

**UNIVERSIDAD NACIONAL
JOSÉ FAUSTINO SÁNCHEZ CARRIÓN**



ESCUELA DE POSGRADO

TESIS

**PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL UTILIZANDO LA TÉCNICA
DE ÁRBOLES DE DECISIÓN EN LA ESCUELA DE POSGRADO DE LA
UNIVERSIDAD NACIONAL JOSÉ FAUSTINO SÁNCHEZ CARRIÓN**

PRESENTADO POR:

Brenda Nicole Diaz Landa

**PARA OPTAR EL GRADO ACADÉMICO DE MAESTRO EN INGENIERÍA DE
SISTEMAS**

ASESOR:

M(o). Erlo Wilfredo Lino Escobar

HUACHO - 2021

**PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL UTILIZANDO
LA TÉCNICA DE ÁRBOLES DE DECISIÓN EN LA ESCUELA DE
POSGRADO DE LA UNIVERSIDAD NACIONAL JOSÉ FAUSTINO
SÁNCHEZ CARRIÓN**

Brenda Nicole Diaz Landa

TESIS DE MAESTRÍA

ASESOR: M(o). Erlo Wilfredo Lino Escobar

**UNIVERSIDAD NACIONAL
JOSÉ FAUSTINO SÁNCHEZ CARRIÓN
ESCUELA DE POSGRADO
MAESTRO EN INGENIERÍA DE SISTEMAS
HUACHO
2021**

DEDICATORIA

La realización de la presente tesis la dedico a Dios por guiar mi camino. A mis abuelas, por siempre haberme apoyado y por el amor incondicional que me dieron. También a mis padres, hermana y demás familiares quienes me brindan cuidado y cariño.

Brenda Nicole Diaz Landa

AGRADECIMIENTO

Deseo agradecer a las autoridades de la Escuela de Posgrado de la Universidad Nacional José Faustino Sánchez Carrión quienes me permitieron el acceso necesario para aplicar el cuestionario de recolección de datos.

También agradezco a mi asesor, el M(o). Erlo Lino Escobar quien pudo guiarme desde el inicio a fin para la culminación del proyecto y el informe final de tesis para lograr este nuevo paso en la carrera profesional.

Brenda Nicole Diaz Landa

ÍNDICE

DEDICATORIA	iii
AGRADECIMIENTO	iv
RESUMEN	x
ABSTRACT	xi

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la realidad problemática	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 de la investigación	4
1.5 Delimitaciones del estudio	5
1.6 Viabilidad del estudio	5

CAPÍTULO II

MARCO TEÓRICO

2.1 Antecedentes de la investigación	6
2.1.1 Investigaciones internacionales	6
2.1.2 Investigaciones nacionales	10
2.2 Bases teóricas	13
2.3 Bases filosóficas	33
2.4 Definición de términos básicos	34
2.5 Hipótesis de investigación	35
2.5.1 Hipótesis general	35
2.5.2 Hipótesis específicas	36
2.6 Operacionalización de las variables	36

CAPÍTULO III

METODOLOGÍA

3.1 Diseño metodológico	38
3.2 Población y muestra	39
3.2.1 Población	39

3.2.2 Muestra	39
3.3 Técnicas de recolección de datos	39
3.4 Técnicas para el procesamiento de la información	40
CAPÍTULO IV	
RESULTADOS	
4.1 Análisis de resultados	42
4.2 Contrastación de hipótesis	64
CAPÍTULO V	
DISCUSIÓN	
5.1 Discusión de resultados	74
CAPÍTULO VI	
CONCLUSIONES Y RECOMENDACIONES	
6.1 Conclusiones	76
6.2 Recomendaciones	77
REFERENCIAS	78
7.1 Fuentes bibliográficas	78
7.2 Fuentes hemerográficas	79
7.3 Fuentes electrónicas	80
ANEXOS	82

ÍNDICE DE TABLAS

Tabla 1 Valor de interpretación de la fuerza de concordancia según el valor del coeficiente Kappa de Cohen.....	41
---	----

ÍNDICE DE FIGURAS

<i>Figura 1.</i> Procedimiento de modelado.	17
<i>Figura 2.</i> Proceso del Data Mining según Microsoft.	21
<i>Figura 3.</i> Árbol de decisión y sus características.	24
<i>Figura 4.</i> Modelo Spady de la deserción estudiantil.	26
<i>Figura 5.</i> Modelo Tinto de la deserción estudiantil.	26
<i>Figura 6.</i> Modelo de la deserción estudiantil en una universidad peruana.	28
<i>Figura 7.</i> Frontis de la Escuela de Posgrado de la UNJFSC.	30
<i>Figura 8.</i> Valoración del índice Kappa.	35
<i>Figura 9.</i> Encuestados según programa de maestría.	42
<i>Figura 10.</i> Deserción de estudiantes encuestados	43
<i>Figura 11.</i> Pregunta 1 – Su rendimiento académico durante la secundaria fue	44
<i>Figura 12.</i> Pregunta 2 – ¿Cuántas asignaturas reprobó en su educación secundaria?	45
<i>Figura 13.</i> Pregunta 3 – ¿Repetió algún año en la educación secundaria?	46
<i>Figura 14.</i> Pregunta 4 – Su rendimiento académico en la carrera de pregrado fue	47
<i>Figura 15.</i> Pregunta 5 – ¿Cuántos cursos desaprobó en su carrera de pregrado?	48
<i>Figura 16.</i> Pregunta 6 - Identifique su sexo	49
<i>Figura 17.</i> Pregunta 7 – Rango de edad	50
<i>Figura 18.</i> Pregunta 8 – Estado civil	51
<i>Figura 19.</i> Pregunta 9 – Percibe que su trabajo se encuentra relacionado con su carrera de pregrado	52
<i>Figura 20.</i> Pregunta 10 – Cantidad de hijos	53
<i>Figura 21.</i> Pregunta 11 – Ingreso familiar	54
<i>Figura 22.</i> Pregunta 12 – ¿Tiene la motivación suficiente para culminar sus estudios de posgrado?	55
<i>Figura 23.</i> Pregunta 13 – Considera que su situación económica actual es	56
<i>Figura 24.</i> Pregunta 14 – Considera que puede financiar sus estudios hasta finalizar la maestría	57
<i>Figura 25.</i> Pregunta 15 – Considera que la disponibilidad de tiempo para realizar sus actividades de estudiante es	58
<i>Figura 26.</i> Pregunta 16 – Considera que su nivel de estrés actual es	59
<i>Figura 27.</i> Pregunta 17 – ¿Considera que la escuela de Posgrado ofrece una infraestructura adecuada para la maestría?	60
<i>Figura 28.</i> Pregunta 18 – ¿Considera que la escuela de Posgrado cuenta con el equipamiento y mobiliario adecuados para la maestría?	61
<i>Figura 29.</i> Pregunta 19 - ¿Considera que la escuela de Posgrado imparte las asignaturas correctas para la realización de sus estudios de posgrado?	62
<i>Figura 30.</i> Pregunta 20 - ¿Considera que la escuela de Posgrado imparte las asignaturas adecuadas para la maestría?	63
<i>Figura 31.</i> Preparación de datos del cuestionario para ingreso como base de datos de entrenamiento al software Weka	65
<i>Figura 32.</i> Carga de datos preparados del cuestionario como base de datos de entrenamiento al software Weka	66
<i>Figura 33.</i> Vista 1 del análisis de la base de entrenamiento con el algoritmo J48	66
<i>Figura 34.</i> Vista 2 del análisis de la base de entrenamiento con el algoritmo J48	67

<i>Figura 35.</i> Vista del análisis de la base de entrenamiento con el algoritmo J48.....	68
<i>Figura 36.</i> Vista del análisis de la base de entrenamiento con el algoritmo J48.....	68
<i>Figura 37.</i> Árbol de decisión generado por el análisis de la base de entrenamiento con el algoritmo J48 por el software Weka.....	70
<i>Figura 38.</i> Ingreso de datos en el software SPSS v. 25.0 para el análisis de coeficiente Kappa en comparación de los casos de deserción reales y los predichos por el software Weka.....	72
<i>Figura 39.</i> Resultado del análisis de coeficiente Kappa por SPSS v. 25.0	73

RESUMEN

La presente tesis analiza el problema de determinar si existe una relación entre las variables de estudio deserción estudiantil y técnicas de árbol de decisión.

Se planteó como hipótesis general la afirmación que es posible aplicar la metodología de minería de datos denominada “árboles de decisión” para crear un modelo de predicción del comportamiento de la deserción estudiantil de la población de la Escuela de Posgrado de la UNJFSC, ello significó la construcción de un modelo de simulación utilizando el software Weka y la captura de datos en relación a los atributos de las circunstancias personales de los estudiantes; para ello se utilizó un cuestionario de 20 ítems el cual fue aplicado a la muestra de 237 estudiantes matriculados en el ciclo 2019-2. Además, se obtuvieron los datos de los matriculados a través de los registros académicos de la Escuela de Posgrado de la UNJFSC para determinar la deserción real de la muestra a la fecha del estudio.

Realizado el modelo, se lograron identificar los principales factores que afectan la deserción estudiantil según el análisis estadístico realizado por el software Weka, los cuales fueron graficados en un árbol de decisiones. Dicho modelo de simulación obtuvo una exactitud del 87,76% y una concordancia, medida a través del índice Kappa de Cohen de un valor de 0,6663, al cual le corresponde una valoración de “concordancia considerable” o “buena” según distintas aproximaciones teóricas.

Como principal conclusión se halló la demostración de la hipótesis general de la investigación al haberse encontrado una concordancia moderada entre el modelo de simulación y los registros de los casos de deserción real registrados por los estudiantes de la muestra a través del índice Kappa con un valor de 0,6663 (concordancia considerable) generado por el software Weka.

Palabras clave: deserción estudiantil, minería de datos, arboles de decisión, J48, Weka, predicción, modelo de simulación.

ABSTRACT

The research analyzes the existence of a correlational relationship between the study variables such as dropouts and decision tree technique through the creation of a model for simulation of the academic performance of master's students in educational programs of the Graduate School of the José Faustino Sánchez Carrion University.

As main hypothesis the possibility of generate a simulation model to predict the dropouts cases of master's students of the Graduate School, which means the generation of a simulation model using Weka software and analysis of data obtained from a 20 items quiz applied to 237 students of the sample compared to the real cases of dropouts from the academical records of the Graduate School.

With the generated model were identified factors affecting the academic performance according the statistical analysis made by the Weka software, those factors were show in a decision three diagram. The model's accuracy through correct prediction percentage was 87,76%, the consistency of the model was measured using the Cohen's Kappa indicator, obtaining an indicator's score of $k = 0,6663$ meaning a “moderate consistency” of the predicted values with the real academic performance.

The main conclusion of the research was the demonstration of the main hypothesis, finding enough statistical significance of model's accuracy and a value $k = 0,6663$ to the prediction model using Weka software.

Keywords: dropouts, decision trees, data mining, J48 algorithm, Weka, prediction model.

INTRODUCCIÓN

La presente investigación presenta el análisis sobre el problema de determinar si es posible realizar una predicción de la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la UNJFSC a través de la técnica de árboles de decisión, siendo relevante ya que permitiría a la Dirección de la Escuela de Posgrado obtener información importante para tomar decisiones en relación a la cantidad de alumnos que continuaría hasta la finalización de sus estudios, generando el correspondiente ingreso por pago de derechos académicos.

En el primer capítulo se describe el problema que la investigación trata, se detalla la problemática en que se encuentra la Escuela de Posgrado de la UNJFSC, la cual no cuenta con una herramienta para el pronóstico de la deserción estudiantil de las maestrías relacionadas con la especialidad de educación, las cuales son las que mantienen la mayor cantidad de alumnado de todos los programas de posgrado que ofrece la institución.

En el segundo capítulo se encuentra el marco teórico para la presente tesis, en donde se definen las variables y dimensiones tomadas en cuenta, además la información sobre la institución en que se realiza el presente estudio.

En el tercer capítulo se menciona la metodología de investigación que sigue la presente tesis, además de las técnicas e instrumento de recolección de datos y la forma del procesamiento de información realizado.

En el capítulo cuarto se muestran los resultados obtenidos en el análisis estadístico descriptivo por medio de gráficos estadísticos los cuales fueron generados en Microsoft Excel. Además, se muestra el resultado del análisis estadístico inferencial realizado para la prueba de hipótesis correspondiente, en el cual se generaron gráficos y tablas mediante el software SPSS en su versión 25.0.

En el capítulo quinto se puede observar la discusión, conclusiones y recomendaciones, por medio de las cuales se complementan, concretan e interpretan los resultados obtenidos en la investigación.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1 Descripción de la realidad problemática

A nivel global, las instituciones educativas en todos los niveles se enfrentan a cierto porcentaje de deserción estudiantil, ello no es diferente en el nivel de Posgrado, pues los motivos para que un estudiante abandone la carrera o programa iniciado son muy diversos y se pueden presentar en todas las etapas de la vida adulta. Es por ello que las instituciones educativas se encuentran cada vez más orientada a obtener y utilizar información relevante para su toma de decisiones, es por tal necesidad de información que se inventaron y difundieron los sistemas de información para diferentes niveles en las organizaciones desde los denominados sistemas transaccionales los cuales recopilan de primera mano la información a medida que se va creando, hasta los sistemas gerenciales de apoyo en la toma de decisión, los cuales brindan soporte al proceso de tomar decisiones estratégicas de la alta gerencia. En la última década, la utilización de dichos sistemas se ha masificado al grado que toda entidad educativa sea estatal o privada requiere un mínimo nivel de automatización de su información actual, sin embargo, con respecto a los datos personales de sus estudiantes, notas, historial, etc. Pese a ello, la utilización de sistemas que permiten mapear escenarios futuros o predecirlos todavía no se ha difundido.

Actualmente, a nivel de América Latina, las universidades estatales y privadas han empezado un proceso de mejora de la calidad del servicio educativo que ofrecen, sea por la competencia o por la renovación de métodos y técnicas de enseñanza, así como por las necesidades del mercado profesional que obligan a modernizar los currículos estudiantiles. Esto dio lugar a procesos de autoevaluación a fin de detectar las debilidades y fortalezas institucionales y generar acciones correctivas de las deficiencias encontradas.

En el Perú, debido a la actualización de su legislación, las universidades estatales buscan la acreditación de sus carreras que permitan la continuidad de sus servicios y se asegure la calidad profesional que se otorga a sus egresados. La Superintendencia Nacional de Educación Superior Universitaria (SUNEDU), es la entidad encargada de dicha supervisión y evaluación, ello con las facultades que le otorgó Ley Universitaria - Ley N° 30220 promulgada el año 2015. Es por ello que las universidades nacionales y privadas iniciaron un proceso de autoevaluación institucional con la finalidad de aprobar los estándares de calidad definidos por dicho ente regulador, sin embargo, tales esfuerzos se ven truncados ante la ocurrencia de la deserción estudiantil, la cual significa que la inversión del Estado traducida en la contratación de los bienes y servicios necesarios para implementar una vacante en las carreras profesionales se desperdicia, pues el estudiante no llega a finalizar la carrera o programa estudiantil. Tal deserción se produce por diversos motivos y algunos de ellos recurrentemente son la falta o carencia de personal docente y administrativo calificado, así como por deficiencias de infraestructura o equipamiento, todo ello relacionado a la asignación de recursos económicos dotado a la universidad estatal por parte del Gobierno Central, el cual correspondió a un 3.8% del Producto Bruto Interno (PBI) del Perú para el año 2019, según la Agencia Peruana de Noticias (2019), sin embargo, la gran mayoría de universidades públicas consideran que dicho monto resulta ser insuficiente para cubrir todas las brechas en infraestructura y personal docente o administrativo.

A nivel local, la Universidad José Faustino Sánchez Carrión también se ve perjudicada por una insuficiente asignación de recursos, sin embargo, en el caso de estudio particular de la Escuela de Posgrado, cuenta con cierta autonomía en cuanto a ingresos, pues además de la asignación presupuestal ordinaria cuenta con una fuente de ingreso de recursos propios, los cuales corresponden a los pagos por matrícula, pensiones mensuales que abonan los estudiantes de los programas de maestrías y doctorados que desarrolla, pagos por los trámites de los estudiantes y egresados. Además, esta casa de estudios tampoco se ve ajena al fenómeno de la deserción estudiantil, el estudio de Claros, Cipriano y Ramirez (2016) quienes realizaron un estudio sobre el índice de riesgo de deserción estudiantil en la Facultad de Ciencias de la Universidad José Faustino Sánchez Carrión concluyendo que el riesgo promedio de deserción era de 20,84 de abandonar la carrera y un 19,17 de abandonar la Universidad (p. 28).

Todo ello pese a que, según informa Gestión (2018) “Uno de los requisitos que la Sunedu contempla para otorgar el licenciamiento a una universidad es que al menos el 25% de sus docentes tengan estudios de posgrado y trabajen a tiempo completo” (p. 1) lo cual permite inferir una creciente demanda en las maestrías del sector educación, sin embargo, todavía existe un riesgo de deserción en la Escuela de Posgrado de la Universidad Nacional José Faustino Sánchez Carrión.

Por todo lo expuesto considero que el generar un modelo de predicción del comportamiento de la deserción estudiantil en la Escuela de Posgrado de la UNJFSC a través de la técnica de “árboles de decisión” permitirá una mayor información a la Dirección de la Escuela de Posgrado con el objeto de planificar sus vacantes, tomar medidas orientadas específicamente a la población susceptible y evitar que se incremente o produzca la deserción de su cuerpo estudiantil.

1.2 Formulación del problema

1.2.1 Problema general

¿Es posible predecir la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la UNJFSC a través de la técnica de minería de datos de árboles de decisión?

1.2.2 Problemas específicos

¿Qué factores permiten predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC?

¿Cuál es la exactitud de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC?

¿Cuál es la medida de concordancia de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC?

1.3 Objetivos de la investigación

1.3.1 Objetivo general

Determinar si es posible aplicar la metodología de minería de datos denominada “árboles de decisión” para crear un modelo de predicción del comportamiento de la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la UNJFSC.

1.3.2 Objetivos específicos

Determinar los factores que permiten predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.

Determinar la exactitud de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.

Calcular la medida de concordancia de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.

1.4 Justificación de la investigación

El presente trabajo de investigación se encuentra justificado en la importancia de mantener una mayor información para la toma de decisiones de la Dirección de la EAP de Posgrado de la UNJFSC en relación al riesgo de deserción estudiantil, ello debido a que la deserción trae consigo una pérdida de recursos de inversión del Estado y la institución educativa que fue necesaria para la contratación de los docentes, infraestructura y servicios que brindó al estudiante que deserta. Esta información predictiva permitirá la formulación de acciones correctivas orientadas a evitar que se eleven los eventos de deserción en los estudiantes coadyuvando a los mismos a la motivación e implementación de circunstancias favorables para que concluyan su programa de maestría en la Escuela de Posgrado de la UNJFSC.

1.5 Delimitaciones del estudio

1.5.1. Delimitación temporal

El trabajo de investigación se realizó en el mes de abril del año 2021.

1.5.2. Delimitación espacial

Se realizó en las oficinas administrativas de la Escuela de Posgrado de la Universidad Nacional José Faustino Sánchez Carrión.

1.5.3. Delimitación social

La tesis se centró en determinar si es posible predecir la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la Universidad Nacional José Faustino Sánchez Carrión.

1.6 Viabilidad del estudio

Para la realización del trabajo de investigación se contaron con los conocimientos teóricos, obtenidos de una formación profesional en Ingeniería de Sistemas, así como los recursos disponibles y acceso para adquirir la información necesaria.

También se contó con la autorización de la Dirección de la Escuela de Posgrado para aplicar el instrumento de medición de la presente tesis lo que me permitió realizar la investigación sin inconvenientes.

CAPÍTULO II

MARCO TEÓRICO

2.1 Antecedentes de la investigación

2.1.1 Investigaciones internacionales

Azoumana (2013), en su trabajo de investigación titulado “Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos”. Objetivo: fue planteado como objetivo general el de analizar los motivos de la deserción estudiantil en la población elegida en relación a la facultad y universidad elegida de Colombia. Metodología: se realizó un estudio descriptivo en el cual se analizó a través del software Weka los datos recopilados de una muestra de 707 estudiantes pertenecientes desde el primer al décimo semestre del programa de Ingeniería de Sistemas, entre los períodos académicos. Resultados: conforme lo publicado, esta investigación halló que se puede destacar que la principal causa de deserción de los estudiantes, de acuerdo a los parámetros establecidos como causales definidas en 5 variables permite afirmar que la causa de la deserción es el factor indeterminado, seguido por dificultades financieras y por otros intereses. Conclusiones: la principal conclusión del estudio es el determinar que las técnicas de minería de datos permiten un conocimiento general del tema y que pueden ser utilizadas para desarrollar trabajos futuros en otras áreas de conocimiento.

Amaya, Barrientos, & Heredia (2014), en su investigación titulada “Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos” desarrollan la construcción de un modelo predictivo de deserción estudiantil, caracterizando a los estudiantes de la Universidad Simón Bolívar de Colombia con el objetivo de poder predecir la probabilidad de deserción de los estudiantes; dicho modelo demostró el desempeño de los algoritmos

presentados para clasificar datos bajo contextos variables y la precisión de uno con respecto al otro. Para la creación del modelo se utilizó la herramienta WEKA que permite de forma muy eficiente el procesamiento y clasificación de los datos con resultados satisfactorios.

Eckert y Suénaga (2015), en su investigación “Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos” desarrollado para la Universidad Gastón Dachary, Departamento de Ingeniería y Ciencias de la Producción en Argentina. Objetivo: identificar factores que influyen sobre la deserción de los estudiantes de la carrera de Ingeniería en Informática de la Universidad Gastón Dachary en Argentina, mediante la aplicación de una técnica de minería de datos. Metodología: el estudio es descriptivo y utiliza la técnica de datos documental, pues analiza los datos proporcionada por los alumnos al ingreso de la universidad en relación a sus antecedentes personales y educativos. Posteriormente utiliza el software Weka para clasificar los antecedentes asignaturas aprobadas, cantidad y resultado de asignaturas cursadas, procedencia y edad de ingreso del estudiante utilizando los algoritmos J48 (C4.5), BayesNet (TAN) y OneR analizando la situación académica de alumnos universitarios en base a los datos de su trayectoria en la UGD. La muestra sobre la cual se trabajó corresponde a los estudiantes de la carrera de Ingeniería en Informática, modalidad presencial, 5 años de duración y la tesis de grado. El período seleccionado para el estudio corresponde a los estudiantes ingresados desde el año 2000 al 2009, totalizando 855 casos analizados. Resultados: se presentan en forma de esquema y gráficamente. A partir de los atributos de entrada, se obtiene como resultado una serie de condiciones representadas de forma escrita mediante un conjunto de reglas, condiciones del tipo si-sino (if-else) y gráfica mediante un árbol de decisión. En la Figura 1 se puede apreciar el conjunto de reglas generado para clasificar los casos de deserción (“Des”) y permanencia (“NoDes”). El nodo o condición inicial representa la cantidad de exámenes finales aprobados correspondientes al primer año de la carrera (“1°Apro”), donde se dividen en dos subclasificaciones, una para cantidades de materias aprobadas menores o iguales a siete y para las mayores a siete. Conclusiones Las herramientas de minería de datos brindan resultados que deben ser interpretados y traducidos a diagnósticos y consecuencias del ámbito real (en este caso la universidad). Esto implica que los resultados de la aplicación de las técnicas se han utilizado para explicar parte del comportamiento de la situación en cuanto a la permanencia y su determinación a partir del desempeño académico de los estudiantes. Las posibles consecuencias y acciones tendientes a la toma de decisiones específicas, está sujeta a consideraciones de otros integrantes del

cuerpo académico de la institución educativa. En el análisis de los resultados obtenidos de los algoritmos de clasificación C4.5 (J48), TAN (BayesNet) y OneR, se pudo observar porcentajes de aciertos similares, sin embargo, no identifican exactamente los mismos atributos.

Aguirre, Valdovinos, Antonio, Alejo y Marcial (2015), en su investigación titulada “Análisis de deserción escolar con minería de datos” desarrollado en la Universidad de Ixtlahuaca CUI en México. Objetivo: plantearon como principal objetivo el identificar las causas de la deserción escolar en la carrera de ingeniería en computación de la Universidad de Ixtlahuaca CUI (UICUI) a través de técnicas de minería de datos. Metodología: el estudio de nivel descriptivo aplicó el proceso general de minería de datos, utilizando una muestra de 497 estudiantes del CUI, utilizando la técnica de la encuesta a través de un cuestionario de seis ítems reactivos que permiten identificar las posibles causas que pudieran propiciar deserción en los estudios universitarios para luego clasificar los datos utilizando el programa J48 correspondiente al algoritmo C4.5. Resultados: luego del análisis de datos señala que un 46% de los alumnos trabaja en el cual solo un 7% es para solventar sus propios gastos, tomando en cuenta que si los alumnos también tienen que aportar dinero a su familia casi a mediados de semestre abandonarían la carrera por seguir ayudando económicamente a su familia. No obstante, en la mayoría del total de los alumnos los padres son quienes solventan sus gastos, al parecer con independencia del hecho de que trabajen o no. Conclusiones: el estudio concluye que los alumnos que caen en la deserción tienen como principal causa que la carrera no era lo que ellos esperaban, y no cumple sus expectativas a futuro.

Cuji, Gavilanes, & Sanchez (2017), quienes en su investigación titulada “Modelo predictivo de deserción estudiantil basado en árboles de decisión” muestran la construcción de un modelo predictivo de deserción estudiantil, para pronosticar la probabilidad, que un estudiante abandone su programa académico, mediante técnicas de clasificación, basadas en árboles de decisión. La metodología utilizada, se basa en Knowledge Discovery in Database (KDD), con cinco etapas: selección, procesamiento, transformación, minería de datos y evaluación. Aplicando el algoritmo, Classification and Regression Tree (CART) de la herramienta R, se construyó un árbol con cuatro niveles de profundidad y mismo número reglas, que evalúan a los posibles desertores. Llevando a concluir que las variables nivel y notas tienen mayor influencia en la deserción.

Sadic, Abdulaziz, Fadl, & Najoua (2018), en su investigación “Educational Data Mining and Analysis of Students Academic Performance Using WEKA” – Data Mining educacional y análisis del rendimiento académico estudiantil usando WEKA. Objetivo general: determinar los factores más influyentes en el rendimiento académico de la población de estudio. Metodología: los autores utilizaron un cuestionario que analizaba un total de 24 factores influyentes en el rendimiento académico, analizaron los datos con el algoritmo J48, PART, Árboles de decisión y redes bayesianas con el software WEKA con la finalidad de determinar la precisión de cada algoritmo analizado. Población y muestra: los autores plantearon un estudio de datos socioeconómicos y demográficos de una muestra 300 estudiantes de tres universidades de la ciudad de Assam – India. Resultados: la investigación obtuvo como principal resultado el determinar que el algoritmo de árboles de decisión fue definitivamente el más preciso pues logró una precisión del 99%, seguido del PART con 74.33%, el J48 con un 73% y de redes bayesianas 65.33%.

Timarán, Caicedo e Hidalgo (2019), en su trabajo de investigación “Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°”. Objetivo general: determinar los factores que se asocian al desempeño académico de estudiantes colombianos de educación secundaria. Metodología: estudio descriptivo con un enfoque cuantitativo, a través de aplicar un diseño no experimental, para ello utilizó la metodología CRISP-DM y utilizó las bases de datos del ICFES la información socioeconómica, académica e institucional de estos estudiantes, para luego diseñar un modelo de predicción usando la herramienta de minería de datos WEKA a través de árboles de decisión con los que identificó patrones asociados al buen o mal desempeño académico de los estudiantes en las pruebas. Población y muestra: la población estuvo determinada por 1.061.680 estudiantes colombianos de grado undécimo de educación media, que presentaron las pruebas Saber 11° en los años 2015 y 2016, tomándose de ellos una muestra poblacional. Resultados: Como conclusión refieren que sí es posible la generación de modelos de predicción consistentes con la realidad y que cuente con el respaldo teórico científico, ello a través del análisis de datos almacenados en las bases de datos a través de WEKA, determinando que los atributos con mayor ganancia de información que forman parte de los patrones descubiertos, asociados a un bajo desempeño académico en las pruebas Saber 11°, están: el estrato socioeconómico bajo, utilizando un índice TIC bajo y el nivel SISBEN 1.

2.1.2 Investigaciones nacionales

Manco (2015), en su investigación publicada para la Universidad Nacional Tecnológica de Lima Sur (Untels) en la ciudad de Lima, la cual tituló “Modelo predictivo para la identificación de patrones de la deserción estudiantil en la Untels”. Objetivo: planteó como objetivo general el determinar los patrones del entorno que impactan en la deserción de los estudiantes de la Universidad Nacional Tecnológica de Lima Sur (Untels). Metodología: para la realización del estudio utilizó una ficha de datos socioeconómica y académica de los estudiantes de la cohorte 2007-I a 2011-I, que incluye los casos de deserción como variable dependiente. Se elaboran seis modelos utilizando el operador Decision Tree de RapidMiner, con y sin validación cruzada, y con parámetros modificados y la implementación de la herramienta Weka, W-J48. Resultados: luego de la construcción del modelo se logró determinar que las siguientes características o patrones del entorno que impactan en la deserción de los estudiantes de la Untels: Número de matrículas en los cuatro semestres consecutivos a su ingreso, Promedio en su segunda matrícula, Edad de ingreso, Promedio en su cuarta matrícula, Año de ingreso, Número de personas dependientes, Semestre de ingreso y Número de cursos aprobados en su primera matrícula. Conclusiones: entre sus principales conclusiones detalla que logró un 90.10% de clasificación correcta, con una desviación estándar de 2.08%. El principal patrón detectado para los desertores, es que el número de matrículas en los cuatro semestres consecutivos a su ingreso sea menor o igual que 3, con una precisión de 88%.

Gonzales y Rodriguez (2017), en su tesis de maestría titulada “Propuesta de un Modelo de Business Intelligence para Identificar el perfil de deserción estudiantil en la Universidad Científica del Sur”. Objetivos: su objetivo principal fue definido como el de aplicar técnicas de inteligencia de negocios y minería de datos para generar un modelo predictivo que permita correlacionar las variables que afectan el proceso de deserción estudiantil en la Universidad Científica del Sur con información basada en datos recopilados en la matrícula del ciclo 2016-2. Metodología: el estudio planteado fue descriptivo y se utilizaron técnicas de minería de datos para elaborar un modelo de predicción a través del software Weka. Resultados: la investigación determina que Se ha identificado los factores de riesgo como: Carrera, Semestre Cursado, Materias cursadas, Modalidad de Ingreso, Materias perdidas, Horario o Programación, Promedio, Edad, Ciudad de procedencia, Domicilio, Sexo, Estado Civil, Nivel de estudios del Padre, Nivel de estudios de la Madre, Ingresos, entre otros.

Conclusiones: existe la posibilidad de deserción de la carrera de Administración de Negocios Internacionales, cuando los cursos son de Turno Noche. En el caso de Artes Escénicas y Teatro, la probabilidad de deserción se da en los alumnos que se encuentran matriculados en 1 curso, y de turno mañana. En el caso de la carrera de Ingeniería de Sistemas, existe la probabilidad de deserción cuando se matriculan en 1 curso en las tardes. Existe la probabilidad de deserción en la carrera de Medicina Humana, cuando se encuentra matriculado en 3 o menos cursos, y se da en el Turno Tarde.

Alania (2018), quien en su trabajo de investigación “Aplicación de técnicas de minería de datos para predecir la deserción estudiantil de la Facultad de ingeniería de la Universidad Nacional Daniel Alcides Carrión” presentado para la Universidad Nacional Daniel Alcides Carrión de Pasco planteó como Objetivo: aplicar la minería de datos para predecir la deserción estudiantil en la Facultad de Ingeniería de la Universidad Nacional Daniel Alcides Carrión. Metodología: es de tipo aplicado, nivel descriptivo; la población estuvo constituida por el total de 1810 alumnos matriculados en el periodo 2017-A en la Facultad de Ingeniería, determinándose una muestra no aleatoria por conveniencia de un total de 218 de alumnos de la Escuela de Sistemas y computación. Resultados: el estudio muestra que existe una diferencia significativa entre la deserción estudiantil y promedio de notas, a un nivel de confianza del 95% y nivel de significancia del 5%. Por tanto, se concluye que promedio de notas influye significativamente en la deserción estudiantil de los alumnos de la escuela de formación profesional de Sistemas y Computación de la UNDAC. Conclusiones: la mayor cantidad de estudiantes que abandonan la carrera es por un bajo rendimiento académico, así como también que la mayoría de estudiantes que desertan son los que aprobaron menos de 24 cursos, también concluye que la eficiencia de los algoritmos de árboles de decisión C4.5 (J48) y RandomTree son muy similares en función de su precisión.

Yamao (2018), en su investigación titulada “Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la escuela profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres, Lima-Perú”. Objetivo general: realizar la predicción del rendimiento académico de los alumnos que ingresaron a la Escuela Profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres en el primer ciclo utilizando minería de datos. Metodología: enfoque cuantitativo y de nivel explicativo y correlacional; de diseño transeccional del tipo correlacional- causal pues busca describir la relación que existe entre

el rendimiento académico y los factores social, económico y académico de los ingresantes que se utilizaron para probar la validez de la predicción del rendimiento académico. Población y muestra: la población del presente estudio son los estudiantes ingresantes a la carrera de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres, de los cuales se obtuvo una muestra probabilística con un intervalo de confianza de 95%, para un total de 1304 ingresantes. Resultados: se realizaron predicciones a través de tres técnicas: regresión lineal, árbol de decisiones y support vector machines, y el mejor resultado de 82.87% se obtuvo utilizando árbol de decisiones. De los diferentes factores, los que más influyeron en el rendimiento académico fueron los siguientes: nota de examen de admisión, género, edad, modalidad de ingreso y distancia desde su casa hasta el centro de estudios. Utilizando minería de datos fue posible realizar predicciones del rendimiento académico de los ingresantes.

Choque (2019), en su tesis “Minería de datos aplicada a la identificación de factores de deserción universitaria en programas de pre grado” presentado para la Universidad Nacional de San Agustín de Arequipa. Objetivo: planteó como aplicar la técnica clustering con el algoritmo ‘K means’ de minería de datos para la identificación de factores de deserción universitaria en programas de pre grado. Metodología: se considera como población un total de 978 alumnos pertenecientes a las universidades particulares de la ciudad del Cusco con modalidad presencial y la muestra fue definida por muestreo no probabilístico denominado muestreo por cuotas. La investigación fue de nivel descriptivo y enfoque cuantitativo, utilizando para ello las técnicas de entrevista y revisión documental; además se utiliza un cuestionario para la recolección de datos. Resultados: en el estudio se elaboró un modelo con la técnica Clustering y el algoritmo K means, el cual fue sometido a cinco pruebas variando la cantidad de clústeres o grupos para evaluar su validez a través de los campos relacionados a los factores potenciales de deserción. Conclusiones: esta investigación identificó como factores para la deserción estudiantil el Plan de Estudios, semestre cursado y distrito al que pertenece.

Mamani (2019), en su tesis titulada “Modelo de minería de datos basado en factores asociados para la predicción de deserción estudiantil universitaria” presentada para la Universidad Nacional de Moquegua. Objetivo: esta investigación planteó como objetivo general el desarrollar e implementar un modelo minería de datos para la predicción de deserción estudiantil universitaria basado en factores asociados. Metodología: la

investigación fue planteada de tipo descriptivo, definiendo como población a los estudiantes matriculados en la Universidad Nacional de Moquegua - Filial Ilo, tomándose como muestra los matriculados del primer ciclo académico en el periodo académico 2017 - 2 de la Universidad Nacional de Moquegua pertenecientes a las carreras profesionales de Ingeniería de Sistemas e Informática, Ambiental y Pesquera. Resultados: Se elaboró un modelo basado en redes neuronales el cual aprende de un cambio en las funciones de activación genera como mínimo el 91% de precisión con una optimización ADAM y obtiene una precisión del 89% en data nueva, mientras que con SGD y RMSprop obtienen entre 60% y 78% respectivamente. Conclusiones: como principal conclusión el estudio declara que se desarrolló e implementó un modelo minería de datos aplicando la hiper parametrización para observar el comportamiento de la ANN en la predicción de deserción estudiantil universitaria, se identificó variables de los factores asociados que aportaron en el ciclo del proyecto.

2.2 Bases teóricas

2.2.1. Minería de datos

La minería de datos puede ser definida como la extracción de información interesante, considerando la palabra “interesante” como no trivial, implícita, previamente desconocida y potencialmente útil. Según la definición que otorga Microsoft (2019):

Data mining is the process of discovering actionable information from large sets of data. Data mining uses mathematical analysis to derive patterns and trends that exist in data. Typically, these patterns cannot be discovered by traditional data exploration because the relationships are too complex or because there is too much data. – *La minería de datos es el proceso de descubrir información útil desde enormes estructuras de datos. La minería de datos utiliza el análisis matemático y estadístico para deducir los patrones y tendencias que existen en los datos. Típicamente, estos patrones no se pueden detectar mediante la exploración tradicional de los datos ya que las relaciones son demasiado complejas o porque existe demasiada data* – traducción propia. (Microsoft, 2019)

Su definición resalta entonces que se ocupa del tratamiento de grandes cantidades de datos para generar conocimiento, es decir, información útil para un propósito. Generalmente se utiliza en escenarios donde una extracción de información deba realizarse obligatoriamente de forma automatizada, pues resulta impracticable realizarla de manera manual. A la minería de datos se le conoce también como “extracción de conocimiento en bases de datos”, “análisis de patrones y datos”, “machine learning”, “data mining” y otros.

Para su desarrollo se utilizan técnicas estadísticas y de inteligencia artificial (algoritmos) para descubrir patrones e irregularidades en los grandes volúmenes de datos. Camborda (2014) realiza un resumen de sus principales características:

Para terminar con la definición de la minería de datos, podemos enumerar también algunos de sus usos:

Descubre hechos y relaciones de datos.

Se necesita poca intervención humana.

Encuentra patrones.

Determina y establece reglas.

Almacena y reutiliza reglas.

Presenta información a los usuarios.

Puede llevar muchas horas.

El usuario final debe ser capaz de analizar resultados. (p. 13)

Su aparición se encuentra muy relacionada al concepto de dinámica y modelamiento de sistemas, que considero indispensable para continuar con la descripción de lo que es data mining.

Dinámica y Modelamiento de Sistemas

La Dinámica de Sistemas fue desarrollada en los años cincuenta en el MIT (Massachusetts Institute of Technology) por el ingeniero Jay Wright Forrester, quien en sus tres obras más trascendentes: “Industrial Dynamics” la cual estudia diversos sistemas comerciales y de gestión como el control de inventarios, la logística y la toma de decisiones; “Urban Dynamics” en el que estudia los problemas de las sociedad urbana, como el hacinamiento y el deterioro de las ciudades; y “World Dynamics” analiza problemas como el crecimiento demográfico y la contaminación a escala global.

Según el enfoque de la Teoría General de Sistemas, para abordar un problema complejo, entender su comportamiento y estimar como cambiará en el futuro, es necesario analizar las partes que lo componen y determinar las relaciones que existen entre ellos. Para llevar a cabo este proceso de abstracción existe una metodología llamada Dinámica de Sistemas, a partir de la cual se obtiene un modelo que representa la realidad del problema y permite analizar su estructura y comportamiento.

Existen muchas definiciones al respecto, sin embargo, la del autor Santa Catalina (2010):

La Dinámica de Sistemas es una metodología para el estudio y manejo de sistemas de realimentación complejos. Una de las características de esta disciplina es el uso del computador para realizar sus simulaciones, lo que ofrece la posibilidad de estudiar el comportamiento y las consecuencias de las múltiples interacciones de los elementos de un sistema a través del tiempo. Esto la hace muy útil para el estudio de fenómenos sociales ya que en ellos están implicados una gran cantidad de elementos e interrelaciones en los que la presencia de no linealidades determina el comportamiento y dificultan una solución analítica. Además, los efectos de las políticas y acciones ejercidas sobre estos sistemas se manifiestan en horizontes temporales diferentes y dilatados. Este hecho dificulta la construcción de laboratorios de experimentación donde se puedan probar diferentes políticas y observar sus consecuencias sobre el sistema. Por tanto, los modelos de simulación dinámica permiten estudiar cómo las políticas, decisiones, estructura y retrasos influyen en el crecimiento y la estabilidad de un sistema. Actualmente su ámbito de aplicación abarca la planificación y diseño de políticas corporativas, la gestión y las políticas públicas, los modelos biológicos y médicos, el área de la energía y el medio ambiente, el desarrollo de la teoría en ciencias naturales y sociales, la toma decisiones y la dinámica no lineal compleja. (pág. 56)

El gran aporte Forrester consistió en transferir el conocimiento de la teoría de control y realimentación de la Ingeniería a otras áreas como la organización y las ciencias sociales, proponiendo una sencilla metáfora hidrodinámica para la representación de un sistema, abstrayendo ecuaciones diferenciales que definen un sistema no lineal. Es decir, la Dinámica de Sistemas representa matemáticamente nuestros modelos mentales. Las principales aplicaciones de software para la Dinámica de Sistemas existentes en la actualidad son los que simulan el correspondiente modelo matemático por medio de métodos numéricos computacionales facilitando el análisis de su comportamiento e incertidumbre a través de una interfaz gráfica.

En Dinámica de Sistemas, la simulación permite obtener trayectorias para las variables incluidas en cualquier modelo mediante la aplicación de técnicas de integración numérica. Sin embargo, estas trayectorias nunca se interpretan como predicciones, sino como proyecciones o tendencias, a través de la elaboración de un modelo.

Un modelo es una abstracción de la realidad que captura la esencia fundamental del sistema, con el detalle suficiente como para que pueda utilizarse en la investigación y experimentación en lugar del sistema real, con menos riesgo, tiempo y coste.

Conforme expresa Osorio Calderón (2010): “La ventaja de la Dinámica de Sistemas consiste en que estas acciones pueden ser simuladas a bajo coste, con lo que es posible valorar sus resultados sin necesidad de ponerlas en práctica sobre el sistema real” (p. 46).

Conforme lo define Rivera Castellanos (2010) existen tres formas de poder escribir un modelo para que sea comprendido por alguien más:

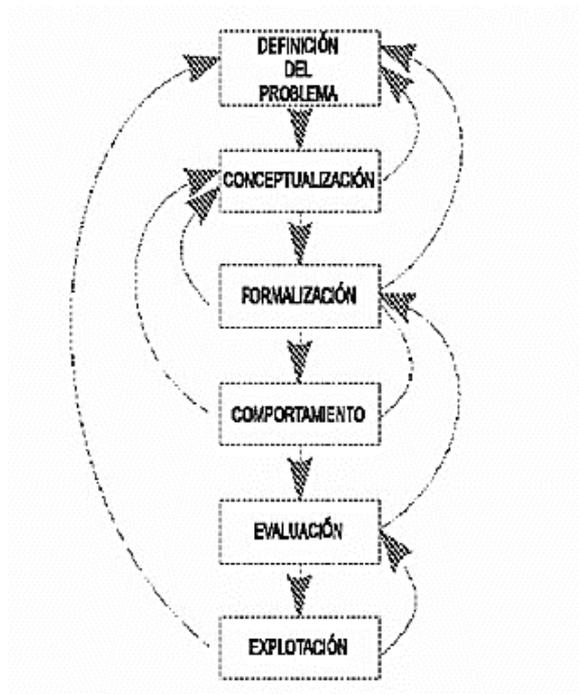
Un modelo verbal se describe mediante palabras que hagan alusión a los componentes y al funcionamiento de un modelo dado. Es la forma más común de explicar un modelo, ya que no necesita ningún otro lenguaje especial para entenderlo.

Un modelo matemático, utiliza las matemáticas como herramienta para explicar un fenómeno, sus partes y predecir su comportamiento. Por lo tanto, los modelos matemáticos son utilizados para explicar formalmente un modelo.

Un modelo grafico es aquel que explica el comportamiento de un fenómeno mediante diagramas, utilizando en parte descripciones verbales y matemáticas. De esta forma con un modelo grafico se puede crear un nexo entre el autor y su audiencia, así como entre los modelos matemáticos y verbales. (pág. 5)

La metodología de la Dinámica de Sistemas difiere de otras técnicas de modelado. En un modelo sistémico la estructura del mismo no está predeterminada por un tipo de modelo matemático previo, sino que la establece un analista dialogando con un experto. Esto le da al modelo un componente heurístico que hace que el modelo se base en el modelo mental que posee el experto sobre el problema. El modelo resultante, aunque al final se traduce en un conjunto de ecuaciones matemáticas, tiene su origen en un punto de vista, con toda la subjetividad que ello implica.

Pese a lo expuesto anteriormente, para la construcción del modelo del presente trabajo de investigación se tomará la secuencia o fases del proceso de modelamiento a través de la dinámica de sistemas conforme lo publicado en la web especializada Weebly, definidas conforme la siguiente figura:



Fuente: (WEEBLY, 2016)

Figura 1. Procedimiento de modelado.

Dichas fases son definidas por (WEEBLY, 2016) como:

Definición del problema. En esta primera fase se trata de definir claramente el problema y de establecer si es adecuado para ser descrito con los útiles sistémicos que hemos desarrollado. Para ello el problema debe ser susceptible de ser analizado en elementos cuya variación a lo largo del tiempo queremos estudiar.

Conceptualización del sistema. En esta segunda fase se trata de acometer dicho estudio, definiendo los distintos elementos que integran la descripción, así como las influencias que se producen entre ellos. El resultado de esta fase es el establecimiento del diagrama de influencias del sistema.

Formalización. En esta fase se pretende convertir el diagrama de influencias, alcanzado en el anterior, en el de Forrester. A partir de este diagrama se pueden escribir las ecuaciones del modelo (algunos entornos informáticos permiten hacerlo directamente). Al final de la fase se dispone de un modelo del sistema programado en un computador.

Comportamiento del modelo. En esta fase se somete el modelo a una serie de ensayos y análisis para evaluar su validez y calidad. Estos análisis son muy variados y comprenden desde la comprobación de la consistencia lógica de las hipótesis que incorpora hasta el estudio del ajuste entre las trayectorias generadas por el modelo y las registradas en la realidad, Así mismo, se incluyen análisis de sensibilidad que permiten determinar la sensibilidad del modelo, y por tanto, de las conclusiones que se extraigan de él, con relación a los valores numéricos de los parámetros que incorpora o las hipótesis estructurales.

Explotación del modelo. En esta última fase el modelo se emplea para analizar políticas alternativas que pueden aplicarse al sistema que se está estudiando. Estas políticas alternativas se definen normalmente mediante escenarios que representan las situaciones a las que debe enfrentarse el usuario del modelo. (p. 2)

En la actualidad se cuenta en el mercado con diversas herramientas de software para la programación o elaboración de modelos de simulación, de las cuales el autor Osorio Calderón (2010) menciona:

3.3.1 Professional DYNAMO. Es el más clásico de los lenguajes. No presenta posibilidades de modelado mediante iconos, pero sin embargo permite tratar ecuaciones de gran dimensión. La mayor parte de los modelos que se encuentran en los libros clásicos de la dinámica de sistemas están escritos en este lenguaje.

3.3.2 STELLA y i-think. Son entornos informáticos de amplia capacidad interactiva que permiten construir modelos empleando procedimientos gráficos, mediante iconos. Ambos poseen una estructura similar, pero mientras el primero se encuentra más orientado hacia usos académicos el segundo lo hace hacia aplicaciones profesionales. Ambos permiten construir los diagramas de Forrester en la pantalla del computador, de modo que al establecer su estructura se generan las ecuaciones. Se pueden agrupar elementos en sus modelos, y posee un zoom que permite desenvolverse con modelos complejos.

3.3.3 PowerSim. Entorno de características análogas a los anteriores (mientras aquellos son americanos, este es europeo — en concreto noruego). Permite desarrollar varios modelos simultáneamente, e interconectarlos posteriormente entre sí.

3.3.4 VenSim. Con respecto a las anteriores presenta algunas ventajas con relación a la organización de datos y a posibilidades de optimización. Se trata de un lenguaje muy potente para el desarrollo de modelos que pueden emplearse tanto en entornos PC como en Unix. Permite documentar automáticamente el modelo según se va construyendo, y crea árboles que permiten seguir las relaciones de causa efecto a lo largo del modelo. Está dotado de instrumentos para realizar análisis estadísticos.

3.3.5 Mosaikk-SimTek. Mosaikk es una herramienta muy sofisticada para PC, que conecta directamente al SimTek, que es un lenguaje de modelado tipo DYNAMO que posee una gran versatilidad. (pp. 46-47)

Proceso de la Minería de datos

Para determinar los pasos o proceso de data mining se tienen diversas aproximaciones, para la presente tesis tomaré los pasos recopilados por Camborda (2014):

1. Selección del conjunto de datos, tanto en lo que se refiere a las variables objetivo (aquellas que se quiere predecir, calcular o inferir), como a las variables independientes (las que sirven para hacer el cálculo o proceso), como posiblemente al muestreo de los registros disponibles.

2. Análisis de las propiedades de los datos, en especial los histogramas, diagramas de dispersión, presencia de valores atípicos y ausencia de datos (valores nulos).

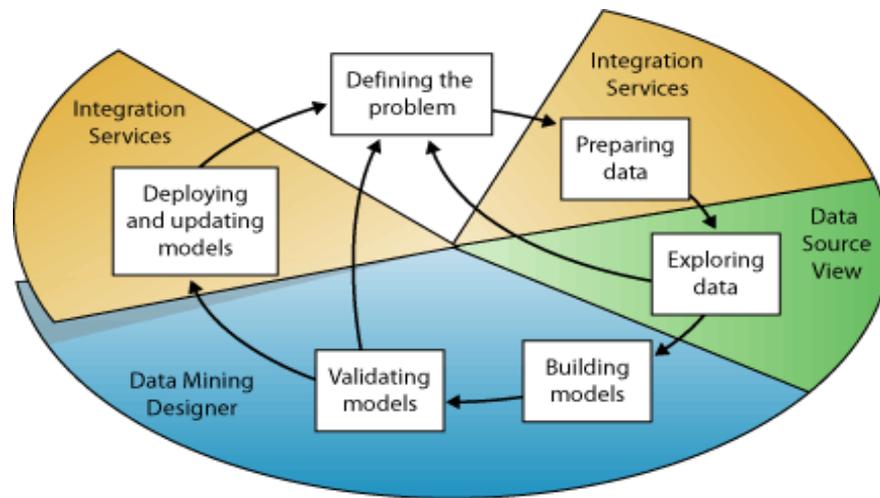
3. Transformación del conjunto de datos de entrada, se realizará de diversas formas en función del análisis previo, con el objetivo de prepararlo para aplicar la técnica de minería de datos que mejor se adapte a los datos y al problema, a este paso también se le conoce como pre procesamiento de los datos.

4. Seleccionar y aplicar la técnica de minería de datos, se construye el modelo predictivo, de clasificación o segmentación.

5. Extracción de conocimiento, mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un pre procesamiento diferente de los datos.

6. Interpretación y evaluación de datos, una vez obtenido el modelo, se debe proceder a su validación comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos. (pp. 14-15)

Sin embargo, a manera de ahondar un poco en otras aproximaciones, describiré también el criterio tomado por Microsoft (2019), el cual se resume en la siguiente figura:



Fuente: (Microsoft, 2019)

Figura 2. Proceso del Data Mining según Microsoft.

Según la aproximación de Microsoft, el Data Mining utiliza seis pasos que resultan ser cíclicos y son Defining the Problem, Preparing Data, Exploring Data, Building Models, Exploring and Validating Models, Deploying and Updating Models traducidos como:

- Definición del problema.
- Preparación de los datos.
- Exploración (análisis) de datos.
- Construcción de modelos.
- Exploración (análisis) y validación de modelos.
- Implementación y actualización de los modelos.

Técnicas de minería de datos

Las técnicas de la minería de datos, son pasos estructurados en el tiempo y provienen del desarrollo de un cruce entre los elementos de inteligencia artificial y de estadística, se trata de algoritmos sofisticados que, aplicados sobre una estructura de datos permiten para obtener unos resultados. Las técnicas más representativas son según la agrupación realizada por Camborda (2014):

Redes neuronales. - Son un paradigma de aprendizaje y procesamiento automático inspirado en la forma en que funciona el sistema nervioso de los animales. Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida. Algunos ejemplos de red neuronal son: el perceptrón y perceptrón multicapa.

Regresión lineal. - Es la más utilizada para formar relaciones entre datos. Rápida y eficaz pero insuficiente en espacios multidimensionales donde puedan relacionarse más de 2 variables.

Árboles de decisión. - Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema. Ejemplos: Algoritmo ID3, Algoritmo C4.5, Chaid, C&RT.

Modelos estadísticos. - Es una expresión simbólica en forma de igualdad o ecuación que se emplea en todos los diseños experimentales y en la regresión para indicar los diferentes factores que modifican la variable de respuesta.

Agrupamiento o Clustering. - Es un procedimiento de agrupación de una serie de vectores según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes. Ejemplos: Algoritmo K-means, o Algoritmo K-medoids.

Reglas de asociación. - Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.
(pp. 16-17)

Técnica de Árboles de decisión

Es una de las técnicas más populares conforme describen Kumar, Baharadwaj, & Pal: “Árboles de decisión es uno de los algoritmos más utilizados en los métodos de aprendizaje supervisado para la exploración de data basada en la técnica divide y vencerás” (pág. 13). Esta técnica fue desarrollada por Morgan y Sonquist en el año 1963, es una técnica para el aprendizaje de modelos comprensibles de decisión elaborados a partir de una muestra de datos disponible.

Es decir, esta técnica construye un “modelo” o “representación” de la regularidad existente en los datos. El término “comprensible” hace referencia al hecho de que estos modelos pueden ser expresados de una manera simbólica, en forma de un conjunto de condiciones o reglas escritas en forma verbal (a diferencia de otros métodos, como las redes neuronales) por tanto, pueden tener como resultado modelos inteligibles para los seres humanos y también para sistemas semiautomáticos que procesen reglas.

Un árbol de decisión tiene unas entradas las cuales pueden ser un objeto o una situación descrita por medio de un conjunto de atributos y a partir de esto devuelve una respuesta la cual en últimas es una decisión que es tomada a partir de las entradas. Los valores que pueden tomar las entradas y las salidas pueden ser valores discretos o continuos. Cuando se utilizan valores discretos en las funciones de una aplicación se denomina clasificación y cuando se utilizan los continuos se denomina regresión.

El árbol de decisión suele contener nodos internos, nodos de probabilidad, nodos hojas y arcos. Un nodo interno contiene un test sobre algún valor de una de las propiedades. Un nodo de probabilidad indica que debe ocurrir un evento aleatorio de acuerdo a la naturaleza del problema, este tipo de nodos es redondo, los demás son cuadrados. Un nodo hoja representa el valor que devolverá el árbol de decisión y finalmente las ramas brindan los posibles caminos que se tienen de acuerdo a la decisión tomada.

En el diseño de aplicaciones informáticas, un árbol de decisión indica las acciones a realizar en función del valor de una o varias variables. Es una representación en forma de árbol cuyas ramas se bifurcan en función de los valores tomados por las variables y que terminan en una acción concreta. Se suele utilizar cuando el número de condiciones no es muy grande (en tal caso, es mejor utilizar una tabla de decisión).



Fuente: (Camborda, 2014, p. 19)

Figura 3. Árbol de decisión y sus características.

Como característica principal de los árboles de decisión se tiene que cada nodo corresponde a un atributo y cada rama al valor posible de ese atributo y una hoja del árbol especifica el valor esperado de la decisión. La explicación de una determinada decisión viene dada por la trayectoria desde la raíz a la hoja de una decisión.

El software Weka

Para Córdoba, L. (2011) “es una herramienta de tipo software para el aprendizaje automático y minería de datos diseñado a base de Java y desarrollado en la universidad de Waikato en Nueva Zelanda en el año 1993, esta herramienta por su nombre en inglés (Waikato Environment for Knowledge Analysis) además es una herramienta de distribución de licencia GNU-GLP o software libre”. WEKA contiene una colección de algoritmos para realizar análisis de datos y modelado predictivo, también tiene herramientas para la visualización de estos datos, además provee una interfaz gráfica que unifica las herramientas para que estén a una mejor disposición.

Sus características más importantes son, según agrupa Canvia (2019):

Es una herramienta muy versátil que soporta muchas tareas estándar de la minería de datos en especial tareas de procesamiento de datos, regresión, clasificación, clustering entre

otras, así mismo permite la visualización y la selección de los datos. Todas las técnicas en WEKA están basadas en la función de datos que están disponibles en un fichero plano o una relación, en donde cada registro de datos esta descrito por un número fijo de atributos nominales o numéricos. Permite el acceso a otras instancias de bases de datos por medio de SQL, gracias al JDBC, además puede procesar un resultado generado a base de una consulta hecha a una base de datos. (pp. 41-42)

2.2.2. Deserción Estudiantil

Conforme recopilan Gonzales y Rodríguez (2017), el concepto de deserción estudiantil que prima sobre los demás es el definido como “la interrupción o desvinculación del alumno dentro del proceso educativo, considerando a los desertores, a aquellas personas, que tienen dos semestres de inactividad” (p. 23), siendo el contexto estándar utilizados en estudios de supervivencia y resultado probabilístico definido con el método de Kaplan-Meier (1958).

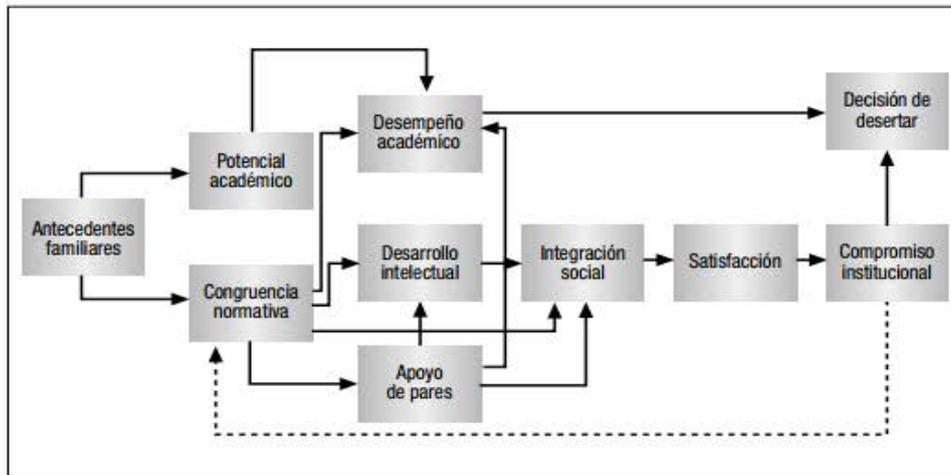
Los autores Vásquez y otros (2003), identifican tres tipologías posibles en la deserción estudiantil, las cuales son agrupadas por Gonzales y Rodríguez (2017):

1. Deserción precoz. El estudiante abandona los estudios antes de haberse matriculado
2. Deserción temprana. El estudiante abandona los estudios durante los primeros cuatro semestres.
3. Deserción tardía. El estudiante abandona los estudios del quinto semestre en adelante. (p. 23)

Otra de las clasificaciones para la deserción estudiantil es ofrecida por Spady en el año 1970, cuyo estudio determinó que una carencia de integración social era el factor preponderante para que el estudiante abandonara su carrera, dicha clasificación es recopilada también por Gonzales y Rodríguez (2017):

1. Deserción por Cohorte. Este tipo de deserción se basa en el conjunto de estudiantes que coinciden en el periodo de ingreso académico en el primer curso en su etapa universitaria.

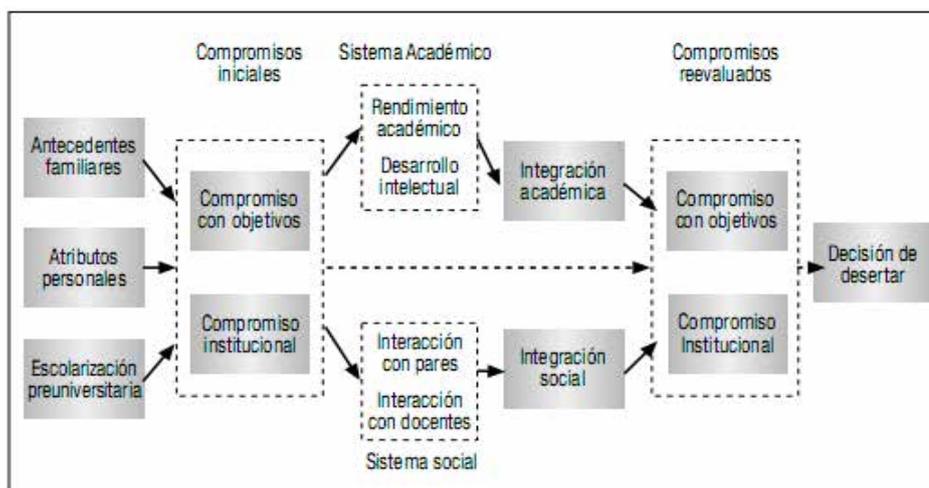
2. Deserción por Periodo. Este tipo de deserción se basa en el número de estudiantes no matriculado en dos semestres. Se le denomina deserción anual. (p. 24)



Fuente: Gonzales y Rodriguez (2017)

Figura 4. Modelo Spady de la deserción estudiantil.

Una teoría diferente es la ofrecida por Tinto, quien refiere que el factor determinante de la deserción se produce en cualquier parte de la línea de tiempo entre ingreso a los estudios hasta la toma de decisión de abortar dicho esfuerzo. Esta teoría concluye que “el estudiante se encuentra en un constante proceso de evaluación de costo beneficio” (Gonzales y Rodriguez, 2017, p. 25).



Fuente: Gonzales y Rodriguez (2017)

Figura 5. Modelo Tinto de la deserción estudiantil.

La deserción estudiantil en el Perú

La deserción estudiantil es un problema que afecta a la educación nacional en todos los niveles, tanto en instituciones educativas privadas como a las regidas por el Estado, pese a ello no se encuentra muy difundido su estudio el cual se ha observado se realiza a través de diferentes indicadores denominados como: “Tasas de abandono, Tasa de término, Tasa anual de deserción, deserción por cohorte de ingreso, etc. con porcentajes muy oscilantes que van desde 7% hasta 50%” (Manco, 2015, p. 80).

Aunque dicho problema ha sido abordado por instituciones internacionales tales UNESCO, OCDE, IELSAC y CINDA (Gonzales & Rodriguez, 2017, p. 23), se estima que en nuestro país “el 43.7% de estudiantes que ingresan a las universidades concluyen con sus estudios. No se dispone de datos de la Tasa de deserción de estudiantes a nivel de Instituciones de Educación Superior Universitaria” (Manco, 2015, p. 80). Pese a ello no se cuenta con una institución o entidad que realice el seguimiento de la deserción estudiantil a nivel universitario por lo cual no contamos con indicadores oficiales sobre dicha problemática para conocer su real magnitud.

Existen muy pocos estudios sobre deserción estudiantil a nivel universitario en Perú, un estudio realizado por la empresa Flanqueo en el año 2013 “halló que un 30% de los alumnos matriculados en universidades e institutos del Perú, se retiraban en el primer semestre” (Gonzales y Rodriguez, 2017, p. 26) en el cual se concluía que la población del sector socioeconómico C es la más propensa a desertar de sus estudios superiores debido principalmente a factores económicos o académicos.

El estudio más reciente es el elaborado por el Ministerio de Educación en el año 2020, el que hace referencia el diario El Comercio (Orbegozo, 2020), en el cual se señaló que la tasa de deserción universitaria alcanzó “18,6% de un total de 955.000 estudiantes en el Perú; un indicador que es seis puntos porcentuales mayor al registrado en el 2019 (12%)” (p. 1). En dicha publicación se señala que “la tasa de deserción actual es de 9,85%; mientras que en las privadas llegó a 22,5%” (p. 1), así también, explicó Jorge Mori, Director General de Educación Superior Universitaria del MINEDU que los factores de deserción resultaron ser multicausales y no solo económicos pues son atribuibles a cuestiones familiares, vocacionales y de exigencia académica en el contexto de la pandemia ocasionada por el COVID-19.

Por tales consideraciones, el presente estudio puede contribuir a sentar bases sobre los factores que lo producen y con posteriores análisis se puedan tomar medidas específicas para contribuir con la calidad educativa y fomentar un mejor aprovechamiento de la enseñanza universitaria.

Dimensiones de la variable deserción estudiantil

Para el presente trabajo de investigación se tomará en consideración el modelo desarrollado por Mori (2012) en el cual combina las teorías de Spady y Tinto adaptándolos a la realidad universitaria peruana, definiendo que las dimensiones para dicha variable resultan ser los factores principales de la deserción estudiantil universitaria tales como:



Fuente: Mori (2012)

Figura 6. Modelo de la deserción estudiantil en una universidad peruana.

Dicho modelo es descrito a detalle por Mori (2012) definiendo cada una de sus dimensiones como:

El Factor académico

Este factor se encuentra referido al aspecto vocacional y personal según Cu Balán (2005). Se encuentra circunscrito al rendimiento académico histórico del estudiante, el cual,

bajo modalidades distintas de aprendizaje, aplicadas por distintos docentes con diferentes estrategias, técnicas y métodos desde su etapa pre universitaria, se concreta en la autopercepción que tiene el alumno en su ingreso a la universidad de su propio desempeño académico, por lo que al percibir un desempeño deficiente, éste se convierte en motivo de preocupación y a la larga resulta un factor determinante para la decisión de desertar.

Factor individual

Se refiere a distintos aspectos muy variables al interior del estudiante, el cual se encuentra afectado por sus “características individuales, las metas, los intereses personales, las motivaciones y la orientación vocacional” (Mori, 2012, p. 13). Es decir, se configura en la consideración de los aspectos internos del estudiante lo cual permite definir su perfil, el cual será útil para efectuar intervenciones y evitar el riesgo de la decisión de desertar.

Factor ambiental

Este factor está centralizado en los “aspectos de relaciones externas y de financiamiento” (Mori, 2012, p. 17); este resulta ser uno de los más significativos y mayoritariamente elegido como principal para la deserción, pues incluye la dificultad o imposibilidad de afrontar los pagos por derechos académicos o sostener la actividad estudiantil. Autores como Sanabria (2002), Cabrera y otros (2004), Canales y Ríos (2007), Ramírez (2009) y Rojas (2009) señalan que el factor ambiental en lo económico es una de las principales razones para la deserción estudiantil.

Factor institucional

Este factor es señalado como uno de los principales para la deserción estudiantil universitaria pues el estudiante se ve afectado por las características de su institución educativa tales como el “plan de estudios, la plana docente, los bajos costos y el tipo de servicios” (Mori, 2012, p. 11). Dichas características únicas brindadas por las universidades en ocasiones no llenan las expectativas o características de la población estudiantil que elige las carreras, y es resultado principalmente de una política orientada a la captación de estudiantes (ingresos) que resultan en ingresos económicos a las universidades y no encontrarse correctamente orientadas a los perfiles y necesidades de los ingresantes a las carreras en cuanto a sus habilidades y capacidades.

2.2.3. Escuela de Posgrado de la Universidad Nacional José Faustino Sánchez Carrión

La Universidad Nacional José Faustino Sánchez Carrión fue creada oficialmente en diciembre de 1968 mediante el Decreto Ley N° 17358 dado por el Gobierno Revolucionario de las Fuerzas Armadas del Perú, mediante el cual, la filial huachana de la Universidad nacional del Centro del Perú se independiza y empieza a funcionar con el nombre actual, iniciando sus funciones bajo la autoridad de su primer rector, el Ing. Luis Felipe Ricci Bohórquez. Actualmente cuenta con 13 Facultades en la que se imparten 37 carreras profesionales además de la Escuela de Posgrado en la que se forman maestros y doctores. (UNJFSC, 2018)

Conforme el estatuto de la UNJFSC (2018), la escuela de posgrado es definida como:

La Unidad de Posgrado es la organización académica - administrativa encargada de integrar las actividades de posgrado de la Facultad y, gestiona los estudios o programas de formación continua que conducen a títulos de segunda especialización, diplomados y cursos que conlleven a certificaciones con nota aprobatoria. (p. 23)

Tiene su sede principal, en la que funcionan sus oficinas administrativas y aulas para el desarrollo de clases en el campus universitario de la UNJFSC en el pabellón “Manuel Angel Mendoza Cruz”.



Fuente: (UNJFSC, 2018)

Figura 7. Frontis de la Escuela de Posgrado de la UNJFSC.

Conforme el estatuto de la UNJFSC, la escuela de Posgrado se encuentra bajo la autoridad de un Director de Escuela, apoyado por la Secretaría Académica, Secretaría Administrativa y la Unidad de Idiomas las cuales tienen entre sus funciones conforme se expone en el Reglamento de Organización y Funciones (UNJFSC, 2016):

Secretaría Académica

Artículo 175°. - Naturaleza y funciones Es la encargada de apoyar al Director de la Escuela en los asuntos académicos de acuerdo al reglamento correspondiente.

Sus funciones son:

1. Supervisar la labor de los docentes de la Escuela de Posgrado.
2. Efectuar la propuesta de los docentes especialistas para el desarrollo de las asignaturas de los programas de maestría y doctorado, en coordinación con las Unidades de Posgrado de las Facultades.
3. Organizar y mantener actualizado los registros de los Grados Académicos de Maestro y Doctor.
4. Emitir constancias Académicas refrendadas por el Director de la Escuela, de acuerdo a la competencia asignada.
5. Formular el Proyecto de Desarrollo Académico de la Escuela de Posgrado.
6. Recepcionar las propuestas de los Directores de las Unidades de Posgrado de las Facultades respecto a la carga lectiva por semestres académicos, las cuales deberán ser de conocimiento oportuno del Director de la Escuela de Posgrado.
7. Coordinar con la Unidad de Grados y Títulos de la Universidad las acciones que corresponden acorde a sus funciones.
8. Coordinar con la Unidad de Biblioteca las acciones que corresponden acorde a sus funciones
9. Coordinar con la Unidad de Registros y Asuntos Académicos las acciones que corresponden a sus funciones.
10. Elaborar y presentar al Director de la Escuela de Posgrado la memoria anual.
11. Asistir a las sesiones de Directorio.

Secretaría Administrativa

Artículo 176°. - Naturaleza y funciones

Es la encargada de apoyar al Director de la Escuela de Posgrado en los asuntos administrativos de acuerdo al Reglamento correspondiente.

Sus funciones son:

1. Asistir a las sesiones ordinarias y extraordinarias del Directorio de la Escuela de Posgrado.
2. Elaborar la memoria anual y presentarla al Director de la Escuela.
3. Llevar los registros actualizados de los estudiantes de las diferentes Maestrías y Doctorados.
4. Fedatear la documentación correspondiente de la Escuela de Posgrado.
5. Elaborar el horario y control de asistencia de los alumnos de la Escuela de Posgrado.
6. Elaborar el presupuesto e informe de ingresos y egresos de la Escuela de Posgrado.
7. Coordinar con el centro de cómputo las acciones que corresponden a la realización de sus actividades regulares.
8. Llevar y mantener al día el Libro de Actas de sesiones del Directorio de la Escuela de Posgrado.
9. Citar a las sesiones ordinarias y extraordinarias del Directorio, transcribir los acuerdos y efectuar la difusión de los mismos.
10. Coordinar con el Jefe del Centro de Idiomas de la Escuela de Posgrado las acciones que se desprenden de la realización de sus labores regulares.
11. Coordinar con la Unidad de Biblioteca las acciones que corresponden acorde a sus funciones.
12. Coordinar con la Unidad de Registros y Asuntos Académicos las acciones que corresponden acorde a sus funciones.
13. Emitir Constancias refrendadas por el Director de la Escuela de Posgrado.

14. Llevar un registro actualizado de los Docentes acreditados por especialidad.
15. Otras que le asigne el Directorio de la Escuela de Posgrado mediante Resolución. (UNJFSC, 2016, pp. 74-75)

Datos estadísticos de población estudiantil

Conforme los datos de población obtenidos desde la página web de la Superintendencia Nacional de Educación Superior Universitaria (SUNEDU), la cual mantiene datos oficiales consolidados en su sistema y se muestran en su totalidad en la sección anexos del presente trabajo obtenidas desde SUNEDU (2019), además de la información obtenida a través de la oficina de Registros académicos de la EAP de Posgrado en relación a las maestrías y años elegidas para el presente estudio.

La base de datos de la SUNEDU muestra la cantidad total de la población de postulantes, ingresados, estudiantes matriculados y egresados por año y por programa (maestría o doctorado) entre los años 2014 y 2017.

2.3 Bases filosóficas

La educación se erige como un aspecto fundamental para conseguir la felicidad del ser humano, la cual toman Marx y Engels como el rol principal de cualquier modo de vida que se elija, ello porque a través de la educación se obtienen capacidades y habilidades que dan ventaja al individuo dentro de una sociedad para conseguir influencia socioeconómica y cultural de la sociedad humana, las cuales a su vez permiten mejores oportunidades de ingresos económicos o estatus social.

Es por ello que Rivero (2017), señala que la filosofía de la educación está orientada a establecer los fundamentos filosóficos que debe formular la denominada “teoría pedagógica” sobre la cual se edificará el proceso de enseñanza orientado al aprendizaje. Dicha filosofía de la educación establecida, claramente debe tomar en consideración los fundamentos sociológicos y psicológicos sobre los cuales se planifican y desarrollan las actividades didácticas.

Ello es importante al considerar que la concepción del propósito de vida y el ideal del ser humano que se quiere al alcanzar individualmente, en su gran mayoría, a la felicidad y sentimiento de realización del individuo. lo cual debe reflejarse en el currículo o programa nacional de todos los países, en el cual debe planificarse lo requerido por los estados en relación a la educación de su población en beneficio de su sociedad.

2.4 Definición de términos básicos

Sistema: Es definido como un objeto o entidad compleja compuesto de partes o componentes que se interrelacionan unos con otros para lograr un objetivo conjunto.

Modelamiento: es una técnica de Dinámica de sistemas, la cual permite la representación de la organización y funcionamiento de un sistema.

Modelo: Es la representación de un sistema funcional, el cual permite simular el resultado de acuerdo al valor de las variables definidas para el mismo.

Minería de datos: es el proceso de búsqueda en grandes bases de datos para encontrar información útil que sirva para la toma de decisiones.

Weka: es un programa informático (software) diseñado a base de Java y desarrollado en la universidad de Waikato en Nueva Zelanda en el año 1993, esta herramienta por su nombre en inglés (Waikato Environment for Knowledge Analysis) además es una herramienta de distribución de licencia GNU-GLP o software libre.

Árboles de decisión: Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.

Exactitud: Se calcula dividiendo el número total de registros correctamente clasificados por el número total de registros incorrectos o de referencia y expresándolo como porcentaje.

Medida de concordancia: Error de medida en la variabilidad, consecuentemente un objetivo de los estudios de fiabilidad consiste en estimar el grado de dicha variabilidad y se utiliza el índice kappa que está relacionada con las distribuciones marginales siguientes:

Valoración del Índice Kappa	
Valor de k	Fuerza de la concordancia
< 0.20	Pobre
0.21 – 0.40	Débil
0.41 – 0.60	Moderada
0.61 – 0.80	Buena
0.81 – 1.00	Muy buena

Fuente: (Camborda, 2014)

Figura 8. Valoración del índice Kappa.

Predicción: A partir de un conjunto de datos históricos con resultado conocido, se pretende modelizar estos datos para poder saber resultados futuros. Es necesario que el entorno de los datos históricos no sea significativamente diferente del entorno del futuro periodo en que se pretende predecir. Un modelo predictivo tiene diversas variables de entrada que han sido seleccionadas por su alta correlación con el resultado histórico, y la salida es el resultado en sí. Algunos ejemplos de técnicas que sirven para crear modelos predictivos son: inducción de reglas, redes neuronales y regresión.

Deserción estudiantil: es una medida a manera de tasa o porcentaje, el cual se refiere a la cantidad de estudiantes que desertan o dejan de lado los estudios emprendidos por distintos factores, los cuales regularmente son de carácter económico o psicológico.

2.5 Hipótesis de investigación

2.5.1 Hipótesis general

Es posible aplicar la metodología de minería de datos denominada “árboles de decisión” para crear un modelo de predicción del comportamiento de la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la UNJFSC.

2.5.2 Hipótesis específicas

Sí es posible determinar los factores que permiten predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.

La exactitud de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC es buena.

La medida de concordancia de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC es buena.

2.6 Operacionalización de las variables

Variable 1: Técnica de árboles de decisión.

Variable 2: Deserción estudiantil.

VARIABLES	DEFINICIÓN CONCEPTUAL	DEFINICIÓN OPERACIONAL	DIMENSIONES	INDICADORES	ITEMS
Variable X: Deserción estudiantil.	La interrupción o desvinculación del alumno dentro del proceso educativo, considerando a los desertores, a aquellas personas, que tienen dos semestres de inactividad. Gonzales y Rodríguez (2017)	Desvinculación de los estudiantes de las maestrías del sector educación de la EAP de Posgrado de la UNJFSC.	Factores Académicos	Rendimiento académico en la educación secundaria Rendimiento académico en el nivel de pregrado	1-3 4,5
			Factores Individuales	Características Personales Características Familiares Motivación	6-8 10-11 9,12
			Factores Ambientales	Financiamiento Relaciones Sociales externas	13,14 15,16
			Factores Institucionales	Infraestructura y equipamiento Plan de estudios Plana docente	17,18 19 20
Variable Y: Técnicas de árbol de decisión	Es una de las técnicas más populares de minería de datos, basada en un algoritmo de los métodos de aprendizaje supervisado para la exploración de data basada en la técnica divide y vencerás. (Kumar, Baharadwaj y Pal, 2012, pág. 13)	Aplicación de la técnica de minería de datos para predecir la deserción estudiantil de los estudiantes de la EAP de Posgrado	Aplicación de la técnica a través de un modelo de simulación	Modelo de simulación mediante la técnica de árboles de decisión	

CAPÍTULO III

METODOLOGÍA

3.1 Diseño metodológico

Por su finalidad y teniendo en cuenta el grado de abstracción, esta investigación es de tipo aplicada, busca aplicar una teoría ya conocida como la minería de datos y una técnica ya desarrollada como la de árboles de decisión para resolver un problema específico de la realidad el cual es conocer si es posible predecir la deserción estudiantil de los estudiantes de los programas de educación de la Escuela de Posgrado de la UNJFSC.

La presente es una investigación correlacional aplicada, pues busca el analizar si es factible relacionar la aplicación de la técnica de árbol de decisión de minería de datos con la deserción estudiantil, ello a través de analizar la información que se obtenga de la población estudiantil de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC y elaborar luego un modelo predictivo que permita predecir su la deserción estudiantil.

La investigación planteada es de diseño no experimental, pues se realiza sin manipular deliberadamente las variables de estudio, pues se realiza el estudio con la población estudiantil de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.

La presente tesis utiliza el corte transversal, pues el instrumento de medición (cuestionario) es aplicado en una sola oportunidad. Además, es de enfoque mixto, ya que utiliza elementos de enfoque cualitativo y cuantitativo.

3.2 Población y muestra

3.2.1 Población

La población es la totalidad de alumnos (237) de los programas de maestría del sector educación matriculados en el ciclo académico 2019-II, las cuales son: “DOCENCIA SUPERIOR E INVESTIGACIÓN UNIVERSITARIA, GERENCIA DE LA EDUCACIÓN y CIENCIAS DE LA GESTIÓN EDUCATIVA CON MENCIÓN EN PEDAGOGÍA”.

3.2.2 Muestra

Para la presente investigación y por conveniencia, se utilizó el muestreo poblacional o censal, es decir, se recopilarán los datos de la totalidad de casos de la población, ello es posible debido a que se cuenta con la totalidad de datos de la base de datos de Registros académicos de Post Grado en relación a dichos programas de maestría y acceso para aplicar la encuesta correspondiente.

3.3 Técnicas de recolección de datos

Para el presente trabajo de investigación se utilizó la técnica de revisión bibliográfica y de análisis documental pues con el fin de obtener datos y técnicas fundamentales para analizar el problema de investigación para este trabajo en estudio se revisaron las fuentes escritas (textos en internet, tesis, revista, etc.) así como los récords académicos del alumnado en los que fue posible averiguar su situación final en relación a si desertaron de los estudios.

También se utilizó la técnica de la encuesta, pues el instrumento de medición para obtener los factores para el modelo de predicción a desarrollar se realizó a través de un cuestionario.

Como instrumento de recolección de datos se realizó la captura de información a través de la aplicación de un cuestionario para la recolección de datos académicos, actitudinales e institucionales, dicho cuestionario fue el utilizado en la investigación del autor Camborda (2014), adaptado a las características de la población de estudio y validado a través del sistema de juicio de expertos.

El cuestionario consta de 20 ítems con el cual se determinaron los valores de las dimensiones de la deserción estudiantil conforme el marco teórico, y ha sido aplicado en el estudio titulado: Predicción del rendimiento académico utilizando la técnica de árboles de decisión en los programas de maestría de educación en la escuela de posgrado de la Universidad Nacional José Faustino Sánchez Carrión (Díaz, 2021) del cual se recopilieron los datos necesarios.

3.4 Técnicas para el procesamiento de la información

Luego de la obtención de datos bibliográficos sobre la organización y funcionamiento de la Escuela de Posgrado de la UNJFSC, se aplicó el cuestionario a los estudiantes de la Escuela de Posgrado matriculados en el ciclo académico 2019 - II. Posteriormente se realizó la compilación de los datos, los cuales fueron estructurados e ingresados al software WEKA, del cual se obtuvieron los datos de análisis de la base de datos de entrenamiento y de predicción que se muestran en el capítulo de Resultados para el análisis.

Se creó una base de datos mediante el programa Ms. Excel en el cual se vincularon los aspectos influyentes en la deserción estudiantil, el cual necesitó ser categorizado conforme el promedio ponderado que obtuvieron según los registros de notas de la totalidad de cursos del ciclo brindados por la Oficina de Registros Académicos de la Escuela de Posgrado, de los cuales se obtuvo el registro de estudiantes que desertaron en el año 2019.

La base de datos obtenida de las respuestas de los cuestionarios fue ingresada en el software Weka, ello a través de una tabulación de datos en Ms. Excel para luego ser convertida a texto simple con el formato .ARFF, codificación ANSI la cual es la utilizada por el software Weka, con la cual se pudo analizar los 20 ítems del cuestionario. La base de datos ingresada fue utilizada como base de entrenamiento para el software.

Se realizó el proceso del archivo, obteniéndose luego una predicción del atributo deserción en base al entrenamiento que se realizó con la base de datos que contiene las deserciones reales, con ello se pudo comprobar el nivel de obtenido alcanzado por la predicción software. El nivel de acierto de la predicción obtenida mediante el software Weka fue analizada a través del índice Kappa de Cohen el cual conforme el valor que tome dicho estadístico corresponderá a la fuerza que existe en el acierto de la predicción otorgada por el software según se muestra a continuación.

Tabla 1 Valor de interpretación de la fuerza de concordancia según el valor del coeficiente Kappa de Cohen

Coeficiente Kappa	Fuerza de concordancia
0	Pobre
0.01 – 0.20	Leve
0.21 – 0.40	Aceptable
0.41 – 0.60	Moderada
0.61 – 0.80	Considerable
0.81 – 1.00	Casi perfecta

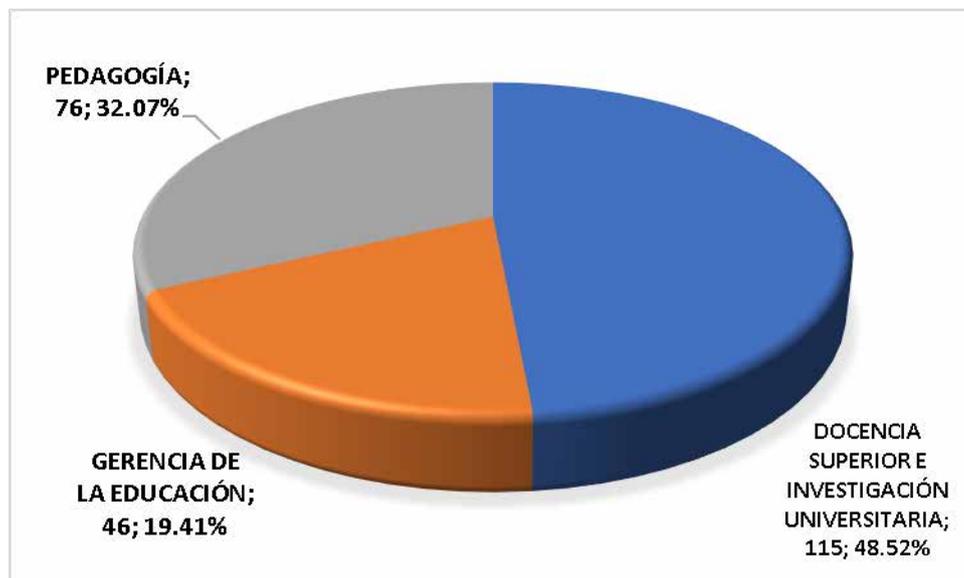
Fuente: (Cerde & Villarroel, 2008)

CAPÍTULO IV

RESULTADOS

4.1 Análisis de resultados

A continuación, se muestra el análisis estadístico de la base de datos confeccionada con las respuestas obtenidas de la aplicación del cuestionario a los 237 estudiantes que conformaron la muestra y cuyo resultado fue el siguiente:

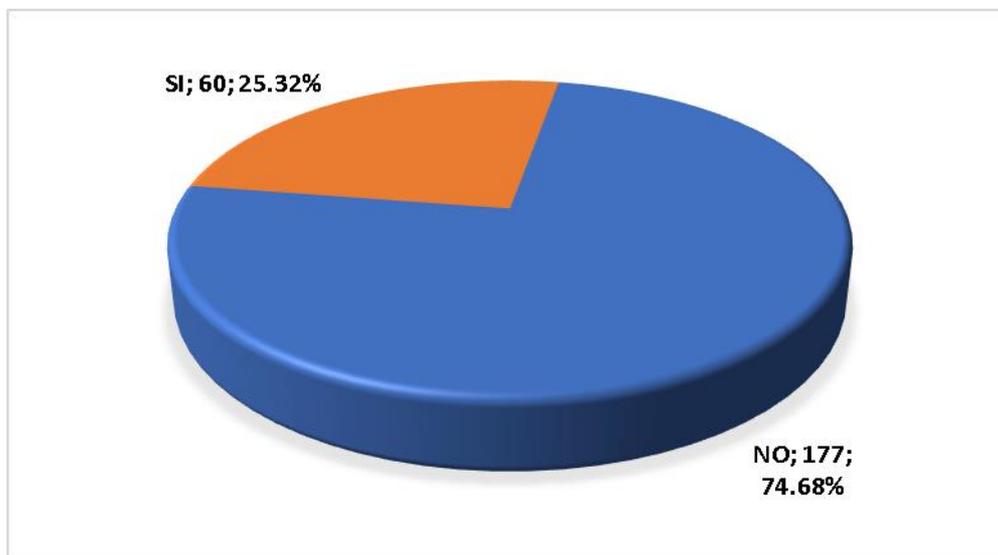


Fuente: Elaboración propia

Figura 9. Encuestados según programa de maestría.

Conforme se observa en la figura anterior, la mayor parte de los encuestados se encuentra cursando la maestría de Docencia Superior e Investigación Universitaria con

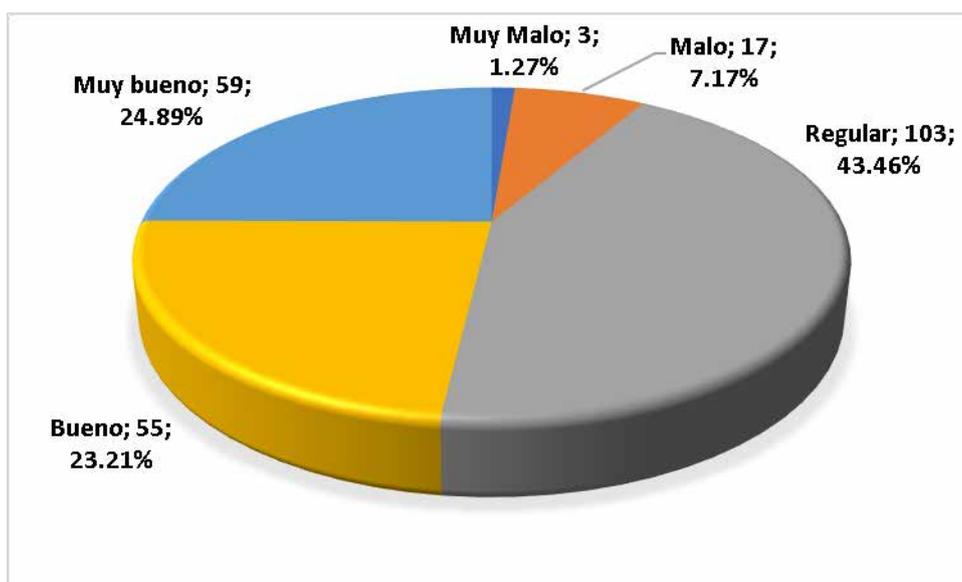
un total de 115 estudiantes que representa un 48,52% de la muestra, seguido por la maestría de Pedagogía la cual mantiene 76 estudiantes que representa un 32,07% de la muestra y por último, la maestría en Gerencia de la Educación con 46 estudiantes que representan un 19,41% .



Fuente: Elaboración propia

Figura 10. Deserción de estudiantes encuestados

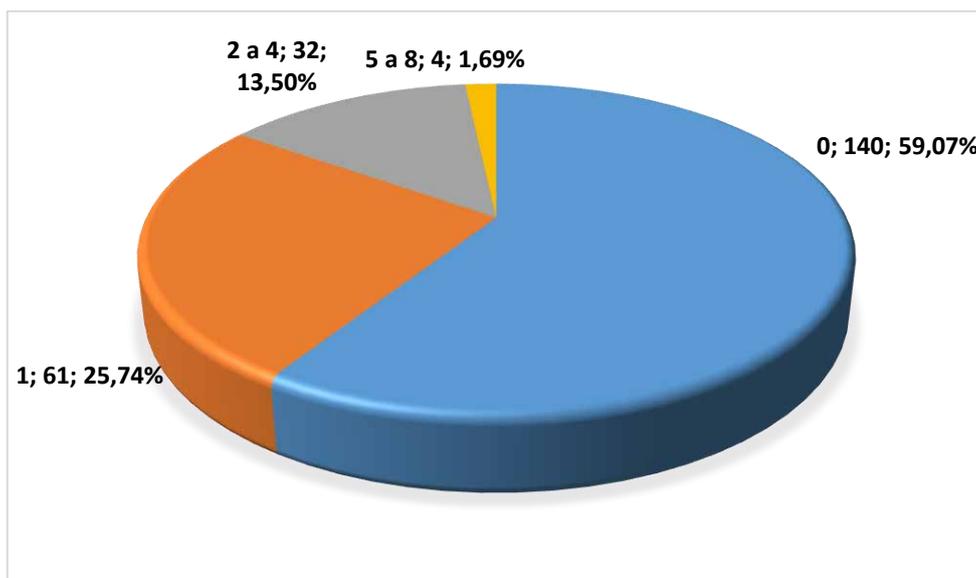
En la figura anterior se muestra la frecuencia y porcentaje de los encuestados matriculados en los ciclos 2019-II que a la fecha de estudio han desertado de sus estudios, observándose que fueron un total de 60 maestristas, que representan el 25,32% de la muestra los que optaron por desertar de sus estudios de posgrado.



Fuente: Elaboración propia

Figura 11. Pregunta 1 – Su rendimiento académico durante la secundaria fue

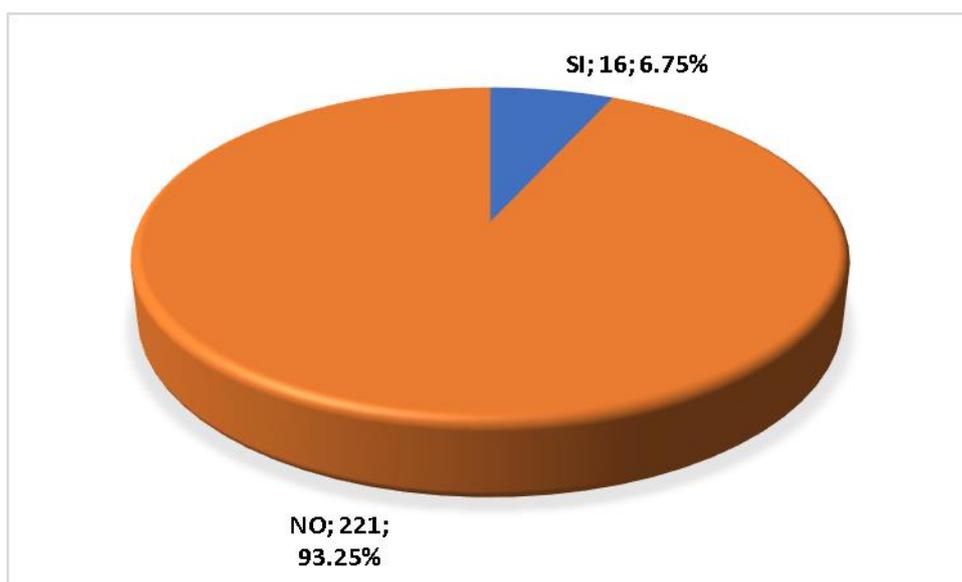
Conforme se observa en la figura anterior, la mayoría de la muestra encuestada en un 43,46% evalúan su rendimiento académico en el nivel de educación secundaria como regular, seguido por un 24.89% que lo evalúan como muy bueno, un 23,21% que lo evalúan como bueno, finalmente como malo y muy malo en un porcentaje de 7,17 y 1,27% de forma correspondiente. El rendimiento académico en esta etapa de formación escolar parece ser un factor que determina el rendimiento académico en los estudios de posgrado pues corresponde a la capacidad de aprender habilidades que servirán a futuro, sin embargo, ello será objeto de análisis en la situación concreta.



Fuente: Elaboración propia

Figura 12. Pregunta 2 – ¿Cuántas asignaturas reprobó en su educación secundaria?

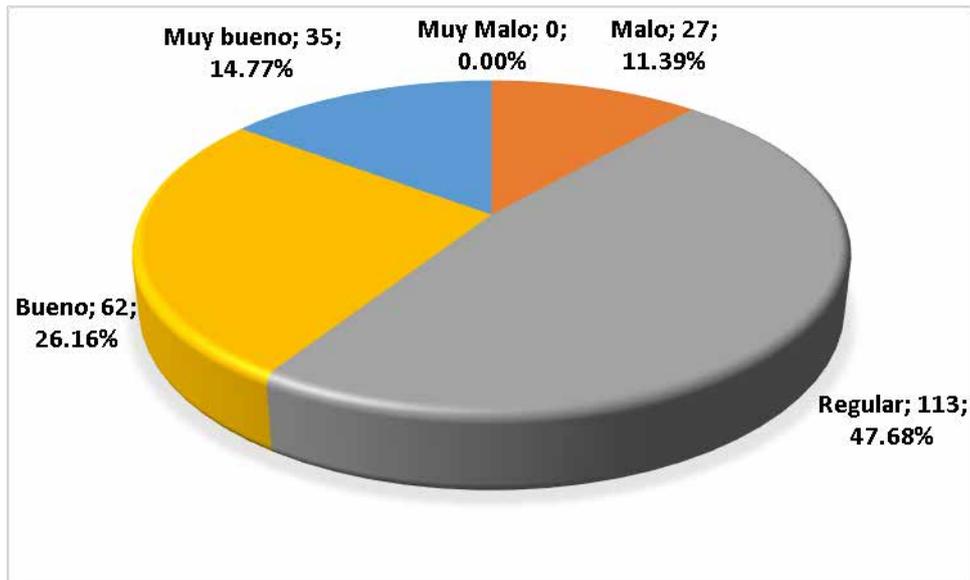
Conforme se observa en la figura anterior, la mayoría de la muestra encuestada en un 59,07% refieren no haber desaprobado ningún curso en la secundaria, así también un 25,74% manifiestan haber desaprobado 1 curso, un 13,50% de 2 a 4 y 1,69% refiere haber desaprobado 5 a 8 cursos. Dicha situación del rendimiento en secundaria parece ser complementaria a la percepción de su rendimiento en dicha etapa, lo cual debe ser analizado para conocer si es un factor de la deserción en los estudios de posgrado.



Fuente: Elaboración propia

Figura 13. Pregunta 3 – ¿Repetió algún año en la educación secundaria?

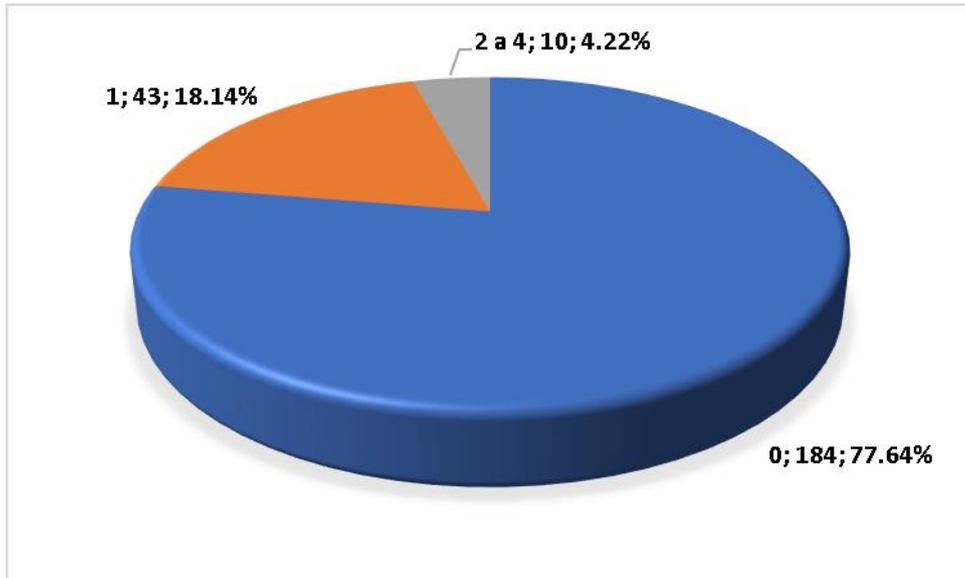
Conforme se observa en la figura anterior, la mayoría de los estudiantes encuestados, en un porcentaje de 93,25% manifestaron no haber repetido ningún año escolar en su educación secundaria, mientras que el 6,75% de la muestra encuestada refirió que sí repitió al menos un año. Dicha circunstancia debe ser analizada a fin de determinar si es un factor que determine la deserción en los estudios de posgrado.



Fuente: Elaboración propia

Figura 14. Pregunta 4 – Su rendimiento académico en la carrera de pregrado fue

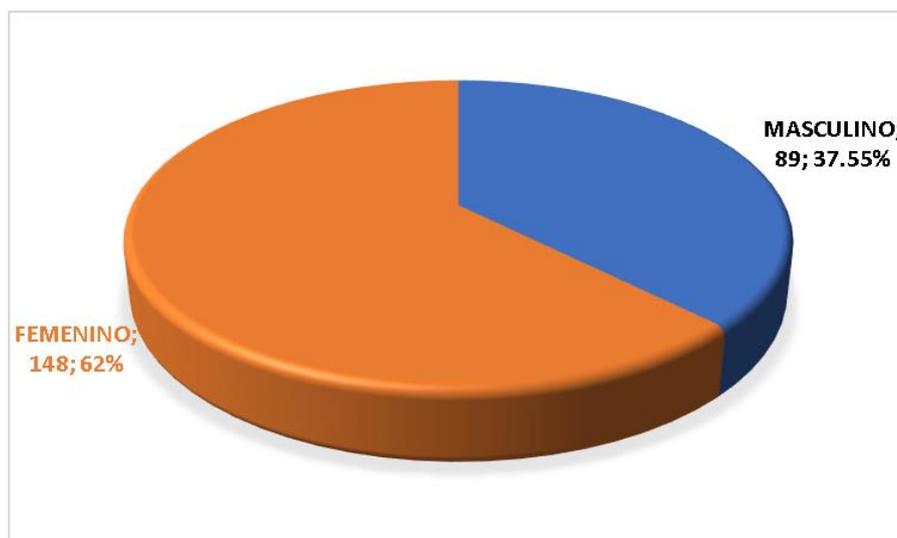
Conforme se observa en la figura anterior, la mayoría de la muestra encuestada en un 47,68% consideran que su rendimiento académico en los estudios de pregrado fue regular, un 26,16% refieren que dicho rendimiento fue bueno, un 14,77% que fue muy bueno y solo un 11,39% aceptan que su rendimiento en los estudios de pregrado fue malo. Al igual que los estudios secundarios, el considerar cómo fue dicho rendimiento en los estudios de pregrado puede ser un indicio de cómo será su rendimiento en los estudios de posgrado, por lo que debe tomarse en cuenta para el análisis de la presente tesis y determinarse si debe formar parte del modelo de simulación.



Fuente: Elaboración propia

Figura 15. Pregunta 5 – ¿Cuántos cursos desaprobó en su carrera de pregrado?

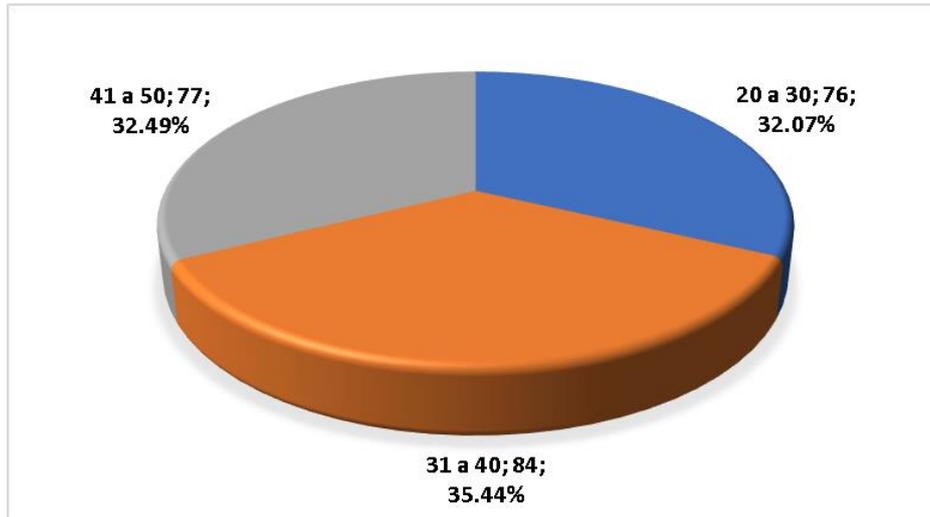
Conforme se observa en la figura anterior, la mayoría de estudiantes en la muestra encuestada en un 77,64% refieren no haber desaprobado ninguna asignatura en los estudios de pregrado, un 18,14% manifiestan haber desaprobado solamente una asignatura y un 4,22% refieren haber desaprobado de 2 a 4 cursos. Tal situación del rendimiento en los estudios de pregrado puede ser determinante para la predicción de su deserción los estudios de posgrado, hecho que merece ser analizado.



Fuente: Elaboración propia

Figura 16. Pregunta 6 - Identifique su sexo

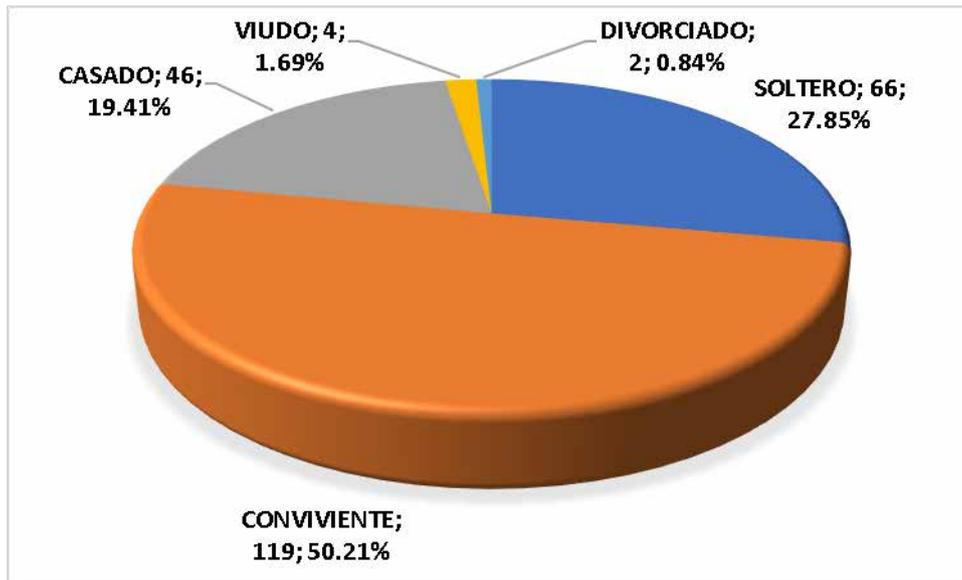
Conforme se observa en la figura anterior, en la pregunta 6 del cuestionario se captó la información sobre el sexo del encuestado, obteniéndose como resultado que la mayoría de estudiantes de la muestra encuestada en un 62% correspondieron a mujeres y un 37,55 % a varones. Dicha distribución será analizada por el software Weka para determinar si tal atributo contribuye a determinar la deserción estudiantil.



Fuente: Elaboración propia

Figura 17. Pregunta 7 – Rango de edad

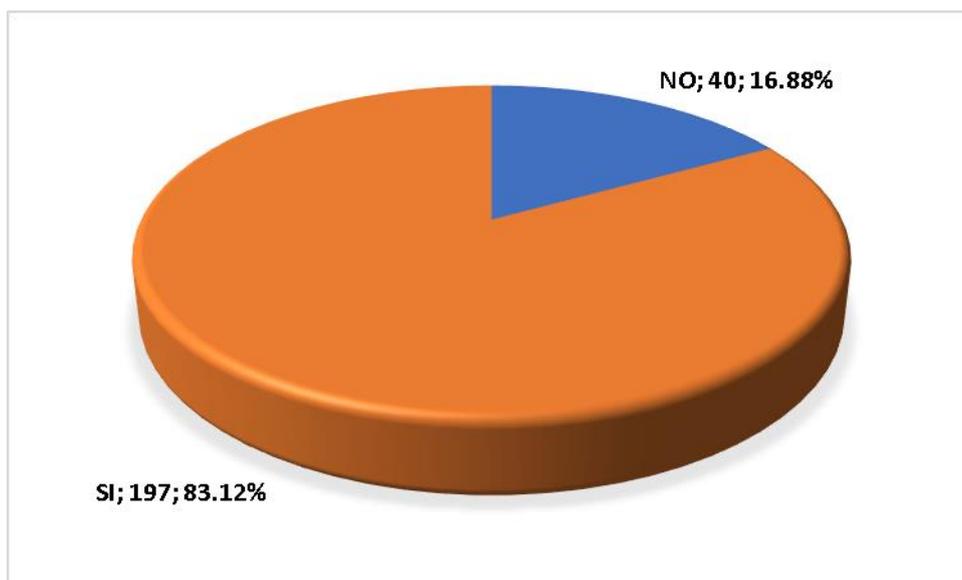
Conforme se observa en la figura anterior, la mayoría de los estudiantes de maestría que conforman la muestra se encuentran en el rango de edad de 31 a 40 años, seguidos del rango entre 41 a 50 años con un 32,49% y finalmente el rango entre 20 a 30 años con un 32,07%; sin embargo, conforme es notable la diferencia entre las frecuencias de dichos rangos etáreos es muy poca.



Fuente: Elaboración propia

Figura 18. Pregunta 8 – Estado civil

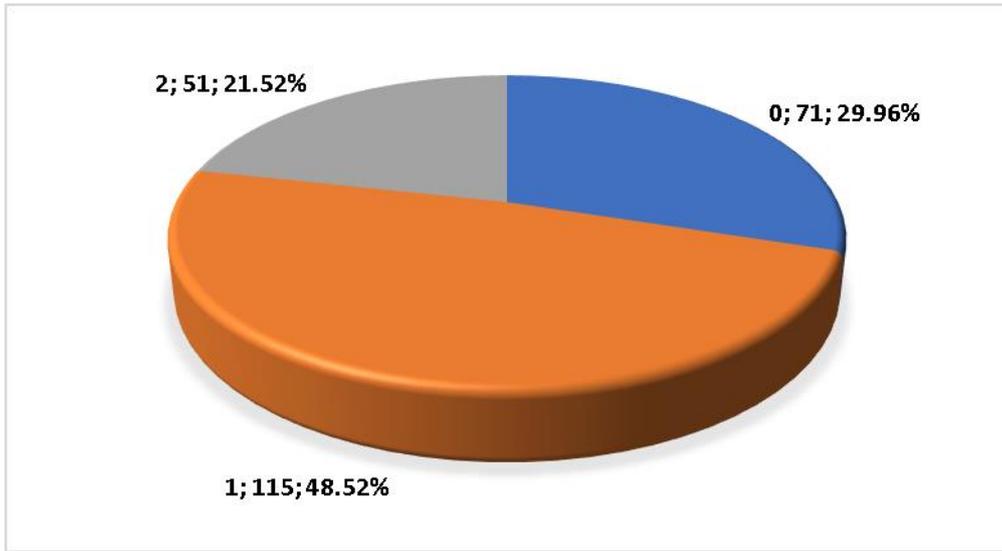
Conforme se observa en la figura anterior, la mayoría de la muestra encuestada en un 50,21% manifestó ser conviviente con un cónyuge, mientras que 27,85% manifestó encontrarse soltero, luego un 19,41% manifestó encontrarse casado. A través del análisis estadístico realizado por el software Weka se determinará si dicho atributo es concluyente para determinar la deserción estudiantil y debe tomarse en cuenta para la predicción del mismo.



Fuente: Elaboración propia

Figura 19. Pregunta 9 – Percibe que su trabajo se encuentra relacionado con su carrera de pregrado

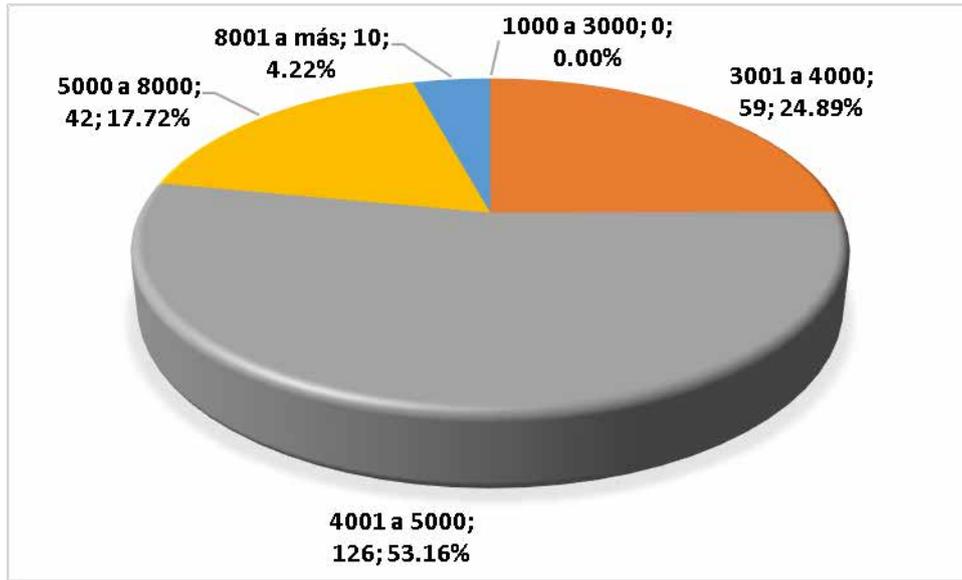
Conforme se observa en la figura anterior, la mayoría de la muestra encuestada en un 83,12% manifiesta que su trabajo actual se relaciona con la carrera de pregrado que estudió, lo cual puede ser indicio que la necesidad de cursar estudios de posgrado en dicha materia. Solo un 16,88% manifestaron que su trabajo no se relaciona con sus estudios de pregrado, lo cual puede ser un factor para determinar su deserción estudiantil en atención a la motivación que ello puede o no ofrecer.



Fuente: Elaboración propia

Figura 20. Pregunta 10 – Cantidad de hijos

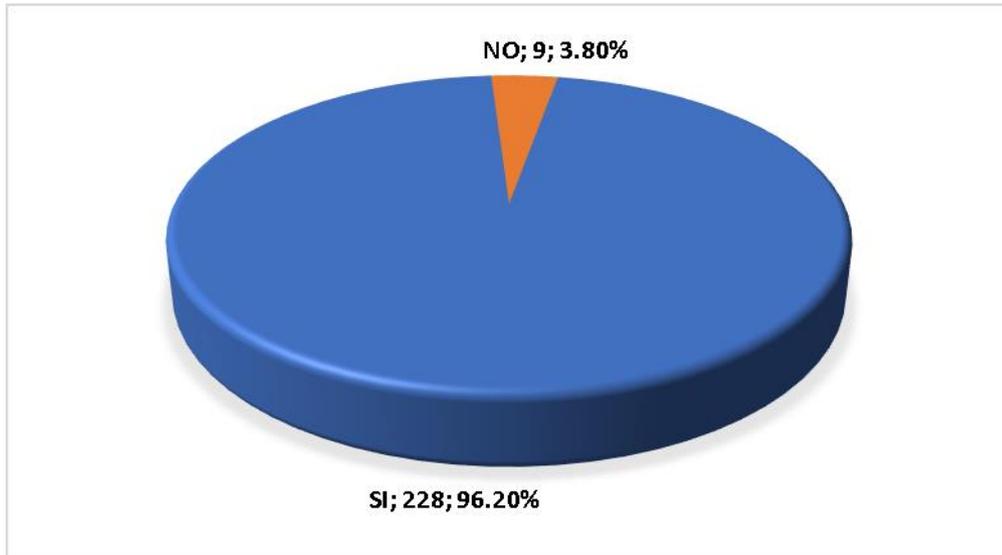
Conforme se observa en la figura anterior, la mayoría de la muestra encuestada en un 48.52% refiere que tiene un hijo/a, seguido de un porcentaje de 29,96% de estudiantes que refiere que no tiene hijos y por último un 21,52% de estudiantes que refieren tener dos hijos. La carga familiar que mantiene una persona puede ser un factor que contribuye a determinar la deserción estudiantil, sin embargo, ello debe ser objeto de análisis.



Fuente: Elaboración propia

Figura 21. Pregunta 11 – Ingreso familiar

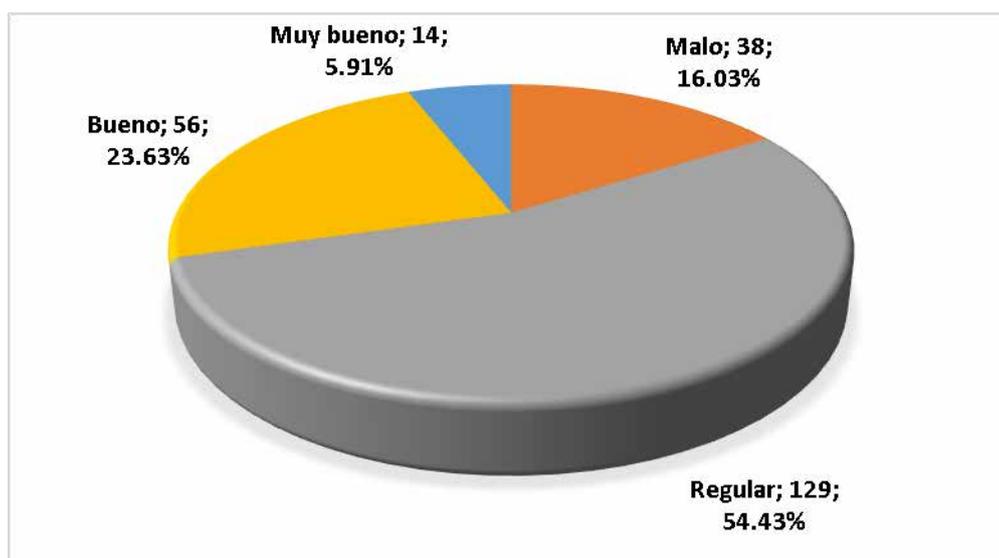
Conforme se observa en la figura anterior, la mayoría de estudiantes de la muestra encuestada en un 53,16% mantiene un ingreso familiar entre 4001 a 5000 soles, seguido por un 24,89 % que mantiene un ingreso familiar entre 3001 a 4000 soles. Ello parece ser relevante pues el contar con una base de ingreso económico sólida para afrontar los gastos que corresponden a derechos académicos y sustento material parece ser a simple vista un hecho determinante para la continuidad de los estudios.



Fuente: Elaboración propia

Figura 22. Pregunta 12 – ¿Tiene la motivación suficiente para culminar sus estudios de posgrado?

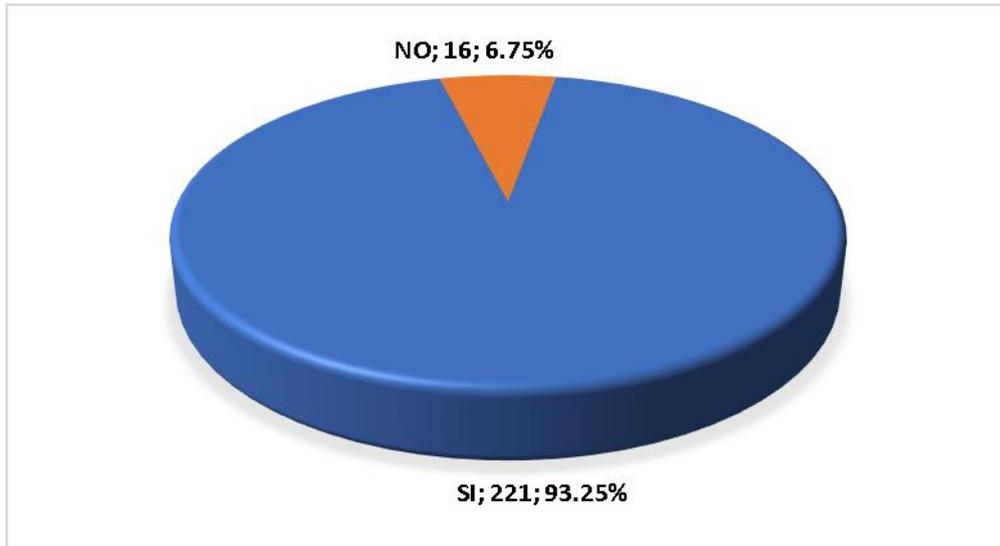
Conforme se observa en la figura anterior, y conforme a la pregunta formulada sobre la tenencia de la motivación suficiente para culminar sus estudios de, en dicha pregunta se buscó averiguar si la motivación resulta un factor en la deserción de estudios. Los resultados mostraron una que la gran mayoría refirieron mantener una motivación adecuada en el 96,20% de los casos con un total de 228 encuestados, y solo el 3,80% de encuestados refirieron que no. En una análisis a priori, el hecho que 60 personas haya desertado efectivamente parece señalar que dicha declaración no es un factor determinante para la deserción de los estudios de posgrado.



Fuente: Elaboración propia

Figura 23. Pregunta 13 – Considera que su situación económica actual es

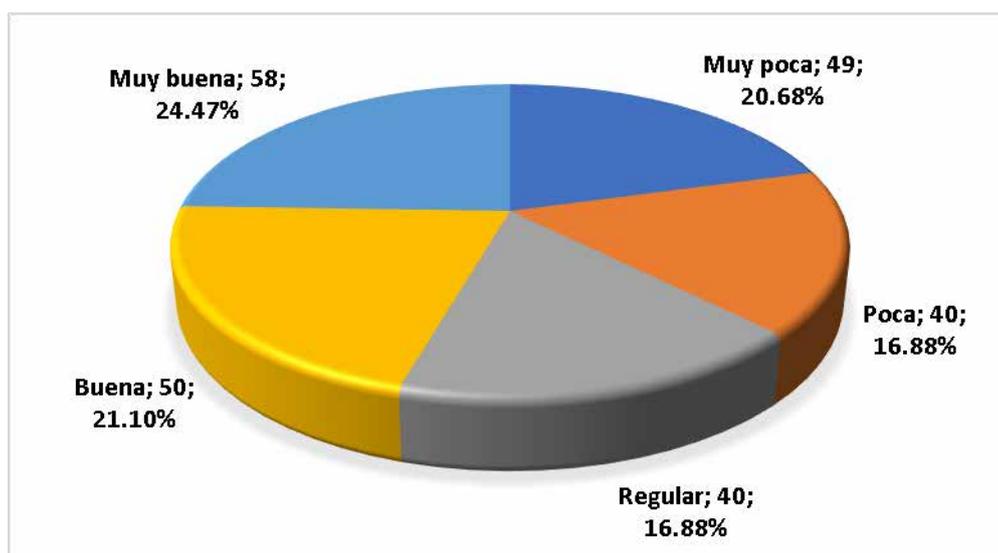
Conforme se observa en la figura anterior, y conforme al cuestionario se buscó determinar la apreciación del encuestado sobre su propia situación económica, a fin de determinar en el posterior análisis si dicha opinión era un factor determinante en la deserción estudiantil en los estudios de posgrado que cursa cada estudiante. Los resultados mostraron que la gran mayoría del 54,43% refirieron mantener una situación económica regular, seguidos del 23,63% quienes opinaron mantener una situación económica buena, para finalmente un 16,03% opinar sobre que mantienen una situación económica mala.



Fuente: Elaboración propia

Figura 24. Pregunta 14 – Considera que puede financiar sus estudios hasta finalizar la maestría

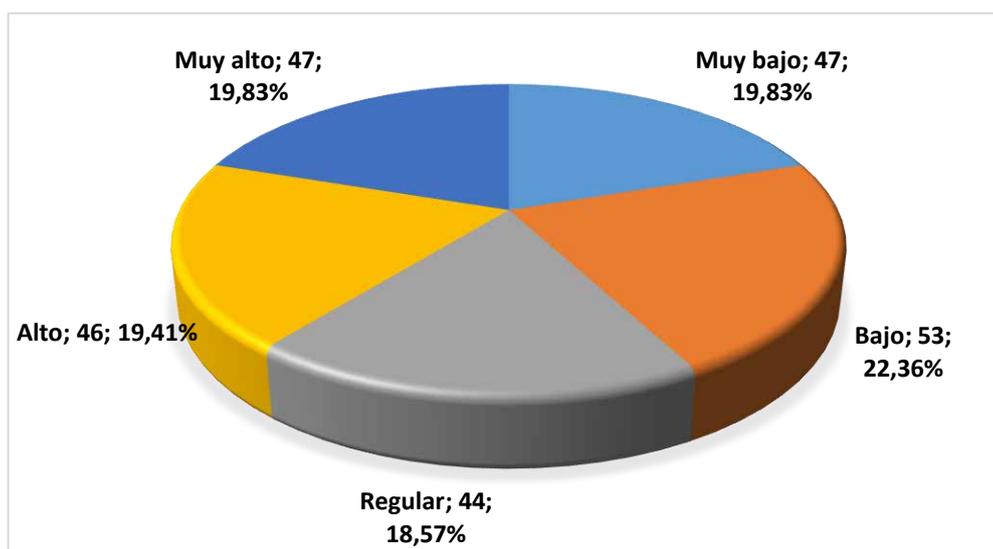
Conforme se observa en la figura anterior, y conforme el cuestionario aplicado buscó determinar la apreciación del encuestado sobre su capacidad financiera para solventar el pago de los derechos económicos que generan sus estudios de posgrado hasta finalizar los mismos. Al respecto se recibió de la gran mayoría con un 93,25% la opinión positiva y solamente un 6,75% refirieron que consideran no tener la posibilidad de financiamiento de sus estudios de posgrado.



Fuente: Elaboración propia

Figura 25. Pregunta 15 – Considera que la disponibilidad de tiempo para realizar sus actividades de estudiante es

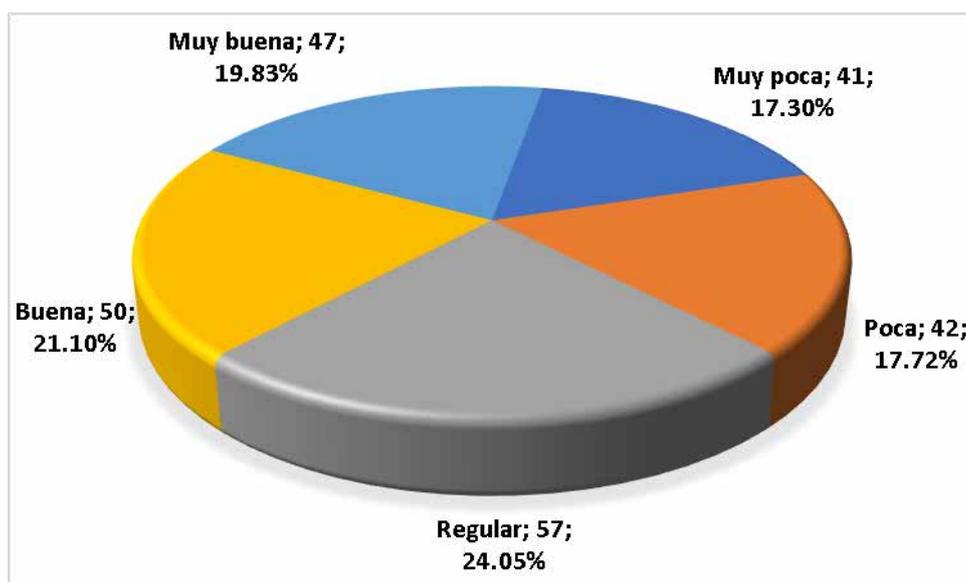
Conforme se observa de la figura anterior, la pregunta formulada busca obtener la apreciación del encuestado en relación a su disponibilidad de tiempo para cumplir sus labores de estudiante de posgrado; se obtuvo de la mayoría con un 24,47% la respuesta que mantienen una “muy buena” disponibilidad de tiempo, seguida de un 21,10% quienes refieren que su disponibilidad es “buena”, con un 20,68% que refieren tener una “muy poca” disponibilidad de tiempo y finalmente un 16,88% para ambas apreciaciones sobre una disponibilidad regular y de poco tiempo para dedicarse a sus estudios.



Fuente: Elaboración propia

Figura 26. Pregunta 16 – Considera que su nivel de estrés actual es

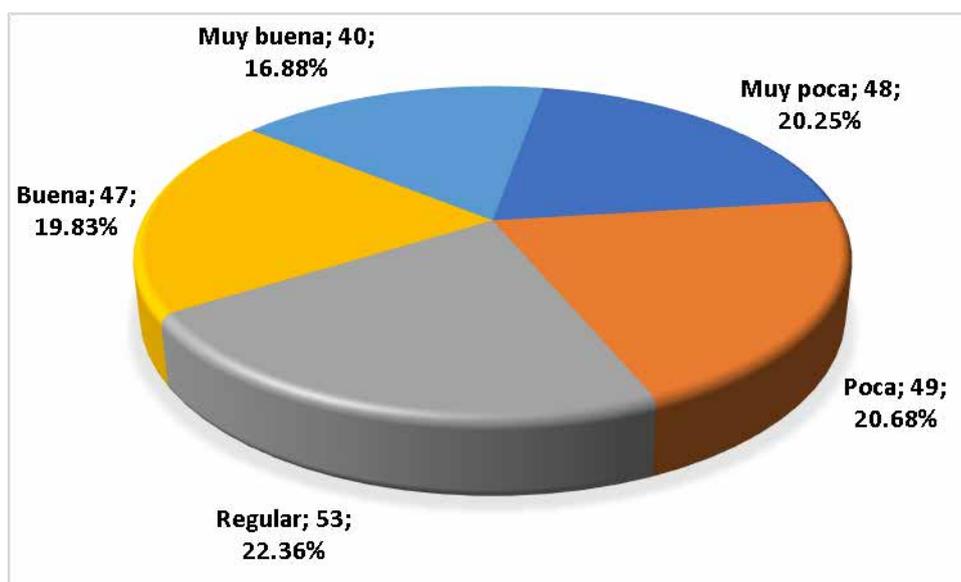
Conforme se observa en la figura anterior, y conforme a la formulación de la premisa de la pregunta se buscó determinar el nivel de estrés percibido por el estudiante de la muestra. Los resultados mostraron una opinión de similar porcentaje en todas las categorías desde bajo nivel de estrés con un 22,36%, seguido por un nivel muy alto y muy bajo con un 19,83%, hasta alto de estrés con un 19,41% y de regular nivel finalmente con 18,57%. Debido a la gran similitud de las frecuencias obtenidas, al parecer tal factor no es un factor determinante para la deserción de los estudios de posgrado.



Fuente: Elaboración propia

Figura 27. Pregunta 17 – ¿Considera que la escuela de Posgrado ofrece una infraestructura adecuada para la maestría?

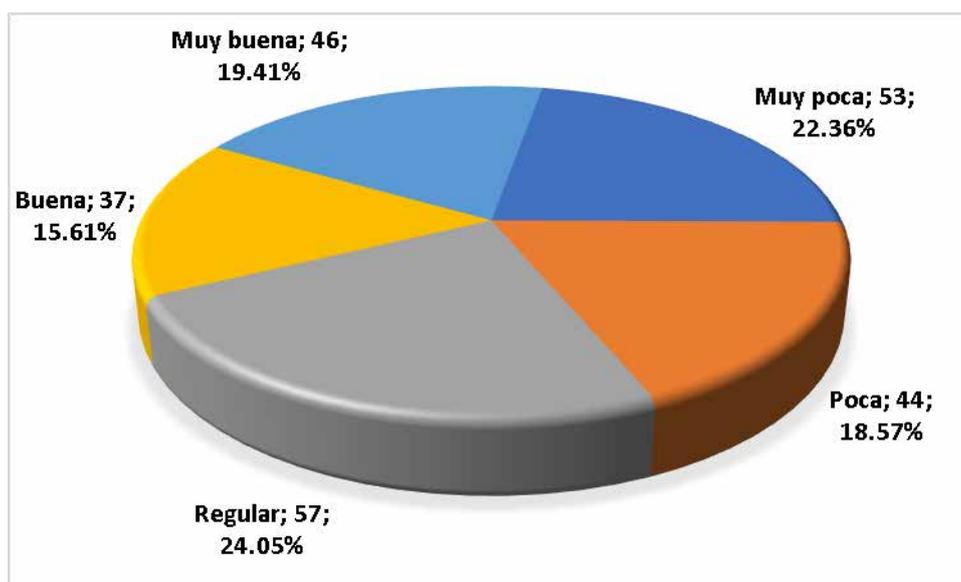
Conforme se observa en la figura anterior en relación a la distribución porcentual de estudiantes que consideran que escuela de posgrado de la UNJFSC mantiene una infraestructura adecuada para la realización de sus estudios de posgrado; los resultados muestran similares cantidades porcentuales en las respuestas desde muy buena con 19,83% a muy poca 17,30%. Dicha similitud en la distribución del porcentaje elegido parece señalar que no es determinante para la deserción estudiantil, sin embargo, ello debe ser materia de análisis estadístico.



Fuente: Elaboración propia

Figura 28. Pregunta 18 – ¿Considera que la escuela de Posgrado cuenta con el equipamiento y mobiliario adecuados para la maestría?

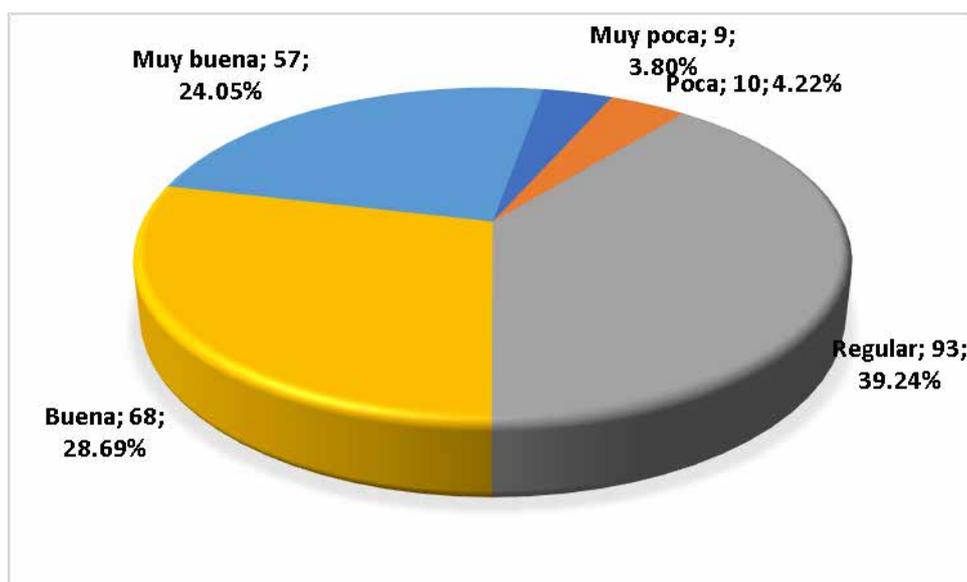
Conforme se observa en la figura anterior en relación a la distribución porcentual de estudiantes que consideran que escuela de posgrado de la UNJFSC cuenta con los equipos adecuados para la realización de sus estudios de posgrado; los resultados muestran similares cantidades porcentuales en las respuestas desde muy buena con 16,88% a muy poca 20,25%. Dicha similitud en la distribución del porcentaje elegido parece señalar que tal opinión no es determinante para la deserción estudiantil, sin embargo, ello debe ser materia de análisis estadístico.



Fuente: Elaboración propia

Figura 29. Pregunta 19 - ¿Considera que la escuela de Posgrado imparte las asignaturas correctas para la realización de sus estudios de posgrado?

Conforme se observa en la figura anterior en relación a la distribución porcentual de estudiantes que consideran que escuela de posgrado de la UNJFSC imparte las asignaturas correctas para la realización de sus estudios de posgrado; los resultados muestran similares cantidades porcentuales en las respuestas desde muy buena con 19,41% a muy poca 22,36%. Dicha similitud en la distribución del porcentaje elegido parece señalar que tal opinión no es determinante para la deserción estudiantil, sin embargo, ello debe ser materia de análisis estadístico.



Fuente: Elaboración propia

Figura 30. Pregunta 20 - ¿Considera que la escuela de Posgrado imparte las asignaturas adecuadas para la maestría?

Conforme se observa en la figura anterior en relación a la distribución porcentual de estudiantes que consideran que escuela de posgrado de la UNJFSC ofrece un buen nivel profesional de los docentes para la realización de sus estudios de posgrado; los resultados muestran similares cantidades porcentuales en las respuestas desde regular con un 39,24%, buena con 28,69%, a muy poca 3,80%. A priori, en atención a la dispersión de los porcentajes es posible que dicha opinión sea determinante para el la deserción estudiantil, sin embargo, ello debe ser materia de análisis estadístico.

4.2 Contrastación de hipótesis

4.2.1. Prueba de las hipótesis específicas

Prueba de hipótesis específicas

Prueba de hipótesis específica N°1: “Sí es posible determinar los factores que permiten predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.”

Conforme lo señalado en el capítulo en que se describe la metodología del presente estudio, se prepararon los datos obtenidos por el cuestionario al software Weka, considerándose los mismos como una base de datos de entrenamiento, es decir, sobre la cual el software analizaría y determinaría las relaciones estadísticas de los valores de los atributos (respuestas del cuestionario) que se conjugan para dar como resultado la predicción de una deserción del estudiante (clase), conforme se muestra en la siguiente figura:

```

Maestra Brenda categoria.arff Bloc de notas
Archivo Edición Formato Ver Ayuda
relation desertaron

@attribute 1 Rendimiento_académico_durante_la_secundaria real
@attribute 2 Asignaturas_reprobó_en_su_educación_secundaria real
@attribute 3 Repitió_año_educación_secundaria real
@attribute 4 Rendimiento_académico_pregrado real
@attribute 5 Cursos_desaprobó_pregrado real
@attribute 6 Sexo real
@attribute 7 Rango_edad real
@attribute 8 Estado_civil real
@attribute 9 Trabajo_relacionado_con_su_carrera_de_pregrado real
@attribute 10 Cantidad_de_hijos real
@attribute 11 Ingreso_familiar real
@attribute 12 Tiene_motivación_suficiente_para_culminar_posgrado real
@attribute 13 Situación_económica_actual real
@attribute 14 Considera_puede_financiar_sus_estudios_hasta_finalizar real
@attribute 15 Disponibilidad_de_tiempo_para_estudiar real
@attribute 16 Nivel_de_estrés_actual real
@attribute 17 Ofrece_una_infraestructura_adecuada real
@attribute 18 Cuenta_con_equipamiento_y_mobiliario_adecuados real
@attribute 19 Imparte_asignaturas_correctas real
@attribute 20 Ofrece_buen_nivel_de_docentes real
@attribute desercion {si,no}

@data
3,1,2,3,1,1,1,2,1,2,1,2,1,5,2,1,5,1,3,00
3,1,2,5,1,2,1,1,2,1,2,1,2,1,3,3,1,2,5,3,00
4,1,2,3,1,1,1,2,2,2,2,1,3,1,3,5,3,4,1,3,00
5,2,2,5,1,2,1,2,1,1,2,1,3,1,2,4,1,5,3,4,00
3,1,2,3,1,2,1,2,2,2,2,1,4,1,3,2,3,3,3,5,00
5,1,2,4,1,1,1,2,1,2,1,3,1,5,3,1,4,5,3,00
5,1,2,3,1,2,1,1,1,1,2,1,3,1,2,4,3,1,3,3,00
3,1,2,3,2,1,1,1,2,2,2,1,2,1,4,1,2,5,2,5,00
3,1,2,3,1,2,1,3,2,2,2,1,2,1,4,5,3,2,1,5,00
5,2,2,3,1,2,1,1,2,1,2,1,2,1,5,5,4,1,1,4,00
3,1,2,3,1,2,1,2,2,1,2,1,2,1,4,3,4,4,3,4,00
4,1,2,3,1,2,1,2,1,1,2,1,3,1,3,5,1,1,1,5,00
3,1,2,3,2,1,1,2,2,2,2,1,3,1,1,2,1,4,1,3,00

```

Fuente: Elaboración propia

Figura 31. Preparación de datos del cuestionario para ingreso como base de datos de entrenamiento al software Weka

Posteriormente, se realizó la carga de datos de la base de datos de entrenamiento, en las que el software Weka identificó las respuestas de las preguntas como un total de 20 atributos que definían una clase (deserción) que corresponde a la que contendrá la predicción sobre si el alumno procede a desertar de sus estudios categorizado (tipo nominal) definido, conforme se muestra en la siguiente figura:

En la figura anterior se muestra la primera parte del análisis realizado a la base de entrenamiento por el software Weka, en dicha sección se muestra las definiciones de los 20 atributos que corresponden a los valores hallados con la toma del cuestionario aplicado, los cuales son en su totalidad numéricos, pues se ha registrado solo la posición de cada alternativa de respuesta marcada en el cuestionario llenado por los estudiantes de la muestra y la clase deserción (mostrada al final de la lista) es de tipo categórica (nominal) conforme los valores de la información sobre su deserción real de los estudios de posgrado y corresponde al valor a predecir por el software.



Fuente: Elaboración propia

Figura 34. Vista 2 del análisis de la base de entrenamiento con el algoritmo J48

En las figuras 25 y 26 se muestra el detalle de la carga de datos realizada por el software Weka, así como el árbol de decisión generado por el software, el cual corresponde a la jerarquización por pesos que realiza Weka para realizar una predicción.

En las figuras anteriores se puede visualizar que el software fue capaz de clasificar correctamente (Correctly Classified instances) un 87,76% y erróneamente un 12,23%, ello corresponde a que de acuerdo al análisis de la base de entrenamiento fue capaz de determinar reglas entre los atributos que definieron la clase buscada (deserción) y realizar una predicción, la cual fue correcta la mayoría de las instancias, es decir, que acertó un total de 208 de las 237 instancias de análisis de las veces, lo cual corresponde a la exactitud del modelo. Sin embargo, el hecho que ha sido posible y que se generó un modelo capaz de realizar predicciones **permiten aceptar la hipótesis específica N° 1** pues “sí es posible determinar los factores que permiten predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC”.

Cabe resaltar que, en relación al análisis inferencial del resultado de las predicciones realizadas en la base de entrenamiento a través del **índice Kappa éste alcanzó un valor de 0,6634** (como se muestra en la figura 28), lo cual conforme la valoración de dicho índice señalada en la tabla 1 corresponde a una concordancia “considerable”, lo cual se refiere a la exactitud del modelo de predicción creado a partir de esta base de entrenamiento.

La base de entrenamiento analizada por el software Weka generó un árbol de decisión con los 20 atributos ingresados, el cual se muestra en la figura siguiente:

El árbol de decisión mostrado en la figura anterior indica que el software Weka encontró una fuerte asociación entre el resultado de la pregunta 3 - ¿Repetió algún año en la educación secundaria?, razón por la cual se encuentra en el primer nivel del árbol, en el segundo nivel determinó establecer a la pregunta 20 - ¿Considera que la escuela de Posgrado ofrece docentes con un buen nivel profesional?, como tercer nivel se ubicó a la pregunta 1 - Su rendimiento académico durante la secundaria fue..., el siguiente nivel se refiere a la pregunta 2 - ¿Cuántas asignaturas reprobó en su educación secundaria?, para continuar nuevamente con las preguntas 1,9,4 y 19 finalmente. Tal jerarquización en niveles permite determinar cuáles de los factores anotados y captados mediante el cuestionario obtuvieron un mayor peso o fueron determinantes para la predicción realizada por el software.

Prueba de hipótesis específica N°2: “La exactitud de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC es buena”.

Para la prueba de esta hipótesis debe recordarse que, conforme lo planteado para el presente trabajo la exactitud se calcula dividiendo el número total de registros correctamente clasificados por el número total de registros incorrectos o de referencia y expresándolo como porcentaje. Dicho resultado tomó el valor de 87,76% conforme se muestra en la figura 28, la cual fue utilizada para la prueba de la hipótesis específica 1. Con tal resultado es posible considerar que para que la predicción alcanzada por el modelo sea “muy buena” debió alcanzar al menos una exactitud del 75%, por lo que se considera que el modelo superó las expectativas de la hipótesis superando, alcanzando hasta una exactitud “muy buena”, ello al considerar la división de muy mala a muy buena en cuartiles de 25% cada una. En consideración a ello **se acepta la hipótesis específica 2**.

Prueba de hipótesis específica N°3: “La medida de concordancia de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC es buena.”

Para la prueba de esta hipótesis debe recordarse que, conforme lo planteado para el presente trabajo la concordancia se relaciona con el coeficiente Kappa de Cohen, el cual relaciona dos medidas realizadas con un método o instrumento distinto a fin de analizar el acierto de uno de ellos en relación al otro.

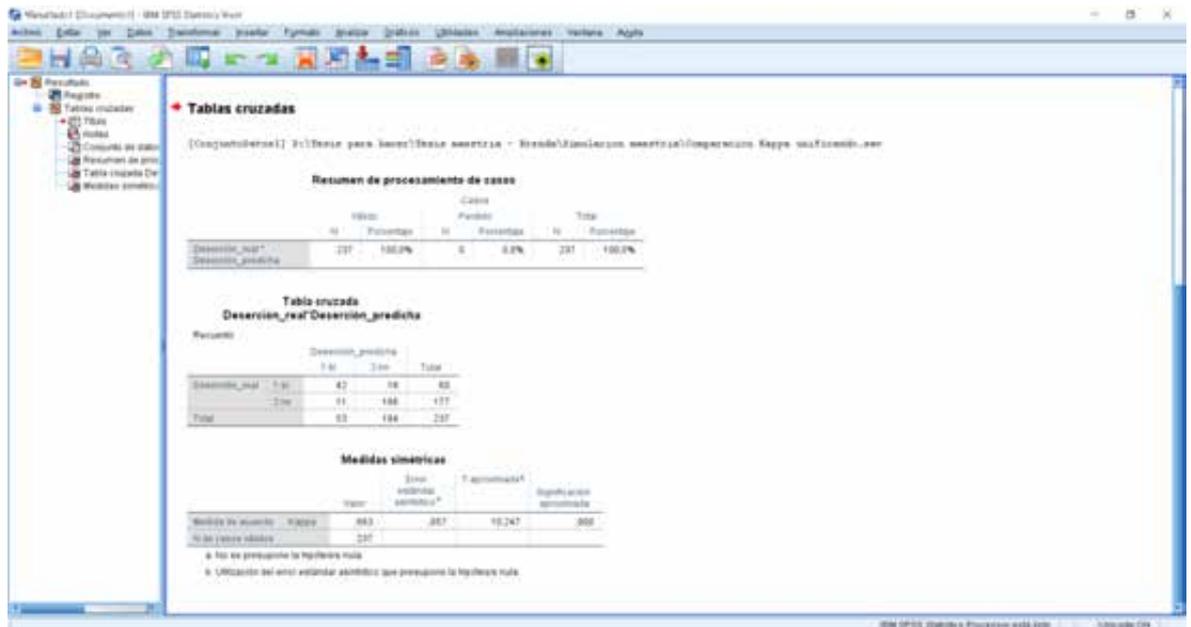


Figura 39. Resultado del análisis de coeficiente Kappa por SPSS v. 25.0

Evaluamos entonces el **valor del índice Kappa obtenido y este fue de 0,663** y corroborando la fuerza de esta relación a través de la tabla 1, se concluye que la concordancia de este modelo corresponde a “considerable” lo cual ya adjudica una más que buena confiabilidad a las predicciones realizadas por el modelo.

Por todo lo expuesto entonces se considera que ha quedado demostrada la hipótesis general de esta tesis.

Prueba de hipótesis General

Conforme la hipótesis general planteada para la presente tesis como:

H1: Es posible aplicar la metodología de minería de datos denominada “árboles de decisión” para crear un modelo de predicción del comportamiento de la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la UNJFSC.

Se considera que ésta ha quedado demostrada con la aceptación de la hipótesis específica 1, la cual señala que sí fue posible determinar los factores que permiten predecir la deserción de la población estudiantil de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC, ello se refuerza a la concordancia “considerable” alcanzada. Ante ello se puede señalar que incluso puede mejorarse dicho nivel con una base de datos de entrenamiento mayor.

CAPÍTULO V

DISCUSIÓN

5.1 Discusión de resultados

En la presente tesis se logró demostrar la hipótesis general formulada, relacionando las variables de estudio deserción estudiantil y técnicas de árbol de decisión a través de la generación de un modelo de predicción utilizando el software Weka y algoritmo de árboles de decisión J48, dicha metodología y resultado fue similar al realizado y obtenido por las investigaciones de Sadic, Abdulaziz, Fadl y Najoua (2018), Amaya, Barrientos y Heredia (2014) y Timarán, Caicedo e Hidalgo (2019); resaltando que no se utilizaron los modelos basados en técnicas multivariantes del estudio realizado por García (2015).

En relación a la exactitud del modelo de predicción generado en el software Weka utilizando la técnica de árboles de decisión para predecir la deserción de los estudiantes fue de un 87,76%, dicho resultado fue superior al obtenido por la investigación de Sadic, Abdulaziz, Fadl, & Najoua (2018) en su investigación que trató sobre la predicción del rendimiento académico estudiantil, obteniendo ellos un 73% con el algoritmo J48. También mayor al 80% de precisión de los modelos para los datos de entrenamiento y un 76% para los datos de validación del modelo generado en la investigación de Orihuela (2019).

También nuestro resultado en exactitud fue ligeramente mayor al porcentaje obtenido en la investigación de Yamao (2018), quien obtuvo un 82.87% de acierto utilizando árbol de decisiones. Pese a dicho resultado en menor porcentaje, se considera que es posible mejorar la exactitud priorizando las preguntas y atributos que analice el modelo para predecir la deserción estudiantil.

En cuanto a la concordancia, ésta fue estimada utilizando el índice Kappa, cuyo valor fue de 0,6634 (como se muestra en la figura 28), al cual le corresponde una concordancia “considerable”, tal consideración no fue analizada por ninguna de las investigaciones señaladas en las investigaciones consignadas en los antecedentes de la presente tesis. Además, se debe tomar en consideración que se generó un modelo de predicción el cual utiliza una jerarquía de pesos estadísticos para generar un árbol de decisiones, éste puede ser sujeto a un mayor entrenamiento con más casos reales a fin de elevar dicha concordancia.

CAPÍTULO VI

CONCLUSIONES Y RECOMENDACIONES

6.1 Conclusiones

1. Se logró demostrar la hipótesis general formulada, ello en razón a que sí fue posible aplicar la metodología de minería de datos denominada “árboles de decisión” para crear un modelo de predicción del comportamiento de la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la UNJFSC, lo cual demuestra que sí existe una correlación entre las variables de estudio deserción estudiantil y técnicas de árbol de decisión.
2. Se demostró la hipótesis específica N°1 por lo tanto, se acepta que “sí es posible determinar los factores que permiten predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC”, los cuales fueron jerarquizados en un árbol de decisiones generados por el software de predicción.
3. La hipótesis específica N°2 formulada como: “La exactitud de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC es buena” fue demostrada pues la exactitud con el modelo de predicción alcanzó a un valor de 87,76%.
4. La hipótesis específica N°3 formulada como: “La medida de concordancia de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC es buena” también fue demostrada pues la concordancia, medida a través del índice Kappa de Cohen alcanzó un valor de 0,66 al que le corresponde una concordancia “considerable” el cual es suficiente para considerarse como “buena”.

6.2 Recomendaciones

1. Al haberse demostrado la hipótesis general, la cual señala que es posible generar un modelo de predicción de la deserción estudiantil en la población de estudio utilizando el software Weka y los atributos identificados y priorizados a través de la técnica de árboles de decisión, establezca los procedimientos necesarios para que se capture la información necesaria a través del cuestionario planteado a los nuevos estudiantes y se efectúe la predicción con los datos obtenidos. Tal información podrá considerarse como un factor de apoyo al proceso de selección de postulantes a las maestrías correspondientes pues resultaría factible identificar los estudiantes que requieren un mayor apoyo psicológico o económico al resultar más propensos a la deserción.
2. Se recomienda además de la captura y realización de la predicción a través de este modelo, se ponga de conocimiento el resultado de la misma al estudiante, ello con la finalidad que sea consciente que estadísticamente presenta elementos que lo predisponen a la deserción por lo que debe tomar las medidas necesarias y cambios en sus actitudes y hábitos para evitar el mismo.
3. Se recomienda difundir los resultados del presente estudio en las facultades de la UNJFSC a fin de que otras escuelas profesionales interesadas puedan generar su propio modelo de predicción de la deserción y obtener más datos para identificar a los estudiantes que, debido a sus situaciones y características personales se encuentren predispuestos a caer en la deserción estudiantil, y tomar las acciones preventivas a tiempo.

REFERENCIAS

7.1 Fuentes bibliográficas

- Alania, P. (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*. Universidad Nacional Daniel Alcides Carrión, Escuela de Posgrado, Pasco.
- Amaya, Y., Barrientos, E., & Heredia, D. (2014). *Modelo predictivo de deserción estudiantil utilizando técnicas*. Obtenido de Redclara: <https://documentos.redclara.net/bitstream/10786/759/1/124-22-3-2014-Modelo%20predictivo%20de%20deserci%C3%B3n%20estudiantil%20utilizando%20t%C3%A9cnicas%20de%20miner%C3%ADa%20de%20datos.pdf>
- Camborda, Z. M. (2014). *Aplicación de árboles de decisión para la predicción del rendimiento académico de los estudiantes de los primeros ciclos de la carrera de ingeniería civil de la Universidad Continental*. Tesis para optar por el grado académico de Magister en Ingeniería de Sistemas , Universidad Nacional del Centro del Perú, Escuela de Posgrado, Huancayo.
- Canvia Oviedo, D. (2019). *Predicción del rendimiento académico de los estudiantes de la UNSAAC a partir de sus datos de ingreso utilizando algoritmos de aprendizaje automático*. Tesis para obtener el grado académico de Maestro en Informática, Universidad Nacional de San Antonio Abad del Cusco, Escuela de Posgrado, Cusco.
- Choque, V. (2019). *Minería de datos aplicada a la identificación de factores de deserción universitaria en programas de pre grado*. Tesis de Maestría, Universidad Nacional San Agustín de Arequipa, Escuela de Posgrado, Arequipa.
- Diaz, B. (2021). *Predicción del rendimiento académico utilizando la técnica de árboles de decisión en los programas de maestría de educación en la escuela de posgrado de la Universidad Nacional José Faustino Sánchez Carrión*. Tesis de grado, Universidad Nacional José Faustino Sánchez Carrión, Huacho. Obtenido de <http://repositorio.unjfsc.edu.pe/handle/UNJFSC/4395>
- Gonzales, C., & Rodriguez, C. (2017). *Propuesta de un Modelo de Business Intelligence para Identificar el perfil de deserción estudiantil en la Universidad Científica del*

- Sur. Tesis de maestría, Universidad Peruana de Ciencias Aplicadas (UPC), Escuela de Posgrado, Lima. Obtenido de <http://hdl.handle.net/10757/622749>
- Mamani, D. (2019). *Modelo de minería de datos basado en factores asociados para la predicción de deserción estudiantil universitaria*. Tesis de Grado, Universidad Nacional de Moquegua, Escuela profesional de Ingeniería de Sistemas e Informática, Moquegua.
- Osorio Calderón, P. (2010). *Propuesta de un Modelo de Simulación como herramienta en la justificación y comprensión de la toma de decisiones en la inversión pública. Un enfoque sistemático*. Tesis para optar por el título de Ingeniero de Sistemas, Universidad Industrial de Santander, Bucaramanga.
- Rivera Castellanos, C. (2010). *Construcción de un Modelo de comportamiento turístico con Dinámica de sistemas*. Trabajo de Graduación, Universidad de San Carlos de Guatemala, Facultad de Ingeniería, Guatemala.
- Santa Catalina, I. (2010). *Modelo de Dinámica de Sistemas para la implantación de Tecnologías de la Información en la Gestión Estratégica Universitaria*. Universidad del País Vasco, Donostia. Obtenido de <http://www.ehu.es/i.morlan/tesis/index.html>
- Vásquez, J., Castaño, E., Gallón, S., & Gómez, K. (2003). *Determinantes de la Deserción*. Universidad de Antioquia, Facultad de Ciencias Económicas, Medellín.
- Yamao, E. (2018). Predicción del rendimiento académico mediante minería de datos en estudiantes del primer ciclo de la escuela profesional de Ingeniería de Computación y Sistemas de la Universidad de San Martín de Porres, Lima-Perú. Universidad San Martín de Porres, Lima. Obtenido de http://www.repositorioacademico.usmp.edu.pe/bitstream/usmp/3555/3/yamao_e.pdf

7.2 Fuentes hemerográficas

- Aguirre, J., Valdovinos, R., Antonio, J., Alejo, R., & Marcial, J. (2015). Análisis de deserción escolar con minería de datos. *Research in Computing Science*(93), 71-82.
- Azoumana, K. (2013). Análisis de la deserción estudiantil en la Universidad Simón Bolívar, facultad Ingeniería de Sistemas, con técnicas de minería de datos. *Pensamiento Americano*, 6(10), 41-51. Obtenido de <http://coruniamericana.edu.co/publicaciones/ojs/index.php/pensamientoamericano>
- Claros, E., Cipriano, J., & Ramirez, F. (2016). Índice del Riesgo de deserción de los Estudiantes de la Facultad de Ciencias de la Universidad Nacional José Faustino

- Sánchez Carrión. *Big Bang Faustiniiano*, 5(2).
doi:<https://doi.org/10.51431/bbf.v5i2.15>
- Cu Balán, G. (2005). El impacto de la escuela de procedencia del nivel medio superior en el desempeño de los alumnos en el nivel universitario. *Revista Electrónica Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 3(1). Obtenido de http://www.ice.deusto.es/rinace/reice/vol3n1_e/Cu.pdf.
- Eckert, K., & Suénaga, R. (2015). Análisis de Deserción-Permanencia de Estudiantes Universitarios Utilizando Técnica de Clasificación en Minería de Datos. *Formación Universitaria*, 8(15). Obtenido de <http://www.redalyc.org/articulo.oa?id=373544192002>
- Kumar, S., Baharadwaj, B., & Pal, S. (2012). Data Mining Applications: A comparative Study for Predicting Student's performance. *International Journal Of Innovative Technology & Creative Engineering*, 1(1), 13-19.
- Manco, M. (2015). Modelo predictivo para la identificación de patrones de la deserción estudiantil en la Untels. *Untelciencia-Perú*, 1(1).
- Mori, M. (2012). Deserción universitaria en estudiantes de una universidad privada de Iquitos. *Docencia universitaria*, 6(1). Obtenido de <https://revistas.upc.edu.pe/index.php/docencia/article/view/42/11>
- Sadic, H., Abdulaziz, D., Fadl, M.-A., & Najoua, R. (Febrero de 2018). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), 447-459. doi:DOI: 10.11591/ijeecs.v9.i2
- Timarán, R., Caicedo, J., & Hidalgo, A. (2019). Árboles de decisión para predecir factores asociados al desempeño académico de estudiantes de bachillerato en las pruebas Saber 11°. *Rev.Investig.Desarro.Innov.*, 9(2), 363-378. doi:10.19053/20278306.v9.n2.2019.9184

7.3 Fuentes electrónicas

- Agencia Peruana de Noticias. (15 de Marzo de 2019). *MEF: Presupuesto público 2019 prioriza sectores de Educación y Salud*. Obtenido de Andina: <https://andina.pe/agencia/noticia-mef-presupuesto-publico-2019-prioriza-sectores-educacion-y-salud-734213.aspx>

- Córdoba, L. (16 de Junio de 2011). *Weka*. Obtenido de Minería de Datos: <http://cor-mineriadatos.blogspot.com/2011/06/weka.html>
- Gestión. (11 de Noviembre de 2018). Estas son las maestrías más solicitadas y sus costos en las principales universidades. *Gestión*. Obtenido de <https://gestion.pe/economia/management-empleo/son-maestrias-solicitadas-peru-costos-249579-noticia/>
- Microsoft. (2019). *Data Mining Concepts*. Recuperado el 27 de Diciembre de 2019, de <https://docs.microsoft.com/en-us/analysis-services/data-mining/data-mining-concepts>
- Orbegozo, A. (28 de Setiembre de 2020). Unos 174.000 estudiantes peruanos dejaron la universidad en lo que va del 2020. *El Comercio*. Obtenido de <https://elcomercio.pe/lima/sucesos/unos-174000-estudiantes-peruanos-dejaron-la-universidad-en-lo-que-va-del-2020-noticia/>
- Rivero, O. (3 de Mayo de 2017). *La filosofía de educación como fundamento teórico de la teoría educativa*. Obtenido de Red social educativa: <https://redsocialeducativa.net/la-filosofia-de-educacion-como-fundamento-teorico-de-la-teoria-educativa>
- SUNEDU. (2019). *Estadísticas de Universidades por Programa de estudios*. Recuperado el 19 de Marzo de 2019, de <https://www.sunedu.gob.pe/sibe/>
- UNJFSC. (2016). *Reglamento de Organización y Funciones*. Recuperado el 19 de Marzo de 2019, de http://www.transparencia.gob.pe/enlaces/pte_transparencia_enlaces.aspx?id_entidad=10045&id_tema=5&ver=D#.XJEKqSJKj3g
- UNJFSC. (2018). *Escuela de Posgrado*. Recuperado el 19 de Marzo de 2019, de Web institucional: <http://www.unjfsc.edu.pe/posgrado.php>
- UNJFSC. (2018). *Estatuto*. Obtenido de <http://www.unjfsc.edu.pe/archivos/emergente/Estatuto-2018-UNJFSC.pdf>
- WEEBLY. (2016). *Construcción, análisis y explotación de modelos*. Obtenido de Dinámica de Sistemas: <https://dinamicadesistemas.weebly.com/unidad-tres.html>

ANEXOS

ANEXO N° 1

TITULO: PREDICCIÓN DE LA DESERCIÓN ESTUDIANTIL UTILIZANDO LA TÉCNICA DE ÁRBOLES DE DECISIÓN EN LA ESCUELA DE POSGRADO DE LA UNIVERSIDAD NACIONAL JOSÉ FAUSTINO SÁNCHEZ CARRIÓN

PROBLEMAS	OBJETIVOS	HIPOTESIS	VARIABLES Y DIMENSIONES	METODOLOGIA
<u>Problema General:</u>	<u>Objetivo General:</u>	<u>Hipótesis General:</u>		
¿Es posible predecir la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la UNJFSC a través de la técnica de minería de datos de árboles de decisión?	Determinar si es posible aplicar la metodología de minería de datos denominada “árboles de decisión” para crear un modelo de predicción del comportamiento de la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la UNJFSC.	Es posible aplicar la metodología de minería de datos denominada “árboles de decisión” para crear un modelo de predicción del comportamiento de la deserción estudiantil de la población de los programas de maestría de educación de la Escuela de Posgrado de la UNJFSC.	<u>Variable 1:</u> Deserción estudiantil	POBLACIÓN: 237 alumnos de las maestrías del sector educación matriculados en el ciclo académico 2019-II de la Escuela de Posgrado de la UNJFSC.
<u>Problemas específicos:</u>	<u>Objetivos específicos:</u>	<u>Hipótesis específicas:</u>	<u>Dimensiones</u> Factores Académicos Factores Individuales Factores Ambientales Factores Institucionales	MUESTRA: Se utilizó el muestreo poblacional o censal.
1.- ¿Qué factores permiten predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC?	1.- Determinar los factores que permiten predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.	1.- Sí es posible determinar los factores que permiten predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.		TIPO: Tipo aplicada, de nivel correlacional, de diseño no experimental de corte transversal y enfoque mixto.
2.- ¿Cuál es la exactitud de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC?	2.- Determinar la exactitud de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.	2.- La exactitud de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC es buena.	<u>Variable 2:</u> Técnicas de árbol de decisión <u>Dimensiones</u> Aplicación de la técnica a través de un modelo de simulación	TECNICAS DE RECOLECCIÓN DE DATOS: Se utilizó el análisis documental y encuesta en relación a los factores de la deserción estudiantil.
3.- ¿Cuál es la medida de concordancia de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC?	3.- Calcular la medida de concordancia de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC.	3.- La medida de concordancia de la técnica de árboles de decisión para predecir la deserción estudiantil de la población de los programas de maestría del sector educación de la Escuela de Posgrado de la UNJFSC es buena.		PRUEBA ESTADÍSTICA: Se utiliza el índice Kappa para analizar la medida de la variabilidad (precisión) alcanzada por el modelo de predicción.

Anexo N° 2

Cuestionario de toma de datos

Gracias por realizar este cuestionario el cual permitirá analizar y señalar los factores de riesgo de deserción estudiantil y permitirá también generar un modelo para predecir la deserción de los alumnos de esta maestría en la Escuela de Posgrado.

Nombres y Apellidos del estudiante: _____

Maestría : _____

(Se guardará la reserva del caso y se usarán los datos solo para fines del presente estudio).

Nro.	Ítem	Marque su respuesta				
Factores académicos						
1	Su rendimiento académico durante la secundaria fue	Muy Malo	Malo	Regular	Bueno	Muy bueno
2	¿Cuántas asignaturas reprobó en su educación secundaria?	0	1	2 a 4	5 a 8	9 a más
3	¿Repetió algún año en la educación secundaria?	SI		NO		
4	Su rendimiento académico en la carrera de pregrado fue	Muy Malo	Malo	Regular	Bueno	Muy bueno
5	¿Cuántos cursos desaprobó en su carrera de pregrado?	0	1	2 a 4	5 a 8	9 a más
Factores individuales						
6	Identifique su sexo	M			F	
7	Rango de edad	20 a 30	31 a 40	41 a 50	51 a 60	61 a más
8	Estado civil	Soltero	Conviviente	Casado	Viudo	Divorciado
9	Percibe que su trabajo se encuentra relacionado con su carrera de pregrado	NO			SI	
10	Cantidad de hijos	0	1	2	3 a 4	5 a más
11	Ingreso familiar (en soles – considerar los ingresos sumados si ambos cónyuges trabajan)	1000 a 3000	3001 a 4000	4001 a 5000	5000 a 8000	8001 a más
12	¿Tiene la motivación suficiente para culminar sus estudios de posgrado?	SI			NO	

Factores ambientales						
13	Considera que su situación económica actual es	Muy Malo	Malo	Regular	Bueno	Muy bueno
14	Considera que puede financiar sus estudios hasta finalizar la maestría	SI		NO		
15	Considera que la disponibilidad de tiempo para realizar sus actividades de estudiante es ..	Muy poca	Poca	Regular	Buena	Muy buena
16	Considera que su nivel de estrés actual es	Muy bajo	Bajo	Regular	Alto	Muy alto
Factores institucionales						
¿Considera que la escuela de Posgrado...						
17	... ofrece una infraestructura adecuada para la maestría?	Muy poca	Poca	Regular	Buena	Muy buena
18	... cuenta con el equipamiento y mobiliario adecuados para la maestría?	Muy poca	Poca	Regular	Buena	Muy buena
19	... imparte las asignaturas adecuadas para la maestría?	Muy poca	Poca	Regular	Buena	Muy buena
20	... ofrece docentes con un buen nivel profesional?	Muy poca	Poca	Regular	Buena	Muy buena

M(o). ERLO WILFREDO LINO ESCOBAR
ASESOR

M(o). EDDY IVAN QUISPE SOTO
PRESIDENTE

M(o). CARLOS ENRIQUE BERNAL VALLADARES
SECRETARIO

M(o). RAÚL CHAVEZ ZAVALA
VOCAL