud, siendo el sexo, y la provincia datos sumamente referenciales para poder predecir a este nivel el rendimiento académico.

Consecuente a ello se puede agregar que difieren los resultados ya que la información, no es la misma a raíz de factores psicológicos, socioeconómicos y pedagógicos distintos en cada lugar de investigación. Se pudo notar la clara diferencia cuando entrenamos la data con regresión logística y clasificación de datos.

Puesto que queda a criterio propio si particionar la data para entrenamiento y pruebas entre un 70 y 30%, o un 80 y 20%, o un 90 y 10% respectivamente.

Finalmente se puede decir que el criterio, metodología, información y tiempo de cada investigación, es sumamente importante para concretar la extracción de conocimiento de los grandes volúmenes de datos e interpretar buenos resultados.

CONCLUSIONES

1. Referente al objetivo “Estimar el aporte de un modelo predictivo aplicando técnicas de minerías de datos a la situación social del alumno de la E.A.P. de ingeniería de sistemas de la UNU” se concluye que las variables distrito y material de vivienda son fundamentales para predecir el rendimiento académico del alumno, fluctuando un promedio entre 11.57 y 12.18.

2. Referente al objetivo “Definir el aporte de un modelo predictivo aplicando técnicas de minerías de datos a la situación económica del alumno de la E.A.P. de ingeniería de sistemas de la UNU” se concluye que las variables condición laboral, el trabajo del padre y el trabajo de la madre son fundamentales para predecir el rendimiento académico del alumno, fluctuando un promedio entre 11.27 y 11.63.

3. Referente al objetivo “Precisar el aporte de un modelo predictivo aplicando técnicas de minerías de datos a la situación personal del alumno de la E.A.P. de ingeniería de sistemas de la UNU” se concluye que las variables deporte, horas de estudio diario y sexo son fundamentales para predecir el rendimiento académico del alumno, fluctuando un promedio entre 11.46 y 13.

4. Referente al objetivo “Especificar el aporte de un modelo predictivo aplicando técnicas de minerías de datos a la situación educacional del alumno de la E.A.P. de ingeniería de sistemas de la UNU” se concluye que las variables método de ingreso a la universidad y el colegio donde estudió

el alumno son fundamentales para predecir el rendimiento académico del alumno, fluctuando un promedio entre 11.27 y 11.63.

RECOMENDACIONES

1. Para este estudio no se tomó en cuenta el aspecto psicológico del alumno, por las limitaciones de tiempo en la investigación, siendo este un factor muy importante que, con su integración en futuras investigaciones, se podría dar resultados mucho más exactos.

2. Hacer uso de la herramienta Weka por su usabilidad, y comodidad al momento de ejecutar los algoritmos de aprendizaje automático, siendo muy útil al momento de realizar investigaciones sobre ciencias de datos.

3. Se recomienda el uso de la metodología KDD, puesto que se ajusta perfectamente al desarrollo de minería de datos, por su fácil entendimiento y su guía en la extracción de información de grandes volúmenes de datos.

4. Se recomienda a la escuela académica profesional de ingeniería de sistemas, la implementación de mecanismos de recopilación de datos socioeconómicos, pedagógicos y psicológicos del alumnado, dado el incremento e importancia de la información generada en estos últimos años, para referir a una data histórica que pueda ser analizada en un futuro.

REFERENCIA BIBLIOGRÁFICA

Alania Ricaldi, P. F. (2018). Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión. *(Tesis de Posgrado).* Universidad Nacional Daniel Alcides Carrión, Pasco.

ARIMETRICS. (10 de Abril de 2019). *ARIMETRICS*. Obtenido de ARIMETRICS: https://www.arimetrics.com/glosario-digital/modelo-predictivo

AUTENTIA. (12 de Diciembre de 2019). *Adictos al trabajo.* Obtenido de adictos al trabajo: https://www.adictosaltrabajo.com/2019/01/17/introduccion-al-machine-learning-con-weka/

Baustista Delgado, L. (23 de junio de 2018). *data-collection-and-reports*. Obtenido de data-collection-and-reports: http://data-collection-and-reports.blogspot.com/2009/05/recoleccion-de-datos.html

Biblioguías - Bibliotecas CEPAL. (3 de julio de 2018). *CEPAL*. Obtenido de CEPAL: https://biblioguias.cepal.org/ProspectivayDesarrollo/Prospectiva

Bugueño, F. (2017). Modelo predictivo para la selección de postulantes destacados a una Institución de Educación Superior. *(Tesis de posgrado).* Universidad de Chile, Santiago.

Camborda Zamudio, M. G. (2014). Aplicación de árboles de decisión para la predicción del rendimiento académico de los estudiantes de los primeros ciclos de la carrera de Ingenieria Civil de la Universidad Continental. *(Tesis de Posgrado).* Universidad Continental, Huancayo.

Candia Oviedo, D. I. (2019). Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando

algoritmos de aprendizaje automático. *(Tesis de Posgrado).* Universidad Nacional de San Antonio Abad del Cusco, Cusco.

Carvajal, l. (17 de julio de 2018). *Lizardo Carvajal*. Obtenido de Lizardo Carvajal: http://www.lizardo-carvajal.com/metodologia-de-investigacion/

Chavienato, I. (2006). *Introducción a la Teoría General de la Administración.* McGraw-Hill Interamericana.

Conceptodefinicion. (26 de Agosto de 2019). *conceptodefinicion*. Obtenido de conceptodefinicion: https://conceptodefinicion.de/excel/

Córdova Fallas, L. (16 de Julio de 2019). *Minería de datos*. Obtenido de Minería de datos: http://cor-mineriadedatos.blogspot.com/2011/06/weka.html

Corral, Y. (2009). Validez y confiabilidad de los instrumentos de investigación para la recolección de datos. *Revista Ciencia de la Educación*, 20.

Diccionario de leyes. (10 de julio de 2018). *Diccionario de Black de leyes*. Obtenido de Diccionario de Black de leyes: https://espanol.thelawdictionary.org/descubrimiento/

EcuRed. (29 de julio de 2018). *EcuRed*. Obtenido de EcuRed: https://www.ecured.cu/Investigaci%C3%B3n

Eiriku, Y. (2018). 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 de Posgrado).* Universidad San Martin de Porres, Lima.

Etzel, M., Walker, B., & Stanton, W. (2004). *Fundamentos de Marketing.* México: Mc Graw Hill. Obtenido de:

https://mercadeo1marthasandino.files.wordpress.com/2015/02/fundamentos-de-marketing-stanton-14edi.pdf

García Herrero, J., & Molina Lopez, J. (2012). *Técnicas de análisis de datos aplicaciones prácticas utilizando Microsoft Excel y Weka.* Madrid - España. Recuperado el 06 de Mayo de 2019

Gestiondeoperaciones. (12 de octubre de 2019). *gestiondeoperaciones.net*. Obtenido de: gestiondeoperaciones.net: https://www.gestiondeoperaciones.net/procesos/arbol-de-decision/

Hernández Cedano, J. Á. (2015). Modelo de minería de datos para identificación de patrones que influyen en el aprovechamiento académico. *(Tesis de posgrado).* Instituto Tecnológico de la Paz, México.

KDNUGGETS. (18 de Octubre de 2019). *From Data Mining to Knowledge Discovery in Databases.* Obtenido de From Data Mining to Knowledge Discovery in Databases: https://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf

La revista informática. (1 de agosto de 2018). *La revista informática*. Obtenido de la revista informática: http://www.larevistainformatica.com/Significado-Computacion.htm

Leaf group. (20 de junio de 2018). *Geniolandia.* Obtenido de geniolandia: https://www.geniolandia.com/13098780/definicion-de-interpretacion-de-datos

Machine Learning Software in Java. (8 de Abril de 2019). *University of Waikato, New Zealand*. Obtenido de University of Waikato, New Zealand: https://www.cs.waikato.ac.nz/ml/weka/index.html

Meneses, J., & Rodríguez, D. (2019). El cuestionario y la revista. *FUOC*, 8-9.

MICROSOFT. (10 de Agosto de 2019). *MICROSOFT.RED.* Obtenido de docs.microsoft.com: https://docs.microsoft.com/en-us/dotnet/framework/data/adonet/dataset-datatable-dataview/

Mohammad, A. (2018). Measuring Academic Performance of Students in Higher Education Using Data Mining Techniques. *(Tesis de posgrado).* Loughborough University, Reino Unido.

Piscoya Ordoñez, L. E. (2016). Aplicación de técnicas de minería de datos para predecir la deserción estudiantil en la educación básica regular en la región de Lambayeque. *(Tesis de Pregrado).* Universidad Señor de Sipan, Lambayeque.

Platzi. (20 de junio de 2018). *Maestros del web*. Obtenido de Maestros del web: http://www.maestrosdelweb.com/que-son-las-bases-de-datos/

Posgradoune. (20 de octubre de 2018). *Posgradoune.* Obtenido de Posgradoune: http://www.postgradoune.edu.pe/pdf/documentos-academicos/ciencias-de-la-educacion/23.pdf

QuestionPro. (20 de Junio de 2018). *QuestionPro.* Obtenido de QuestionPro: https://www.questionpro.com/es/una-encuesta.html

Real Académia Española. (3 de julio de 2018). *Real Académia Española.* Obtenido de Real académia española: http://dle.rae.es/?id=29kah9l

Reglamento Académico UNU. (15 de Junio de 2017). *UNU*. Obtenido de www.unu.edu.pe: https://www.unu.edu.pe/portal/pdf/2017/ReglamentoAcademicoV1.5_2017.pdf

Searchdatacenter. (1 de agosto de 2018). *Search datacenter en español*. Obtenido de search datacenter en español: https://searchdatacenter.techtarget.com/es/definicion/Analisis-de-Datos

Tecnologías información. (16 de junio de 2018). *Tecnologías información.* Obtenido de tecnologías información: https://www.tecnologias-informacion.com/agregacion.html

Tecnologías información. (15 de julio de 2018). *Tecnologías información*. Obtenido de Tecnologías información: https://www.tecnologias-informacion.com/transformacion.html

Textos sobre la lupa. (1 de agosto de 2018). *Textos sobre la lupa*. Obtenido de Textos sobre la lupa: http://textossobrelalupa.blogspot.com/2012/11/preparacion-y-analisis-de-datos-cap-3.html

UBUNLOG. (07 de Julio de 2019). *UBUNLOG*. Obtenido de UBUNLOG: https://ubunlog.com/mysql-workbench-bases-datos/

Universidad Politécnica de Cartagena. (10 de julio de 2018). *Crai biblioteca.* Obtenido de crai biblioteca: http://www.bib.upct.es/datos-de-investigacion

Vásquez, J. (2016). Modelo predictivo para estimar la deserción de estudiantes en una Institución de Educación Superior. *(Tesis de posgrado).* Universidad de Chile, Santiago.

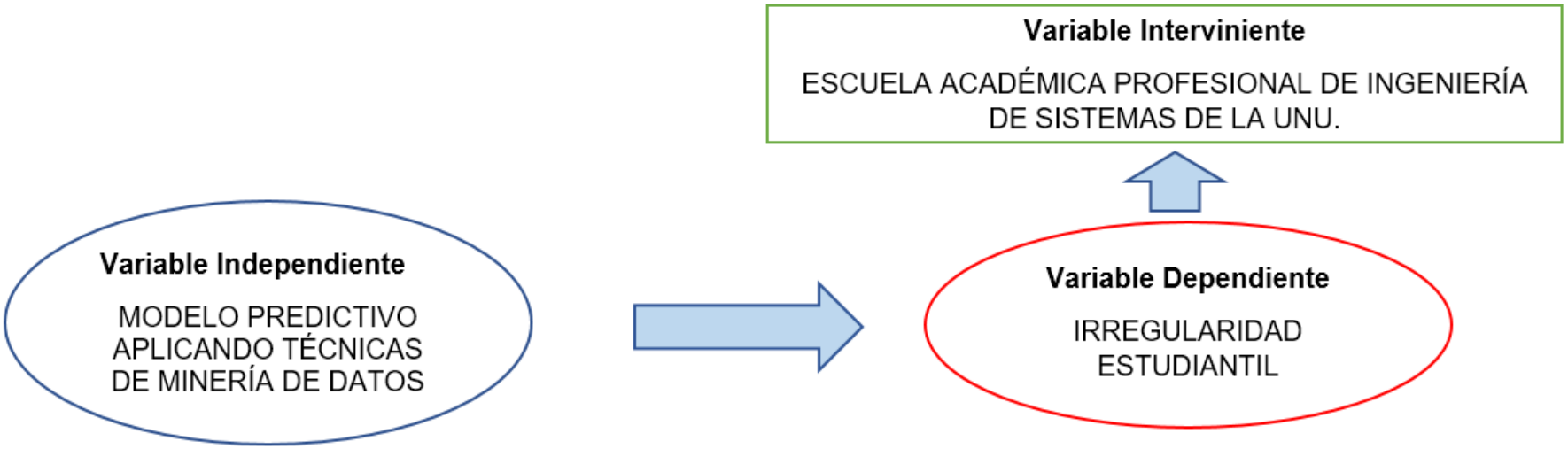
Whitley, L. A. (2018). Educational data mining and its uses to predict the most. *(Tesis de posgrado).* East Carolina University, Estados Unidos.

ANEXO

ANEXO 1: MATRIZ DE IDENTIFICACIÓN DEL PROBLEMA

N°	SÍNTOMAS	CAUSAS	CONSECUENCIAS	PRONÓSTICO
01	El alumno se sobrecarga distribuyendo su tiempo en el trabajo y el estudio.	La baja situación socioeconómica en la que se encuentra el alumno.	El alumno pierde horas de estudio, por estar trabajando.	El alumno debe distribuir y responsabilizarse de su horario de trabajo y de estudio.
02	El bajo nivel de enseñanza que reciben los alumnos en los colegios, en áreas como lenguaje y matemáticas.	El estado peruano tiene funcionarios con falta de principios y ética en el sector educación.	Desaprobación de cursos constantemente por el alumno.	El estado peruano debe capacitar constantemente a los docentes y brindarles un buen salario.
03	Optar por estudiar ingeniería de sistemas con falta de vocación.	La necesidad de formalizarse como profesional y no perder tiempo sin estudiar.	Falta de compromiso por parte del alumno en el transcurso de su formación profesional.	Terminar la carrera profesional, optando posteriormente por la carrera de su agrado.
04	Desconocimiento de tecnologías que brinden predicción de datos para la deserción de los alumnos.	No investigar temas científicos y tecnológicos que puedan ayudar a comprender el apoyo que tenemos en minería de datos	No explotar la información con la que se cuenta para encontrar patrones de deserción estudiantil.	Investigar y aprender esta tecnología(DM) para un buen uso de la información con la que se cuenta.
“IRREGULARIDAD ESTUDIANTIL EN LA E.A.P. DE INGENIERÍA DE SISTEMA DE LA UNU”				

ANEXO 2: HOJA DE RUTA DE LAS VARIABLES



	MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS	IRREGULARIDAD ESTUDIANTIL	ESCUELA ACADÉMICA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNU.
TÍTULO	Modelo predictivo aplicando técnicas de minería de datos para la irregularidad estudiantil en la escuela académica profesional de ingeniería de sistemas de la UNU.		
FORMULACIÓN DEL PROBLEMA	¿De qué manera un modelo predictivo aplicando técnicas de minería de datos determina la irregularidad estudiantil en la escuela académica profesional de ingeniería de sistemas de la UNU?		
OBJETIVO	Establecer un modelo predictivo aplicando técnicas de minerías de datos para determinar la irregularidad estudiantil en la escuela académica profesional de ingeniería de sistemas de la UNU.		
HIPÓTESIS	El modelo predictivo aplicando técnicas de minería de datos determina la irregularidad estudiantil en la escuela académica profesional de ingeniería de sistemas de la UNU.		

ANEXO 3: MATRIZ DE OPERACIONALIZACIÓN DE LAS VARIABLES

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES
MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS	(ARIMETRICS, 2019). Es el agrupamiento de funciones matemáticas y de la estadística inferencial para predecir resultados.	El modelo predictivo para este estudio comprende estas cinco dimensiones: selección de datos, pre-procesamiento, transformación de datos, minería de datos e interpretación.	Selección de datos	• Analizar los datos obtenidos.
			Pre - procesamiento	• Preparación y limpieza de datos.
			Transformación de datos	• Tratamiento dimensional de los datos.
			Minería de datos	• Extracción de patrones de datos.
			Interpretación	• Interpretación de los datos.
IRREGULARIDAD ESTUDIANTIL	(Reglamento Académico UNU, 2017). Referente a los que están matriculados por condición, los que tienen desaprobados una o más materias, los observados, los que se matriculan con menos de doce créditos.	La irregularidad estudiantil para este estudio comprende de cuatro dimensiones: situación social, situación económica, situación personal y situación educacional del alumno.	Situación social	• Entorno físico vivienda • Lugar vivienda • Zona vivienda • Material de vivienda • Tiempo que radica en Pucallpa • Número de hijos • Edad de hijos • Estado civil
			Situación económica	• Condición laboral • Dependencia económica • Cuánto dinero percibe mensual • Nivel educativo padre y madre • Trabajo padre y madre
			Situación personal	• Edad • Sexo • Horas que dedica a estudiar • Días que estudia fuera de la universidad • Transporte • Deporte que practica • vocación
			Situación educacional	• Año que termino la secundaria • Tipo de colegio • Nombre de colegio • Preparación para ingreso a la universidad • Año de ingreso a la universidad • Método de ingreso a la universidad • Realizó traslado de otra universidad • Universidad de procedencia

ANEXO 4: MATRIZ DE CONSISTENCIA SISTÉMICA

<table>
<tr><th>FORMULACIÓN DEL PROBLEMA</th><th>OBJETIVOS</th><th>HIPÓTESIS</th><th>VARIABLES, DIMENSIONES E INDICADORES</th><th>METODOLOGÍA DE LA INVESTIGACIÓN</th></tr>
<tr>
<td>¿De qué manera un modelo predictivo aplicando técnicas de minería de datos determina la irregularidad estudiantil en la escuela académica profesional de Ingeniería de Sistemas de la UNU?</td>
<td>Establecer un modelo predictivo aplicando técnicas de minería de datos para determinar la irregularidad estudiantil en la escuela académica profesional de Ingeniería de Sistemas de la UNU</td>
<td>El modelo predictivo aplicando técnicas de minería de datos determina la irregularidad estudiantil en la escuela académica profesional de Ingeniería de Sistemas de la UNU.</td>
<td rowspan="2">VARIABLE 1: MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS
DIMENSIÓN: SELECCIÓN DE DATOS
• Analizar los datos obtenidos previamente.
DIMENCIÓN: PRE-PROCESAMIENTO
• Preparación y limpieza de datos obtenidos con anterioridad.
DIMENSIÓN: TRANSFORMACIÓN
• Tratamiento preliminar de los datos obtenidos con anterioridad.
DIMENSIÓN: MINERÍA DE DATOS
• Extracción de patrones de los datos procesados.
VARIABLE 2: IRREGULARIDAD ESTUDIANTIL
DIMENSIÓN: SITUACIÓN SOCIAL
• Entorno físico.
• Lugar vivienda.
• Zona vivienda.
• Material de vivienda.
• Tiempo que radica en Pucallpa.
• Número de hijos.
• Edad de hijos.
• Estado civil.
DIMENSIÓN: SITACIÓN ECONÓMICA
• Condición laboral.
• Depende económicamente.
• Cuánto dinero percibe mensual.
• Nivel educativo del padre.
• Nivel educativo de la madre.
• Trabajo del padre.
• Trabajo de la madre.
DIMENSIÓN: SITUACIÓN PERSONAL
• Edad.
• Sexo.
• Horas que dedica a estudiar.
• Días que estudia en la universidad.
• Transporte.
• Deporte que practica.
• Vocación.
DIMENSIÓN: SITUACIÓN EDUCACIONAL
• Año que termino la secundaria.
• Tipo de colegio.
• Nombre del colegio.
• Preparación para ingreso a la universidad.
• Año de ingreso a la universidad.
• Método de ingreso a la universidad.
• Realizó traslado de otra universidad.
• Universidad de procedencia.</td>
<td rowspan="2">TIPO DE INVESTIGACIÓN
APLICADA.

NIVEL DE INVESTIGACIÓN
EXPLORATORIO.

POBLACIÓN
La población de estudio en esta investigación está constituida por los registros académicos de los alumnos de la escuela académica profesional de ingeniería de sistemas de la UNU, entre los periodos 2015 y 2018.

MUESTRA
Muestra no probabilística-intencionada, conformada por los registros académicos de 596 estudiantes de la escuela académica profesional de ingeniería de sistemas UNU registrados entre los periodos 2015 y 2018.

DISEÑO DE LA INVESTIGACIÓN
Diseño no experimental: transeccional exploratorio.
ESQUEMA


Donde:
- X: Modelo predictivo aplicando técnicas de minería de datos.
- O: Irregularidad estudiantil en la E.A.P. de ingeniería de sistemas de la UNU

SOFTWARE PARA PROCESAR DATOS:
Weka

<table>
<tr><th>Fuentes</th><th>Técnicas</th><th>Herramientas</th></tr>
<tr><td>Primaria</td><td>Encuesta</td><td>Cuestionario</td></tr>
<tr><td>Secundaria</td><td>Análisis de Datos</td><td>Procesamiento en minería de datos.</td></tr>
</table>
</td>
</tr>
<tr>
<td>a) ¿Cuál es la estimación de la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación social del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU?

b) ¿Cuál es la estimación de la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación económica del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU?

c) ¿Cuál es la estimación de la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación personal del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU?

d) ¿Cuál es la estimación de la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación educacional del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU?</td>
<td>a) Estimar la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación social del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

b) Estimar la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación económica del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

c) Estimar la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación personal del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

d) Estimar la irregularidad estudiantil empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación educacional del alumno de la E.A.P. de ingeniería de sistemas de la UNU.</td>
<td>a) La estimación de la irregularidad estudiantil es significativa empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación social del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

b) La estimación de la irregularidad estudiantil es significativa empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación económica del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

c) La estimación de la irregularidad estudiantil es significativa empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación personal del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.

d) La estimación de la irregularidad estudiantil es significativa empleando un modelo predictivo aplicando técnicas de minería de datos considerando la situación educacional del alumno de la E.A.P. de Ingeniería de Sistemas de la UNU.</td>
</tr>
</table>

ANEXO 5: CUESTIONARIO DE PREGUNTAS DE RECOLECCIÓN DE DATOS

Número de Cuestionario: _____

Fecha: _____

Instrucciones: Sírvase responder marcando con una (X), de acuerdo a su apreciación.

Alumno: ……

Ciclo : ……...

PREGUNTAS ESPECÍFICAS

VARIABLE 2: IRREGULARIDAD ESTUDIANTIL

DIMENSIÓN: SITUACIÓN SOCIAL

01. Entorno físico de vivienda

a) Propio **b)** Alquilado **c)** Cedido **d) especifique: ……**

02. Lugar de vivienda

Distrito: ……..……..

03. Zona de vivienda: **a)** Rural **b)** Céntrica **d)** Urbana

04. Material de vivienda **a)** Noble **b)** Madera **d)** Estera **e)** Seminoble

05. Tiempo que radica en Pucallpa …….. Años ……. Meses

06. ¿Cuántos hijos tiene? (si tuviera) ……..……..……

07. De tenerlos, ¿cuáles son sus edades? ……..……

08. Estado Civil ……..……..…

DIMENSIÓN: SITUACIÓN ECONÓMICA

09. ¿Cuál es su condición laborar actual?

a) solo estudio **b)** trabajo y estudio

10. ¿De quién depende económicamente?

a) Padres **b)** Trabajo **c)** Padres y trabajo

11. ¿Cuánto percibe mensualmente? S/. ……..……..

12. Nivel educativo del padre o tutor

a) Educación primaria **b)** Educación secundaria

c) Educación superior (Tecnológico) **d)** Educación superior (Universitario)

13. Nivel educativo de la madre o tutora

a) Educación primaria **b)** Educación secundaria

c) Educación superior (Tecnológico) **d)** Educación superior (Universitario)

14. A que se dedica sus padres o tutor

Papá: …….

Mamá: ……

DIMENSIÓN: SITUACIÓN PERSONAL

15. Tu edad: ……

16. **Tu sexo:** **a)** Masculino **b)** Femenino **c)** Otros

17. ¿Cuántas horas dedica a estudiar al día? ……..

18. Marque los días que dedica a estudiar fuera de la universidad

a) Lunes **b)** Martes **c)** Miércoles **d)** Jueves **e)** Viernes **f)** Sábado **e)** Domingo **f)** Todos los días

19. Transporte que usa para movilizarse a la U.N.U.

a) Servicio Público **b)** Propia **c)** Otros (especifique) ……..

20. ¿Qué disciplina deportiva practica?

a) Fútbol **b)** Vóley **c)** Basquetbol **d)** Tae kwon Do **e)** Ajedrez **f)** Ninguno

21. ¿Usted tiene vocación por la carrera de Ingeniería de Sistemas?

a) Si **b)** No

DIMENSIÓN: SITUACIÓN EDUCACIONAL

22. Año que terminó la secundaria: ……

23. Tipo de colegio en que estudió: **a)** Estatal **b)** Particular

24. Nombre de colegio: ……..

25. Preparación para el ingreso en la UNU

a) CEPREUNU **b)** Autoestudio **c)** Otros (especifique)……..

26. Año de ingreso a la universidad: …….

27. ¿Cómo ingresó usted a la universidad?

a) Examen de admisión **b)** Exoneración **c)** Título profesional **d)** CEPREUNU

e) Otros (especifique) ……..

28. ¿Realizó traslado de otra universidad?

a) Si **b)** No

29. Universidad de procedencia (Si hubiese marcado sí anteriormente)

……

ANEXO 6: VALIDACIÓN DE LOS INSTRUMENTOS POR EXPERTOS

I. DATOS PERSONALES:

1.1. APELLIDOS Y NOMBRES DEL INFORMANTE : Pandro Padilla Euclides

1.2. GRADO ACADÉMICO : Msc. en Ingeniería de Sistemas

1.3. INSTITUCIÓN DONDE LABORA : Universidad Nacional de Ucayali

1.4. TITULO DE LA INVESTIGACIÓN : MODELO PREDICTIVO UTILIZANDO TÉCNICAS DE MINERIA DE DATOS PARA DETERMINAR LA IRREGULARIDAD ESTUDIANTIL EN LA E.A.P DE ING. DE SISTEMAS EN LA UNU.

1.5. AUTOR DEL INSTRUMENTO : PORTUGUEZ DURAND, ENRIQUE IGLESIAS

1.6. NOMBRE DEL INSTRUMENTO : ENCUESTA

II. ASPECTOS A EVALUAR: (CALIFICACIÓN CUANTITATIVA)

INDICADORES DEL INSTRUMENTO DE EVALUACIÓN	CRITERIOS CUALITATIVOS CUANTITATIVOS	Deficiente (01-09)	Regular (10-13)	Bueno (14-16)	Muy bueno (17-18)	Excelente (19-20)
01 CLARIDAD	Está formulado con lenguaje apropiado				17	
02 OBJETIVIDAD	Está expresado con conductas observables			15		
03 ACTUALIDAD	Adecuado al avance de la ciencia y calidad				18	
04 ORGANIZACIÓN	Existe una organización lógica del instrumento				17	
05SUFICIENCIA	Valora los aspectos en cantidad y calidad				17	
06 INTENCIONALIDAD	Adecuado para cumplir con los objetivos			16		
07 CONSISTENCIA	Basado en el aspecto teórico científico del tema de estudios				18	
08 COHERENCIA	Entre las hipótesis, dimensiones e indicadores			16		
09 METODOLOGÍA	Las estrategias responden al propósito del estudio				17	
10 OPORTUNIDAD	Genera nuevas pautas para la investigación y construcción de teorías				18	
Sub Total				47	122	
Total						169

Valoración cuantitativa (total x0.4) 17

Valoración cualitativa ..... aceptable

Valoración de aplicabilidad ..... válido

Leyenda:
01-13 improcedente
14-16 Aceptable con recomendación
17-20 Aceptable

lugar y fecha: 12.06.18

Firma y Post – Firma del Experto: .....

DNI: 40820337

I. DATOS PERSONALES:

1.1. APELLIDOS Y NOMBRES DEL INFORMANTE : Fernández Granados Freddy Elio

1.2. GRADO ACADÉMICO : Maestro en Docencia Universitaria y G.F.

1.3. INSTITUCIÓN DONDE LABORA : UNU.

1.4. TITULO DE LA INVESTIGACIÓN : MODELO PREDICTIVO UTILIZANDO TÉCNICAS DE MINERÍA DE DATOS PARA DETERMINAR LA IRREGULARIDAD ESTUDIANTIL EN LA E.A.P. DE ING. DE SISTEMAS EN LA UNU.

1.5. AUTOR DEL INSTRUMENTO : PORTUGUEZ DURAND, ENRIQUE IGLESIAS

1.6. NOMBRE DEL INSTRUMENTO : ENCUESTA

II. ASPECTOS A EVALUAR: (CALIFICACIÓN CUANTITATIVA)

INDICADORES DEL INSTRUMENTO DE EVALUACIÓN	CRITERIOS CUALITATIVOS CUANTITATIVOS	Deficiente (01-09)	Regular (10-13)	Bueno (14-16)	Muy bueno (17-18)	Excelente (19-20)
01 CLARIDAD	Está formulado con lenguaje apropiado				18	
02 OBJETIVIDAD	Está expresado con conductas observables			16		
03 ACTUALIDAD	Adecuado al avance de la ciencia y calidad				17	
04 ORGANIZACIÓN	Existe una organización lógica del instrumento				18	
05SUFICIENCIA	Valora los aspectos en cantidad y calidad				18	
06 INTENCIONALIDAD	Adecuado para cumplir con los objetivos			15		
07 CONSISTENCIA	Basado en el aspecto teórico científico del tema de estudios				18	
08 COHERENCIA	Entre las hipótesis, dimensiones e indicadores				17	
09 METODOLOGÍA	Las estrategias responden al propósito del estudio			16		
10 OPORTUNIDAD	Genera nuevas pautas para la investigación y construcción de teorías				18	
Sub Total				47	124	171
Total						

Valoración cuantitativa (total x0.4) 17

Valoración cualitativa ..... aceptable

Valoración de aplicabilidad ..... Valido

Leyenda:
01-13 improcedente
14-16 Aceptable con recomendación
17-20 Aceptable

Lugar y fecha: 14/06/2018

Firma y Post – Firma del Experto: .....

DNI: 40077294

ANEXO 7: BASE DE DATOS RECOLECTADA



Data_thesis.sql

ANEXO 8: DOCUMENTO DE AUTORIZACIÓN DE TRATAMIENTO DE DATOS DE LA UNU

UNIVERSIDAD NACIONAL DE UCAYALI
FACULTAD DE INGENIERÍA DE SISTEMAS E INGENIERIA CIVIL
DIRECCIÓN DE ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

AUTORIZACIÓN

Yo, Mg. Arturo Yupanqui Villanueva, director de la escuela académica profesional de ingeniería de sistemas.

Autorizo que el estudiante Enrrique Iglesias Portuguez Durand, de la carrera profesional de ingeniería de sistemas con código estudiantil 0002130355, utilice los datos académicos de los alumnos registrados en nuestra memoria de la Facultad de Ingeniería de sistemas e ingeniería civil, comprendidos entre los periodos 2015 – 2018, en su investigación de tesis titulada "MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS PARA LA IRREGULARIDAD ESTUDIANTIL EN LA ESCUELA ACADÉMICA PROFESIONAL DE INGENIERIA DE SISTEMAS DE LA UNU, 2018"

Por lo cual la dirección de escuela de ingeniería de sistemas dará las facilidades respectivas.

Se entrega la siguiente autorización para los fines que estimen por conveniente.

Pucallpa, 31 de octubre del 2018

Atentamente,

Ing. Mg. ARTURO YUPANQUI VILLANUEVA
Director Esc. Prof. Ing. Sistemas
Fac. de Ing. de Sistemas y de ing. Civil

ANEXO 9: CONSTANCIA POR PARTE DE LA ESCUELA ACADÉMICA PROFESIONAL DE INGENIERÍA DE SISTEMAS

UNIVERSIDAD NACIONAL DE UCAYALI
FACULTAD DE INGENIERÍA DE SISTEMAS E INGENIERIA CIVIL
DIRECCIÓN DE ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS

CONSTANCIA

El Director de la Escuela Profesional de Ingeniería de Sistemas, Arturo Yupanqui Villanueva, quien suscribe.

HACE CONSTAR:

Que, **ENRRIQUE IGLESIAS PORTUGUEZ DURAND**, identificado con DNI **N° 73372296,** con Código de estudiante **N° 0002130355**, ex alumno de la carrera profesional de ingeniería de sistemas, desarrolló su investigación de tesis titulada **"MODELO PREDICTIVO APLICANDO TÉCNICAS DE MINERÍA DE DATOS PARA LA IRREGULARIDAD ESTUDIANTIL EN LA ESCUELA ACADÉMICA PROFESIONAL DE INGENIERÍA DE SISTEMAS DE LA UNU, 2018"**, con datos académicos de la memoria de la Facultad de Ingeniería de Sistemas e Ingeniería Civil, para lo cual nos referimos a los resultados que demuestra su modelo predictivo en base a sus dimensiones **SOCIALES, ECONÓMICAS, PERSONAL Y EDUCACIONAL**, las cuales nos permite comprender de manera detallada el rendimiento académico de nuestra plana estudiantil.

Se expide la presente, a solicitud del (la) interesado (a) para los fines que estime por conveniente.

Pucallpa, 10 de mayo del 2019

Ing. Mg. ARTURO YUPANQUI VILLANUEVA
Director Esc. Prof. Ing. Sistemas
Fac. de Ing. de Sistemas y de ing. Civil