UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD: CIENCIAS DE LA INGENIERÍA Y APLICADAS

CARRERA: INGENIERÍA EN INFORMATICA Y SISTEMAS COMPUTACIONALES

PROYECTO DE INVESTIGACIÓN

Modelo de retención universitaria: Un enfoque de Machine Learning

AUTORES:

Urgiles Urgiles José Luis

Vásquez Mullo Marcia Salome

TUTORA:

Dr. Albán Taipe Mayra Susana

LATACUNGA – ECUADOR

DECLARATORIA DE AUTORÍA

Nosotros, Marcia Salome Vásquez Mullo con CI: 050263164-1 y José Luis Urgiles Urgiles con CI: 150053638-1, declaramos ser autores del presente proyecto de investigación: "Modelo de retención universitaria: Un enfoque de Machine Learning", siendo Ing. MSc. Mayra Susana Albán Taipe, con cédula de ciudadanía N°. 050231198-8 tutora del presente trabajo: y eximo expresamente a la Universidad Técnica de Cotopaxi y a sus representantes legales de posibles reclamos o acciones legales.

Además, certificamos que las ideas, conceptos, procedimientos y resultados vertidos en el presente trabajo investigativo, son de nuestra exclusiva responsabilidad.

Atentamente

_____	_____
Marcia Salome Vásquez Mullo	José Luis Urgiles Urgiles
CI:050263164-1	CI: 150053638-1

AVAL DEL TUTOR DE PROYECTO DE TITULACIÓN

En calidad de Tutora del Trabajo de Investigación sobre el título: "Modelo de retención universitaria: Un enfoque de Machine Learning", de los señores estudiantes Urgiles Urgiles José Luis, con cédula de ciudadanía 150053638-1 y Vásquez Mullo Marcia Salome, con cédula de ciudadanía 050263164-1 de la carrera de Ingeniería en Informática y Sistemas Computacionales, considero que dicho informe Investigativo cumple con los requerimientos metodológicos y aportes científicos-técnicas suficientes para ser sometidos a la evaluación del Tribunal de Validación de Proyecto que el Honorable Consejo Académico de la Facultad de Ciencias de la Ingeniería y Aplicadas de la Universidad Técnica de Cotopaxi designe, para su correspondiente estudio y calificación.

Latacunga 17 de septiembre, 2020

Tutora de Titulación
Dr. Mayra Albán Taipe
CC: 050231198-8

UNIVERSIDAD TÉCNICA DE COTOPAXI

Ingeniería Informatica Y Sistemas Computacionales

APROBACIÓN DEL TRIBUNAL DE TITULACIÓN

En calidad de Tribunal de Lectores, aprueban el presente Informe de Investigación de acuerdo a las disposiciones reglamentarias emitidas por la Universidad Técnica de Cotopaxi, y por la Facultad de Ciencias de la Ingeniería y Aplicadas; por cuanto, los postulantes: Vásquez Mullo Marcia Salome, con cédula de ciudadanía N° 050263164-1 y Urgiles Urgiles José Luis, con cédula de ciudadanía 150053638-1, con el título de proyecto de titulación: "Modelo de retención universitaria: Un enfoque de Machine Learning", han considerado las recomendaciones emitidas oportunamente y reúnen los méritos suficientes paras ser sometido al acto de Sustentación de Proyecto.

Por lo antes expuesto, se autoriza realizar los empastados correspondientes, según la normativa institucional.

Latacunga, 17 de septiembre, 2020

Para constancia firman:

Lector 1(Presidente)
Nombre: Ing. Llano Casa Alex Christian
CC: 050258986-4

Lector 2
Nombre: Ing. Mg. Cadena Moreno José
CC: 050155279-8

Lector 3
Nombre: Ing. Mg. Manuel William Villa Quishpe
CC: 180338695-0

AGRADECIMIENTO

A Dios por haberme brindado una familia maravillosa, a mi padre y hermanos por apoyarme en todo lo que podían, a mi abuelita por sus oraciones, y a mi tía, agradecerle a todo por su apoyo en mis estudios. Agradecer a mis amigos por enseñarme y darme ánimos en momentos de tristeza, por el apoyo y su amistad que me supieron brindar.

A mis maestros por haberme brindado su conocimiento, que es un tesoro para cualquier estudiante. La paciencia que tuvieron al enseñar en cada uno de sus clases.

José Luis Urgiles Urgile

AGRADECIMIENTO

A Dios, por permitir mi existencia, y darme la sabiduría y la fuerza para lograr las metas establecidas, a mis padres que siempre me han apoyado sin dejarme desfallecer, y por la comprensión, el apoyo y compañía de mi mujercita durante las noches agotadoras.

También agradezco a mis maestros por orientar mi formación académica especialmente a la Ing. MSc. Mayra Albán, como mentora, me brindo su conocimiento y apoyo que guio todas las etapas de este proyecto de investigación para lograr mis objetivos deseados.

Marcia Salomé Vásquez Mullo

DEDICATORIA

Mi investigación actual está dedicada a Dios, porque me da sabiduría y me guía cuando me encuentro en dificultades, estas dificultades me han permitido mostrar la capacidad de superarme.

A mi madre Amada Vásquez quien estuvo al tanto de este proceso y me ánimo a concluir sin dejarme desfallecer, y a mi padre Marcos Laica que a pesar de la distancia ha estado presente para inculcarme y animarme a seguir adelante en situaciones difíciles, gracias por su ayuda, y todo el apoyo incondicional que me brindaron y me enseñaron que mientras haya dedicación todo se puede lograr.

Finalmente, quiero dedicar mi proyecto de investigación a mi mujercita Katherine Casa, ella quien me brindo su comprensión permanente, porque mi inspiración viene de su presencia animándome a seguir adelante, todo esfuerzo vale la pena mi niña, gracias por tu comprensión y tu amor incondicional

Marcia Salomé Vásquez Mullo

DEDICATORIA

Este proyecto de investigación, se la dedico a mi padre Luis Hernán Urgiles Urgiles y mi abuelita Luz Marina Urgiles Urgiles quienes fueron los que estaban presentes en toda mi carrera de formación.

Finalmente, dedico a mis hermanos, por su contribución que me brindaron en toda esta etapa de mi formación profesional.

José Luis Urgiles Urgiles

ÍNDICE DE CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE FIGURAS

INDICIE DE ECUACIONES

UNIVERSIDAD TÉCNICA DE COTOPAXI

FACULTAD: CIENCIAS DE LA INGENIERÍA Y APLICADAS

CARRERA: INGENIERÍA EN INFORMATICA Y SISTEMAS COMPUTACIONALES

TITULO: Modelo de retención universitaria: Un enfoque de Machine Learning

Autores

Urgiles Urgiles José Luis

Vásquez Mullo Marcia Salome

RESUMEN

La retención universitaria se ha convertido en un fenómeno reconocido mundialmente, debido a su complejidad y múltiples causas que deben ser tratadas en el entorno universitario, la disminución de sus tasas genera dificultades de orden académico y de gestión para las Instituciones de Educación Superior. Se considera importante analizar los temas de retención como medio para mitigar problemas que afectan al estudiante y que permita la culminación con éxito de una carrera profesional. Por tal razón, se propone un modelo para identificar factores de retención estudiantil universitaria basada en la aplicación de técnicas de Machine Learnign. Los datos se obtienen de un proceso de levantamiento de información por medio de una encuesta a 294 estudiantes de una universidad pública del Ecuador, para el desarrollo de la investigación se utiliza la metodología Knowledge Discovery in Database (KDD) y algoritmos de aprendizaje supervisado como redes neuronales. Los resultados permiten diseñar un modelo conceptual basado en 7 factores que influyen en la retención de los estudiantes en las universidades utilizando Regresión Lineal. Para el proceso de predicción se utilizó algoritmos Clúster y Redes Neuronales, dando como resultado una tasa de precisión del 94.2% con el modelo Multilayer Perceptrom, lo que permite determinar que la investigación desarrollada se sustenta bajo el procedimiento experimental que comprueba la validez del modelo conceptual propuesto.

Palabras clave: retención de estudiantes, Machine Learning, K-means, clúster, redes neuronales.

COTOPAXI TECHNICAL UNIVERSITY

FACULTY: ENGINEERING SCIENCES AND APPLIED

CAREER: ENGINEERING IN INFORMATICS AND COMPUTER SYSTEMS

TITLE: College Retention Model: A Machine Learning Approach

Authors

Urgiles Urgiles José Luis

Vásquez Mullo Marcia Salome

ABSTRACT

University retention has become a globally recognized phenomenon, due to its complexity and multiple causes that must be addressed in the university environment; the decrease in its rates generates academic and management difficulties for Higher Education Institutions. It is considered important to analyze retention issues as a means of mitigating problems that affect the student and that allow the successful completion of a career. For this reason, a model is proposed to identify university student retention factors based on the application of Machine Learnign techniques. The data is obtained from an information survey process through a survey of 294 students from a public university in Ecuador, for the development of research the Knowledge Discovery in Database (KDD) methodology and supervised learning algorithms with neural networks are used. The results allow to design a conceptual model based on 7 factors that influence the retention of students in universities in universities using Linear Regression. For the prediction, Cluster and Neural Networks algorithms were used, resulting in an accuracy rate of the proposed models of 94.2% with the Multilayer Perceptrom model, which allows to determine that the research developed is based under the experimental procedure that checks the validity of the proposed conceptual model.

Keywords: student retention, Machine Learning, K-means, cluster, neural networks,

Universidad Técnica de Cotopaxi

CENTRO DE IDIOMAS

AVAL DE TRADUCCIÓN

En calidad de Docente del Idioma Inglés del Centro de Idiomas de la Universidad Técnica de Cotopaxi; en forma legal **CERTIFICO** que: La traducción del tema de tesis al Idioma Inglés presentado por los señores egresados de la **CARRERA DE INGENERÍA EN INFORMÁTICA Y SISTEMAS COMPUTACIONALES DE LA FACULTAD DE CIENCIAS DE LA INGENIERÍA Y APLICADAS: URGILES URGILES JOSÉ LUIS, VÁSQUEZ MULLO MARCIA SALOME**, cuyo título versa **"MODELO DE RETENCIÓN UNIVERSITARIA: UN ENFOQUE DE MACHINE LEARNING"**, lo realizaron bajo mi supervisión y cumple con una correcta estructura gramatical del Idioma.

Es todo cuanto puedo certificar en honor a la verdad y autorizo a los peticionarios hacer uso del presente certificado de la manera ética que estimaren conveniente.

Latacunga, septiembre del 2020

Atentamente,

MSc. Alison Mena Barthelotty
DOCENTE CENTRO DE IDIOMAS
C.C. 0501801252

CENTRO DE IDIOMAS

1. INFORMACIÓN GENERAL

Título:

Modelo de Retención Universitaria: Un enfoque de Machine Learning

Fecha de inicio:

Mayo 2020

Fecha de finalización:

Septiembre 2020

Lugar de ejecución:

Universidad Técnica de Cotopaxi Facultad de Ciencias de la Ingeniería y Aplicadas

Facultad que auspicia

Ciencias de la Ingeniería y Aplicadas

Carrera que auspicia:

Ingeniería en Informática y Sistemas Computacionales

Proyecto de investigación vinculado:

Modelamiento de Algoritmos para Sistemas de Información

Equipo de Trabajo:

Tutora de Titulación:

Dr. Mayra Susana Albán Taipe

Autores:

José Luis Urgiles Urgiles

Marcia Salome Vásquez Mullo

Área de Conocimiento: Ciencias

Sub- Área: Informática

Línea de investigación:

Tecnologías de la Información y Comunicación (TICS)

Sub líneas de investigación de la Carrera:

Inteligencia Artificial y Robótica

2. DESCRIPCION DEL PROYECTO

La retención estudiantil ha cobrado relevancia en el campo educativo, debido a la sentida necesidad de generar alternativas para la permanencia y graduación de los estudiantes, así como también, por la necesidad de establecer estrategias que permitan contribuir a la consecución de metas académicas universitarias. La propuesta de investigación se basa en el diseño de un modelo conceptual determinado por factores que influyen en la retención de los estudiantes en las universidades, además, se aplica un enfoque de Machine Learning a través del uso de algoritmos de aprendizaje supervisado y no supervisado tales como Clúster y Redes Neuronales mediante el algoritmo Multilayer Perceptrom y Voted. El diseño de la investigación es descriptivo y de corte mixto, cuyo análisis se centra en la descripción de factores para la construcción del modelo conceptual y cuya aplicación servirá como una herramienta de apoyo para la toma de decisiones de los directivos de las instituciones universitarias, para establecer, estrategias o actividades encaminadas a enfrentar las dificultades que se presentan en el contexto académico universitario.

3. JUSTIFICACIÓN DEL PROYECTO

La retención universitaria para los autores Armijo, Zárate & Carvajal (2019) es la persistencia como un fenómeno que ha permitido generar estrategias de aprendizaje y focalizar nuevos procesos de participación curricular. En su estudio, realizado en Chile se establece que la

retención de los estudiantes en las universidades puede contribuir a mejorar el sistema de Educación Superior a través del desarrollo integral, curricular, y social del estudiante; así como también, la unificación de conocimientos adquiridos conjuntamente con los mentores, académicos y el equipo profesional para responder a las necesidades y permanencia en las universidades. Se puede señalar, que en los últimos tiempos el estado chileno ha promovido diferentes recursos, como implementar programas de mejoramiento de la calidad, así como también el financiamiento para más estrategias y nivelar las oportunidades de conocimiento en los estudiantes, contribuyendo a enfatizar la permanencia estudiantil en la Educación Superior (SIES, 2014).

Mientras que Urbina & Ovalles (2016) señala que la permanencia universitaria es considerada como la motivación, parte fundamental que impulsa la integración académica y social que permite incrementar avances a partir de la adaptación curricular en los estudiantes, así como también, innovadores proyectos educativos implementadas en las Instituciones de Educación Superior, que contribuye a obtener mayor conocimiento y éxito mejorando el acercamiento a la educación profesional.

De igual manera, el autor Frutos (2017) señala que es necesario diseñar formas de asegurar el aumento de la retención estudiantil en las carreras universitarias, por medio de estrategias y herramientas que permitan al estudiante culminar su formación profesional, así como también, menciona que la tasa de retención es un indicador de evaluación y acreditación de carreras, que permite establecer estrategias que contribuyen al aumento del índice de persistencia.

Dentro de este análisis de la retención universitaria se puede resaltar el modelo predictivo propuesto por Aguilera, Campos & Vera, (2017) quienes buscan predecir la permanencia a partir de atributos con estudiantes previo al ingreso a la universidad en el primer semestre del año 2017. El proceso estuvo constituido con el análisis estadístico de regresión lineal, para determinar la variable del objeto de estudio, los autores realizan un proceso de análisis basados en 58 estudios a estudiantes que ingresaron a la universidad a través del Programa de Acompañamiento y Acceso Efectivo en la Educación Superior. Se realizó el análisis a partir de características personales y de educación formativas para la universidad, así como también, el compromiso del estudiante, el sistema académico, y por ultimó la integración universitaria. Para el procesamiento de datos se utilizó el programa R-Project que contribuyo a identificar factores

que influyen en la retención estudiantil universitaria en la Universidad de Playa Ancha de Ciencias de la Educación en Chile.

Otro de los estudios relevantes dentro de la literatura es el de Ferrão & Almeida (2018) que presentan un modelo multinivel de persistencia con la análisis estadísticos que incluyen técnicas de regresión logística de multinivel de datos. El estudio se basó en 2.697 datos de estudiantes matriculados en el primer año en la universidad pública de Minho en el año académico 2015-2106. Se analizó los atributos persistencia, admisión, sexo, semestre de carrera, trayectoria escolar, estatus socioeconómico, entre otros. El procedimiento experimental comprendía entornos basados en la estadística mediante el software MLwinN2.3.1, el cual permitió conocer las condiciones de admisión, y el semestre que cursa el estudiante contribuyendo con factores importantes para la persistencia en la Educación Superior en Portugal.

Por lo expuesto anteriormente se considera importante el desarrollo de un modelo de retención universitaria que permita la determinación de factores de éxito en la retención estudiantil universitaria, el mismo que será utilizado como una herramienta de apoyo para la toma de decisiones oportunas por parte de los administradores de las universidades. El modelo consta de experimentos validados bajo el entorno de la regresión lineal simple y la construcción de modelos de predicción mediante el uso del clúster y redes neuronales, modelos predictivos validados mediante métricas de precisión de la predicción que permiten obtener una alta fiabilidad en sus resultados y que determinan la confiabilidad de los modelos propuesto resultados.

4. BENEFICIARIOS DEL PROYECTO

Se considera como beneficiarios directos del proyecto los estudiantes que son la razón de ser de las universidades. Los beneficiarios indirectos corresponden a toda la comunidad universitaria y padres de familia.

5. PROBLEMA DE INVESTIGACIÓN

La retención universitaria se ha convertido en un fenómeno reconocido mundialmente, debido a su complejidad y múltiples causas que influyen en el contexto del desarrollo universitario

(Celada & Lattuada, 2018). Es importante mencionar que las instituciones de educación superior deben considerar a la permanencia estudiantil como un indicador de gestión, debido a que la disminución de sus tasas no permiten alcanzar las metas académicas trazadas para la conclusión de la etapa profesional y la obtención de un título universitario (Velázquez & González, 2017). Además, se identifica como problema principal de este análisis objeto de estudio la cantidad de estudiantes que no continúan con su proceso de formación académica, de cada cien estudiantes cincuenta completan su formación profesional terciaria (Andrea, 2018).

La literatura ha permitido identificar investigaciones relacionadas con la retención estudiantil desde hace algunas décadas atrás. Tal es el caso, del estudio realizado por Spady (1970) quien sostiene que la falta de integración de los alumnos al sistema de Educación Superior probablemente influye en la decisión de abandono de los estudiantes en las aulas universitarias. Tinto (1973) presenta un modelo de deserción universitaria en el que analiza factores de integración del estudiante en el entorno académico interno y externo, el autor hace referencia a las dificultades de integración que pueden suscitarse en el desarrollo académico del estudiante que pueden afectar a los procesos de retención estudiantil y a la obtención de las metas académicas propuesta por los alumnos.

Así también, Fishbein & Ajzen (1975) argumenta que las actitudes, intenciones y conductas podrían influir en el comportamiento de los estudiantes induciendo a la disminución de la permanencia en los estudios universitarios. De igual manera, Tinto (1975) identifica al abandono universitario como base principal de su estudio, por medio de comportamientos que busca identificar las causas por las cuales el estudiante decide retirarse de sus programas académicos de estudio. Bean (1980) señala la relación de compromiso que tienen los estudiantes con su proceso de formación profesional recibida y la satisfacción de los beneficios académicos que otorga la institución buscando la trasferencia para el cambio de institución, según el autor una disminución en la satisfacción percibida de su entorno académico e institucional podría influir en la disección de los alumnos de cambiar de institución superior. Ethington (1990) considera que el rendimiento académico, influye en actores sociales y curriculares que afecta el nivel de aspiraciones y futuro de los estudiantes lo que podría influir en una posible disminución de la permanencia estudiantil que causa efectos negativos para los estudiantes, padres de familia e instituciones. Por otro lado, Tinto (1993) sostiene en su estudio que los

factores de integración social y académica son determinantes a la hora de tomar una decisión sobre la permanencia estudiantil del alumno.

Según, datos del Banco Mundial en el año 2017 citado por Munizaga, Cifuentes y Beltrán (2018) en América Latina y el Caribe la tasa de retención universitaria corresponde aproximadamente al 46%, el 32% de estudiantes se presume que pueden retornar a las actividades académicas en años posteriores, mientras que el 22% restante abandonaron las Instituciones de Educación Superior. En un estudio realizado en la Pontificia Universidad de Chile se obtiene una tasa de retención del 55%, lo que señala que existen limitaciones en los procesos de permanencia estudiantil y en la no consecución de la obtención del título académico para los estudiantes (García., et. all., 2017).

Además, el Departamento de Educación de los Estados Unidos, Centro Nacional de Educación Estadísticas (2015) citado por Carut (2018) en este país se registró una tasa retención del 59% durante los años 2011- 2012. Durante los últimos 20 años la tasa de persistencia no ha cambiado a pesar del análisis y estudio continuo de aspectos que influyen negativamente para elevar la tasa de retención. En México, el acceso a la educación superior es considerado uno de los más bajos de acuerdo a datos obtenidos de la Organización para la Cooperación y el Desarrollo Económico OCDE, ya que se estima que solo un tercio de la población puede acceder a la universidad, de cada 100 niños que empiezan la educación primaria solo 21 se gradúan de la universidad. Los problemas más relevantes que afectan a la educación terciaria en este país se encuentran relacionados con la inequidad social, costo de la educación, calidad de la educación superior, y la perspectiva que tiene el estudiante sobre la inserción en el campo laboral (Castiello-Gutiérrez, 2019).

En Ecuador, según datos del SENESCYT en el año 2016 señalado por Ganuza, Rodríguez & Auccahualpac (2017) 7 de cada 10 alumnos continúan en la universidad después de sus dos primeros años de estudio, sin embargo, estas cantidades de estudiantes suelen disminuir en el transcurso de la formación académica. Se podría pensar que en el sistema educativo superior ecuatoriano la medición de la eficiencia se identifica en la cantidad de alumnos matriculados en relación al número de títulos obtenidos en el tiempo oficial.

Sin embargo, en los países europeos se presentan problemas relacionados con el acceso y la retención de estudiantes universitarios no tradicionales tanto en jóvenes como adultos. El autor destaca que la tasa de retención es un determinante en la eficiencia de la educación superior. Además, encuentra problemas relacionados con la población estudiantil diversa, la perspectiva institucional, ampliación del aprendizaje permanente, que se ve involucrado con los estudiantes, gobierno estatal y las instituciones de educación superior mencionado en el proyecto Ranlhe (Merrill, 2011).

Por otro lado, en Buenos Aires según Lattuada (2017) en el debate universitario indica que la persistencia y la graduación de estudiantes son problemas que se presentan en la mayoría de las universidades y de acuerdo a las tasas que estos presenten dependen la existencia o desaparición de unas cuantas universidades. También inciden las matrículas en las universidades privadas y en el presupuesto en el caso de universidades públicas. Además, las altas tasas de deserción se manifiestan en los primeros años de la formación académica en las Instituciones de Educación Superior.

Se pueden determinar otros factores que influyen en la persistencia estudiantil y que están relacionados con la aptitud, preparación previa, compromiso, responsabilidad, percepción de dificultad de la carrera, integración académica, habilidades de comunicación, clima institucional, proceso de admisión, calidad académica, entre otras. Estos factores identificados pueden ser considerados como indicadores sociales, cognitivos e institucionales, que afectan a la retención estudiantil (Chanchí et al., 2019).

Cabe destacar, que la literatura aporta diferentes enfoques de investigación, entre los que mencionaremos a los autores Ayala & Atencio (2018) y el modelo explicativo, en el que señala que los antecedentes académicos, familiares y económicos son factores determinantes para la retención estudiantil. Además, los investigadores encontraron cambios en los efectos causales que están estrechamente relacionados con la retención. También, se analiza los problemas relacionados con la institución universitaria, debido a las dificultades que se presentan hasta el proceso de titulación de los estudiantes en la educación superior.

Adicionalmente se pudo identificar, modelos de retención estudiantil basados en enfoques de machine learning para tratar de solventar problemas relacionados con la retención estudiantil,

se evidencia un incremento de la producción científica al año 2020 relacionado con el tema de estudio, lo que indica la importancia de tratar a la retención estudiantil como un tema de importancia dentro de la comunidad científica para incrementar sus tasas y una adecuado toma de decisiones por parte de los administradores de las instituciones de educación superior.

5.1 Formulación del Problema

¿Cuáles son los factores influyen en la retención estudiantil universitaria?
¿Cuáles son los enfoques de machine learning para tratar la retención estudiantil universitaria?

6. OBJETIVOS

a) General

Desarrollar un modelo para determinar factores de éxito de retención universitaria a través del uso de algoritmos de aprendizaje supervisado y no supervisado.

b) Específicos

- ✓ Revisar sistemáticamente la literatura para conocer el estado actual del tema de objeto de estudio en torno al desarrollo de conceptos y teorías que permitan sustentar teóricamente las preguntas de investigación e hipótesis.

- ✓ Diseñar un modelo conceptual para determinar las causales que afectan a la retención estudiantil universitaria basado en la aplicación de regresión lineal simple.

- ✓ Validar el modelo teórico a partir de la construcción de modelos de predicción y clasificación mediante el uso de clúster y redes neuronales y establecer la tasa de precisión de la predicción de los modelos propuestos.

7. ACTIVIDADES Y TAREAS DE LOS OBJETIVOS

Tabla 1: Tareas planteadas de la investigación

Objetivos	Tareas	Resultados de las Tareas	Instrumentos
Revisar sistemáticamente la literatura para conocer el estado actual del tema de objeto de estudio en torno al desarrollo de conceptos y teorías que permitan sustentar teóricamente las preguntas de investigación e hipótesis.	Búsqueda de información en fuentes primarias de investigación científica. Selección de documentos acorde al tema de investigación Análisis de contenidos de los documentos seleccionados	Documentos seleccionados Marco Teórico	Metodología de Bárbara Kitchenham para revisiones del estado del arte
Diseñar un modelo conceptual para determinar las causales que afectan a la retención estudiantil universitaria basado en la aplicación de regresión lineal simple.	-Selección de variables. Pre-procesamiento (Limpieza y transformación de datos) Análisis de coeficientes Interpretación de variables	Modelo original Modelo ajustado Modelo Teórico	Se utilizó la técnica descriptiva de regresión lineal en el software Spss
Validar el modelo teórico a partir de la construcción de modelos de predicción y clasificación mediante el uso de clúster y redes neuronales y establecer la tasa de precisión de la predicción de los modelos propuestos.	Extracción del conocimiento (Minería de datos) Interpretación y evaluación.	Aplicación de métricas de evaluación Evaluación de resultados Tasas de Predicción	Se utilizó técnicas no supervisadas Clúster k-means, y técnicas supervisadas redes (Perceptrom Multilayer, Voted Perceptrom) en la predicción Métricas de evaluación: técnica accuracy, thresold Curve.

Fuente: grupo de trabajo

8. FUNDAMENTACIÓN CIENTÍFICO TÉCNICA.

8.1 Antecedentes.

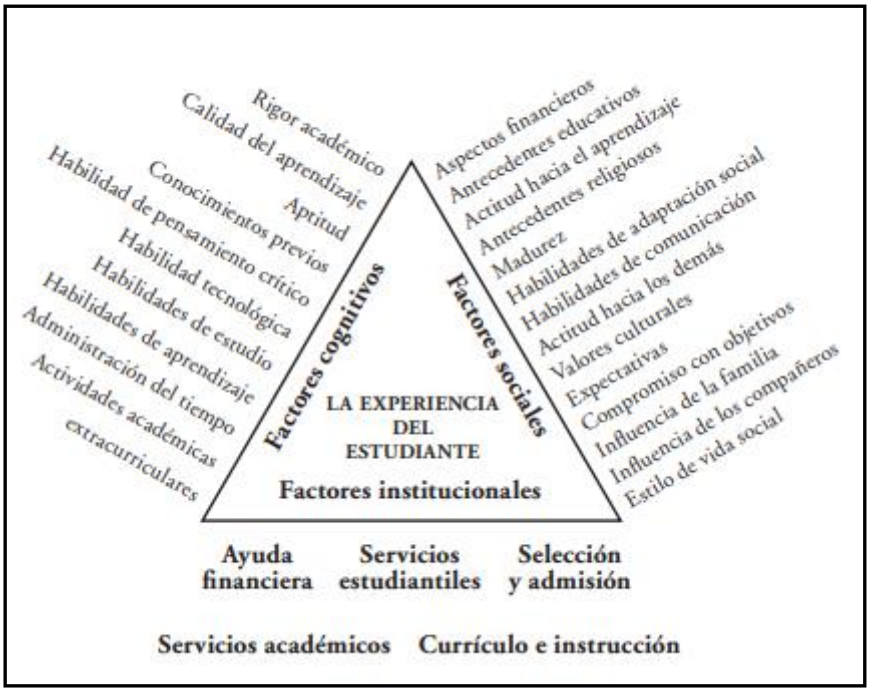
Las primeras indagaciones en la retención estudiantil a nivel mundial aproximadamente parten desde la década de los años 70, tal es el caso de Spady (1970) que presenta con respecto del

Modelo Sociólogo en cual manifiesta la dificultad que tienen los estudiantes de integrarse a la educación superior. Tinto (1973) señala el modelo organizacional que hace referencia a la importancia que prestan las autoridades en cuanto al aumento a la tasa de abandono, así mismo, menciona que los costos que implica la educación estudiantil también infieren en la decisión de abandonar o permanecer en la educación superior. Así también, Fishbein & Ajzen (1975) muestra el modelo psicológico que determina factores como las actitudes, intenciones y conductas que influyen en el comportamiento de los estudiantes induciendo al abandono o continuación con la preparación académica. De igual manera, Tinto (1975) en su investigación hace referencia a un modelo teórico que busca entender los motivos por los cuales los estudiantes abandonan la educación superior, en esta investigación se identifica, factores como el retiro voluntario, que presenta un impacto muy relevante en la institución educativa, y en la consecución de estrategias que permita contribuir con la flexibilidad de la educación superior. Ethington (1990) postula el modelo teórico mediante el estudio de variables socioeconómicas, logros previos y aspiraciones de grado, encontrando un fuerte impacto de las variables o factores determinados como: valor otorgado por la universidad, la asistencia y la expectativa de éxito que tiene una influencia directa en la persistencia del estudiante universitario. Por su parte, Tinto (1993) en su teoría de la retención busca la integración institucional, y social del estudiante, así como también, resalta la importancia que tiene experiencia adquirida por el estudiante en el aula, el aprendizaje permite la persistencia del estudiante para alcanzar las metas académicas.

8.2 Principales Referentes Teóricos

En la investigación se destaca el modelo geométrico de durabilidad y consecución de objetivos, propuesto Swail (1995), menciona la retención a partir de la relación existente entre las capacidades adquiridas por el estudiante con las cuales llega al campus universitario y las capacidades que a fututo se desarrollaran en la institución a través de su formación académica. Según el autor, el modelo propuesto discute las dinámicas entre aspectos cognitivos, sociales y factores institucionales, una adecuada combinación de estos elementos de manera positiva puede contribuir a que el estudiante consiga su proceso de graduación desarrollo y persistencia.

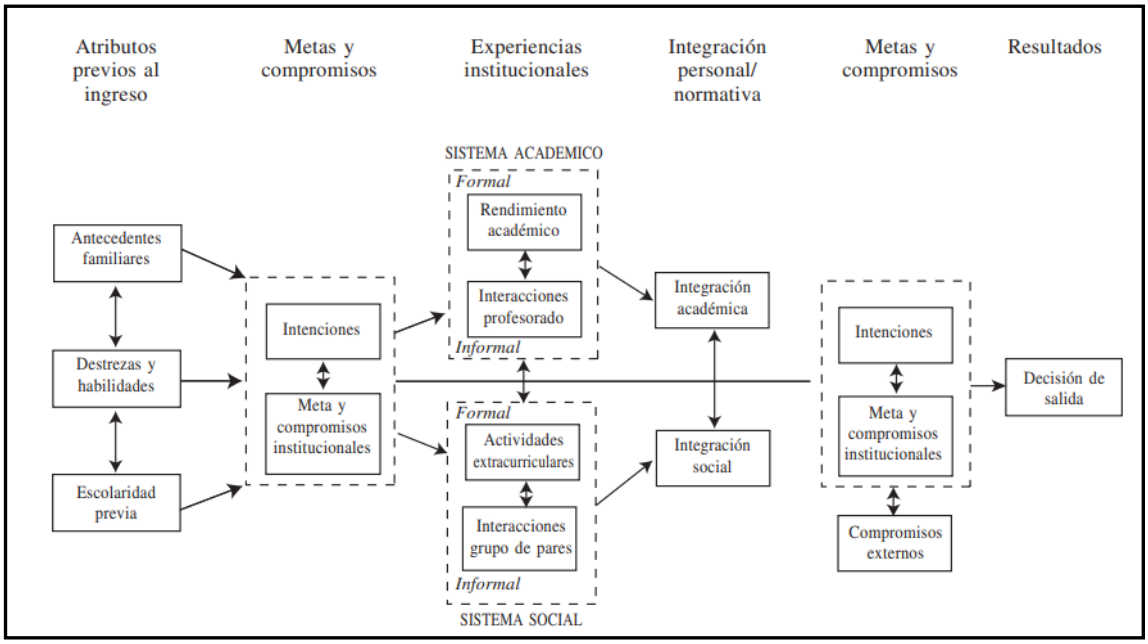
Figura 1: Modelo Geométrico de Persistencia y Logro de los Objetivos.



Fuente: Tomado de Torres (2012)

Otro de los modelos establecidos para analizar la retención de los estudiantes en las universidades y que se sido muy apreciado en la literatura es el modelo de integración de estudiantes de Vicent Tinto (1993) que analiza atributos propios del estudiante antes de la entrada a la universidad, metas académicas, experiencias institucionales basadas en el sistema académico y el sistema social, metas institucionales y la integración académica y social del estudiante. En el modelo se analiza el ámbito de relaciones sociales entre los estudiantes, aspectos de los estilos de enseñanza y participación de estudiantes en el aula, la concepción del aprendizaje y su rendimiento académico; además, se realiza una concepción de la forma de integración del estudiante para desarrollar actividades académicas con sus compañeros, debido a que se concibe que la conformación de redes de estudio y la integración personal del alumno puede tener influencia positiva en la retención de los estudiantes.

Figura 2: Modelo de Integración de Vincent Tinto.



Fuente: Tomado de Donoso y Schiefelbein (2007)

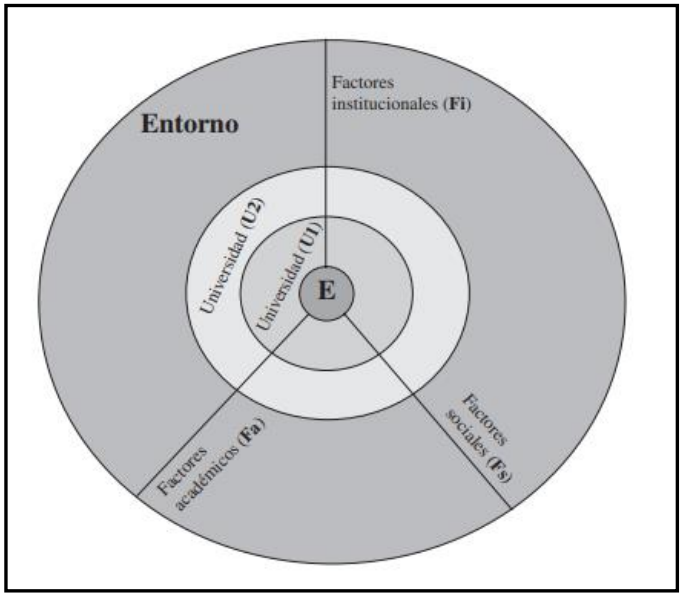
En el mismo sentido, el modelo propuesto por Martin & Arendale (1992) llamado Instrucción Suplementaria, este modelo en lugar de prestar atención con tanta frecuencia a estudiantes de alto riesgo, se centra en contenidos evaluados como de alta complejidad, dominados cursos de alto riego, busca brindar ayuda académica dependiendo del curso. La atención se centra en el proceso y el contenido, el autor considera aspectos importantes necesarios para enfatizar la práctica, hacer más efectiva la asimilación de contenidos y tener el privilegio de fortalecer las habilidades aprendidas, aprovechar el hecho de que los estudiantes se encuentran en la etapa inicial y utilizar los materiales de clase para fortalecer la adquisición de estrategias de aprendizaje y las habilidades de dominar las estrategias.

El Modelo de Amsterdan propuesto por Jong, Sikkema & Dronkers (1997) basado en la teoría de Vicent Tinto y Capital Humano, que afirma que en un contexto dado los individuos entienden las reglas de juegos sociales de cierta manera, de modo que pueden predecir el resultado de acciones alternativas hasta cierto punto. Se incluyen los factores institucionales, al igual que los factores relacionados con el desarrollo personal y factores sociales, ya que reconoce que, si bien los programas educativos son procesos personales, dependen en cierta medida de los elementos del entorno institucional. La motivación es otro factor importante, que puede considerarse como el objetivo general del estudiante, hay dos intenciones: intención de éxito e

intención de éxito dentro de un año. Este es el objetivo comercial del estudiante, pero se reconoce que los antecedentes y la educación del estudiante también afectarán la intención del estudiante de ingresar a la educación superior. Estos métodos se basan en el supuesto de que la educación es una inversión en capital humano y, por lo tanto, el gasto en educación se recuperará posteriormente a través de salarios más altos, es decir, a mayor costo de educación, menor probabilidad de que una persona decida comenzar a aprender, y el aumento del bienestar hace que las personas tengan más probabilidades en su proceso de aprendizaje.

Modelo del equilibrio dinámico centrado en el estudiante de permanencia en la universidad, es un modelo conceptual desarrollado por el autor Díaz Peralta (2008), en donde usa una matriz topológica para agrupar las variables según sus autores obteniendo una relación entre autores y categorías. Como factor principal se menciona la motivación, donde si es positiva aumenta la posibilidad de permanecer en la institución, mientras que, si es negativa aumenta la probabilidad de abandonar la carrera, relacionando así con la integración académica y social. Se considera tres ejes como la situación laboral, compromiso social y metas sociales, técnicas de estudio, calidad de salud, que se adaptan la motivación de permanecer en la universidad, considerando al estudiante como el elemento fundamental sobre el que se rige el desarrollo de la universidad, este modelo permite la dependencia entre los factores para mantener el equilibrio.

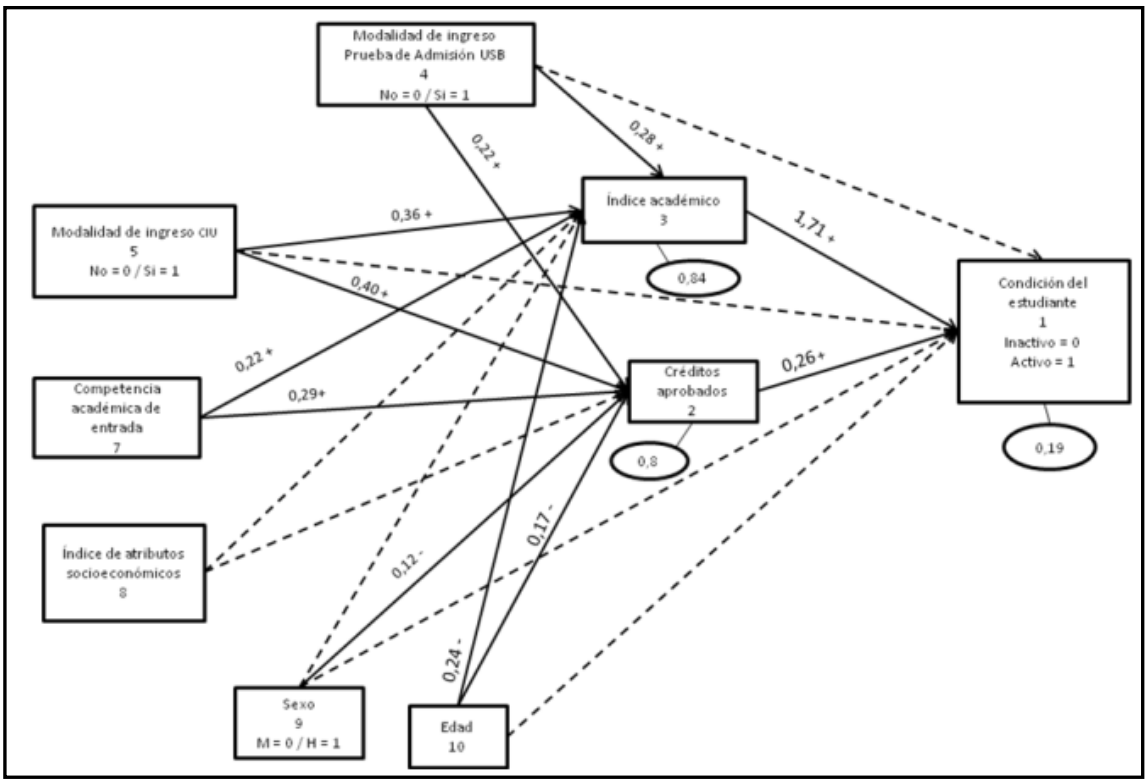
Figura 3: Modelo de Equilibrio Dinámico Centrado en el Estudiante.



Fuente: Tomado de Días (2018)

El modelo de ruta permanente para estudiantes universitarios desarrollado por Fernández de Morgado (2012) se analizan datos socioeconómicos, demográficos, académicos y endógenos y exógenos. El modelo se basa en un modelo de ruta teórico sugerido que consiste en colocar primero el símbolo del coeficiente de la ruta, luego el número de variables endógenas, el número de variables exógenas y finalmente el símbolo que identifica la dirección del cambio. Las variables que pueden predecir el estatus del estudiante incluyen dos variables directas, y cinco indirectas, donde a mayor crédito aprobado y mayor índice, mayor posibilidad de mantenerse activo. La edad es un factor directo, según el autor a menor edad, mayor índice académico y mayor posibilidad de permanecer en la escuela. Por otro lado, los atributos socioeconómicos no afectan directa e indirectamente el estado de los estudiantes.

Figura 4: Modelo de Ruta Permanente.



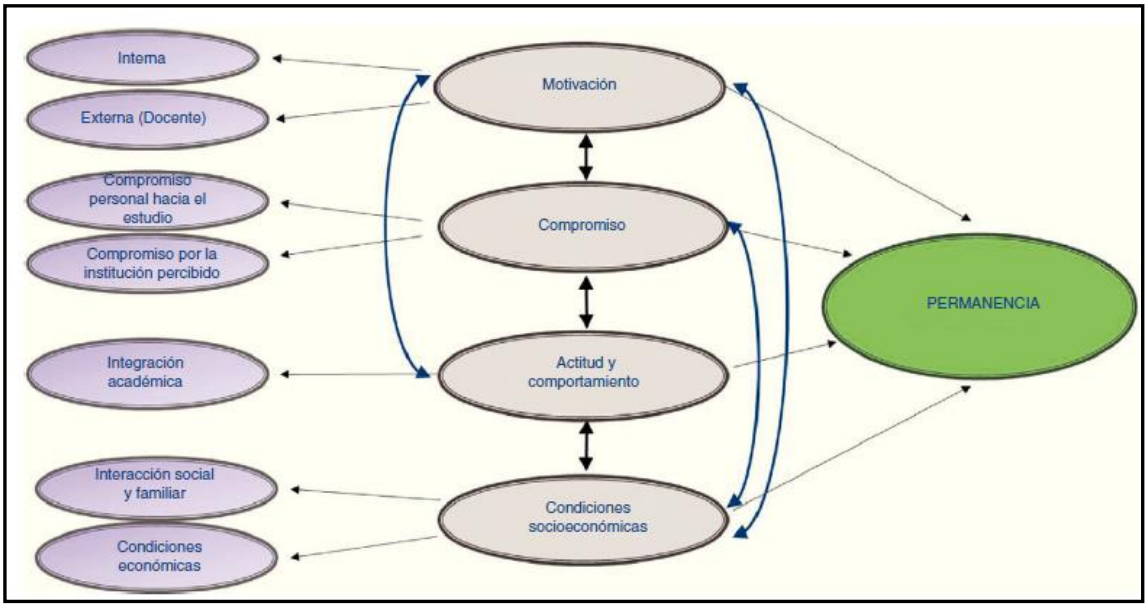
Fuente: Tomado de Fernández de Morgado (2012)

Podemos incluir el Modelo de Apoyo Académico propuesto por Meneses., et.all (2015) en donde se identifica factores académicos que contribuye a la aprobación, rendimiento y retención del estudiante. El sistema emplea antecedentes de perfil identificado a través da la vulnerabilidad académica, el cual permite elaborar un plan de tutorías, talleres de adaptación universitaria, el programa es voluntario o por solicitud del docente o Director de Carrera. Los

resultados obtenidos indican que al iniciar el programa de retención se evidencia avances consecutivos llegando a obtenerse al año 2014 el 89% de retención en colaboración de los programas desarrollados. Además, se identifica la satisfacción de los estudiantes con la participación de las actividades que propone el programa, calificando al mismo como soporte de adaptación a la universidad, así como también, un apoyo para enfrentar los nuevos cambios en la educación universitaria

Velázquez & González (2017), desarrollan un modelo teórico conceptual sobre permanencia estudiantil, determinado en cuatro factores como la motivación, el compromiso, la actitud, el comportamiento y condiciones socioeconómicas, se dividen en tres categorías como la aprobación de materia dentro de un tiempo establecido, asistencia regular y continuidad interrumpida del estudio, Las asociaciones directas entre las variables independientes respecto a la variable dependiente, se presentaron en los factores actitud y comportamiento, compromiso, y condiciones socio económicas; mientras que la variable motivación se encuentra asociada de manera indirecta a la permanecía. Respecto a los factores socioeconómicos, se encuentra la categoría interacción social y familiar con mayor peso.

Figura 5: Modelo Teórico Conceptual sobre Permanencia Estudiantil



Fuente: Tomado de Velázquez Narváez & González Medina (2017)

De igual manera, Rodríguez, Gonzales & Aguilera (2017) desarrollan un modelo predictivo para la permanencia en la Educación Superior, el modelo analiza atributos previos al ingreso a

la universidad, monitoreando así los indicadores para el acceso a la educación terciaria y la optimización de estas acciones. Se analiza datos de 58 estudiantes de las cuales se estudian características como el origen, la residencia, nivel de estudio de sus padres, las notas previas, compromiso del estudiante y el sistema académico. Para desarrollar el modelo se utiliza el programa R-Studio con técnicas de regresión, y red jerárquica para encontrar factores con mayor influencia en la permanecía, se utiliza la variable rendimiento medio con base para la predicción, dentro de las cuales sobresalen las notas de enseñanza, al igual que la asistencia y materias aprobadas después de ingresar a la universidad.

Así mismo, en la investigación realizada por Peña (2017) se estudia la influencia del apoyo que recibe el estudiante de su entorno para fortalecer los procesos de aprendizaje. Los periodos analizados para la investigación comprenden los años 2011 al 2015 con un total de 13304 registros, para analizar estos datos utiliza un análisis descriptivo, regresiones lineales y logísticas. En las regresiones lineales se analiza los datos a partir del promedio acumulado y las veces que estudiante recibió un apoyo por semestre, mientras que el análisis por regresión logística se empleó para determinar la relación entre el apoyo y la calificación. Concluyendo que el apoyo tanto académico como económico influye en el rendimiento académico**.**

El modelo de ecuaciones estructurales en donde se obtienen factores influyentes como los socio-económicos y antecedentes familiares, esta relación contribuye a la retención a través del aporte familiar y de los beneficios económicos que pueden permitir al estudiante permanecer en su formación curricular. Utilizando técnicas de análisis estadístico, el modelo establece las relaciones de las variables a través del uso de regresión. Además el modelo puede incorporar variables de medición, motivación, que permitan obtener resultados favorables en la permanencia estudiantil en la educación superior (Ayala & Atencio, 2018).

De igual manera, el modelo propuesto por Ariño (2018) denominado Modelo Gradual Bidimensional de Integración de Factores, que combina elementos relacionados con la adaptación de los estudiantes a la institución denominado por el autor ajuste personal y ajuste institucional, que hace referencia a la adaptación de las instituciones a los estudiantes y viceversa Este modelo se caracteriza porque se centran en los contenidos con alta complejidad denominado clase de alto riesgo, provee asistencia académica en el primero y segundo año según el curso, este modelo aprovecha que los estudiantes están en fases iniciales de su carrera

para reforzar su adquisición de estrategias de aprendizaje. Los procesos de evaluación y retroalimentación están planteados como estrategias de mejora permanente para garantizar una formación académica adecuada hasta los procesos de graduación, así como también, se considera procesos relevantes de capacitación continua y la intervención de actores académicos en todo el proceso de formación del estudiante.

8.3 Retención como Indicador de la Calidad

Según el Modelo de Evaluación Externa de las Universidades y Escuelas Politécnicas del Ecuador año 2019 propuesto por el Consejo de Aseguramiento de la Calidad de la Educación Superior, dentro de la componente estudiantado del modelo se analiza la permanencia estudiantil como un indicador para la evaluación y acreditación de las Universidades y Escuelas Politécnicas del Ecuador, con el objetivo de promover una adecuada gestión de calidad para el desarrollo universitario. La tasa de permanencia estudiantil está relacionada con la tasa de estudiantes matriculados a la educación terciaria en el proceso de evaluación y que fueron admitidos dos años antes en relación a los estudiantes matriculados que ingresaron a la universidad dos años antes del proceso de evaluación.

Se puede considerar que el organismo de control pretende evaluar las acciones que las instituciones de educación superior han ejecutado para fortalecer la permanencia de los estudiantes con miras a un proceso de titulación en los tiempos oficiales. Entorno a la permanencia se puede evidenciar que existen aspectos de análisis que se relacionan con el objeto de estudio como: procesos de tutorías académicas, tutorías específicas para la titulación, participación de los estudiantes en ayudantías de cátedras y proyectos de investigación y vinculación. Así como también, la participación que tienen los estudiantes en la toma de decisiones respecto al accionar de la universidad.

Como consecuencia del proceso de una adecuada permanencia estudiantil se obtiene como resultado la consecución de las metas académicas de los estudiantes a través de su proceso de titulación y la satisfacción emocional del estudiante. La tasa de graduación tanto de pregrado como de posgrado también es considerada como un indicador de calidad en la gestión académica de una institución de educación superior y está relacionada con cohortes de matrículas y el total de estudiantes de grado en un periodo académico.

En el Modelo Genérico de Evaluación de Aprendizaje de Carreras en el Ecuador (2017) desarrollado por el Consejo de Evaluación y Acreditación y Aseguramiento de la Calidad de la Educación Superior, la tasa de retención es considerada como un indicador de evaluación y se encuentra en el criterio Estudiantes en donde se analiza aspectos relacionados con las garantías que deben brindar las universidades para establecer las condiciones de bienestar de los estudiantes y la eficiencia académica. Según el modelo al menos el 80% de los estudiantes que ingresaron a las universidades en una determinada cohorte deben mantenerse en la institución en un tiempo definido, en este indicador se enlazan aspectos sobre condiciones y características institucionales, y las estrategias implementadas para conservar a los estudiantes y el bienestar universitario.

Según Muñoz & Ramírez (2018) en su investigación sobre articulación de los sistemas de la calidad, Consejo Nacional de Acreditación (CNA) y Normas NTC-ISO 9001 para Programas Académicos de Educación Superior en Instituciones Públicas de Colombia, en su investigación el autor unifica los requerimientos de las normas CNA y NTC generando una matriz de congruencia que representa los elementos concordantes entre ambos, logrando como resultado un modelo de articulación de sistemas de calidad para la orientación de elementos sistémicos para una gestión de procesos educativos. En el modelo propuesto por el autor la permanencia y retención estudiantil se encuentra inserta en el indicador Bienestar Institucional. Los criterios considerados en este modelo hacen referencia a los servicios de bienestar universitario, el apoyo a la formación integral que debe ser suficiente y accesible para los estudiantes a través de políticas integrales que beneficien el desarrollo académico de los estudiantes las mismas que deben estar definidas por la institución. También, son considerados los recursos de apoyo para garantizar un ambiente de trabajo adecuado y un aprendizaje académico de excelencia a través de mecanismos de mejora continua que aporten a la calidad y en la determinación de objetivos institucionales.

En el trabajo presentado por Passarini (2018) que hace referencia a las trayectorias, egreso y acreditación de Carreras en el Mercosur, en el documento propuesto por el autor se analiza las dimensiones y componentes establecidos por Arco Sur para evaluar las carreras, dentro del indicador comunidad universitaria se encuentra el criterio de estudiantes comprendidos para procesos de evaluación y acreditación la misma que relacionada entre otros aspectos la casusas

de desvinculación de los estudiantes, acciones para procesos de retención, apoyo y orientación de los alumnos, implementación de dispositivos de becas, entre otros. Por consiguiente, las estructuras de apoyo a estudiantes son consideradas como fortalezas traducidas en acciones concretas que fortalecen la consolidación de estructuras de apoyo a la enseñanza para cumplir las exigencias de los procesos de acreditación y mejorar los procesos de permanencia estudiantil.

Cáceres (2015) señala que una buena práctica para alcanzar la calidad en la educación superior se relaciona con el trabajo conjunto entre institución y la comunidad universitaria, el seguimiento continuo de los procesos permitirá un desarrollo constante en el contexto de la educación. El aseguramiento de la calidad para Rodríguez & Pedraja (2011) es determinado como un proceso concebido desde la planificación estratégica debido a que es una actividad de gestión determinada en un medio académico competitivo, sin embargo, se debe considerar que este enfoque de calidad también debe estar enlazado con la eficiencia académica y la retención de los estudiantes considerando que éstos son la razón de ser la universidad. Al analizar los modelos de evaluación para Carreras, Universidades y escuelas Politécnicas del Ecuador se puede establecer la necesidad de fortalecer estrategias que permitan alcanzar estos indicadores de manera adecuada.

Por otro lado, Kuh, Kinzie, Schuch & Whitt (2005) señalan que la permanencia estudiantil tiene estrecha relación con el nivel de reto académico de los programas y las instituciones, que implica desafiar al estudiante tanto en lo intelectual como en lo creativo, así como también en el aprendizaje activo y colaborativo, en el que desempeña un papel preponderante el contacto con pares. La interacción de los estudiantes con docentes y otros miembros de su comunidad, en la que se exalta la vinculación de los alumnos con actividades más allá del aula de clase, principalmente en el ámbito investigativo. Las oportunidades educativas enriquecedoras, que estimulen vínculos entre el estudiante y su contexto social, y ambientes de apoyo en el campus, que fomenten tanto la buena realización de actividades académicas como el cultivo de relaciones sociales.

Los servicios de apoyo como las tutorías y el acompañamiento estudiantil que se ofrecen en las instituciones de educación superior tienen el objetivo principal de facilitarle la integración social y académica al estudiante en el contexto universitario. Los servicios de apoyo están dirigidos a

la transición, adaptación, desarrollo y retención de los estudiantes en el contexto universitario (Byron, 2012). La vinculación estructural del estudiante con las funciones sustantivas de la universidad brinda oportunidades a los estudiantes para que participen en el funcionamiento de la universidad, para personalizar las experiencias educativas en términos de estudio y satisfacción del estudiante y sus experiencias de aprendizaje (Schibrowsky & Ackerman, 2007).

8.4 Bases Teóricas

8.4.1 Retención Universitaria.

El termino retención se deriva del latín retention onis, que significa acción y efecto de retener. Para los autores del MEN (2010) describen que la retención es el camino de éxito que el estudiante toma para culminar su formación académica, además el autor Terraza menciona que el rumbo que toma el estudiante se encuentra con dificultades para alcanzar su titulación exitosa, por lo tanto, si se interrumpe la formación académica, no sería trayectoria exitosa de preparación (Terraza-Beleño, 2019).

Según Torres en su investigación teórica sobre la retención, encuentra autores que definen dos perspectivas. La persistencia del estudiante y la retención institucional, donde la persistencia es la capacidad del estudiante para conservar o lograr sus objetivos, y la tasa de retención es la capacidad de la institución en mantener a los estudiantes en el aula hasta la graduación. (Torres, 2010). Mientras que, Pineda & Pedraza (2015) describe que la retención genera escenarios para que las instituciones desarrollen estrategias, que promuevan la persistencia y permanecía de los estudiantes, garantizando la finalización de ciclos y niveles en los tiempos establecidos por la institución. Según Himmel (2002) considera dos aspectos para definir la retención; el tiempo establecido por la institución para obtener el título, o no considerar el tiempo debido a diversos factores como demora, suspensión o cargas académica, por lo cual define como la persistencia del estudiante para obtener un título.

8.5 Factores de Retención Universitaria

En la literatura se puede identificar investigaciones en las que se centra el análisis de la influencia positiva y negativa de factores de retención estudiantil, que hacen referencia a

factores asociados con el sistema administrativo, financiero, académico y la calidad de servicios para los estudiantes en las instituciones de educación terciaria (Aljohani, 2016). Así mismo, los autores Velázquez, Narváez & González (2017) consideran necesario establecer factores de la retención universitaria que afectan directamente al estudiante, de los cuales se destacan la integración académica, compromiso por la institución, las interacciones sociales tanto como las familiares y por último la motivación externa. De la misma manera, Rico, Pedraza & Moreno (2017) menciona que la determinación de factores ligados a la permanencia estudiantil universitaria se relaciona con indicadores de índole socioeconómicos, institucionales, académicos; de los cuales, se destacan la vocación, la motivación personal, el agrado por la carrera seleccionada como los más importantes.

8.6 Clasificación de los Factores de Retención

Por otra parte, Guerra, Rivero, Díaz, & Arciniegas (2019) en su investigación de tendencias de los modelos de información identifican variables como: motivación, compromiso, autoconfianza, integración académica, capital social, servicios estudiantiles, académicos, ayuda financiera, de las cuales, están relacionadas con factores cognitivos, sociales, y organizacionales que permiten destacar la influencia en la retención estudiantil.

8.6.1 Factores Económicos

La literatura menciona el factor económico asociado a las variables socioeconómicas, situación laboral, financiamiento, campo laboral que podrían atender los desafíos financieros, que el estudiante enfrenta en la formación de sus estudios superiores, la presencia de estos factores exigen cambios en las instituciones que permitan promover la permanencia estudiantil universitaria (Mellado, R, Cifuentes, M, & Beltrán, 2017).

8.6.2 Factores Institucionales.

Por otra parte, Torres (2010) menciona que los factores institucionales están ligados a las estrategias, prácticas y cultura de la universidad en donde los más grandes desafíos se presentan en la iniciación de la carrera profesional, y la capacidad que tiene la institución para enfrentar

este problema mediante el apoyo a estudiantes en capacitaciones, servicios académicos que permitan contribuir a la retención y permanencia de los estudiantes de la educación superior.

8.6.3 Factores Personales

También se menciona que los factores personales integran características personales o individuales donde se determinan la vocación, el gusto por la carrera, la motivación, y la influencia que tiene la relación con la familia. Estos factores son determinantes en la decisión de permanecer o no en la carrera. El autor agrupa estos aspectos como factores de protección para llegar a titularse en el tiempo ideal manteniendo así la permanecía estudiantil en las Instituciones de Educación Superior (Parada Rico, 2017).

8.6.4 Factores Académicos

Los factores académicos educativos pueden tener efectos negativos o positivos, sobre la integración del estudiante a la universidad que influyen de manera positiva sobre la retención de estudiantes (Ayala & Atencio, 2018) Así, mismo los factores que se identifican como académicos son: recursos informáticos, carga académica, acceso a las bibliotecas, tiempo para el estudio, estos indicadores permiten un impacto positivo en la permanecía de la educación superior (Parada Rico, 2017).

Tabla 2: Factores de Retención Estudiantil Universitaria basados en la Revisión de la Literatura

Factor	Indicador	Fuente
Económico	Becas	(Velázquez & González, 2017) (Bordón et al., 2015) (Navarro et al., 2019)
	Decisión de trabajar y estudiar	(Verónica, 2020)
	Laboral	(Mellado, René Cifuentes Orellana, Beatriz Beltrán Gabriel, & Jacob, 2017)
	Fondo económico	(Espinosa, Leydi, & Mariño, 2018.)
	Facilitar empleo a padres o estudiantes	(Espinosa, Leydi, & Mariño, 2018.)
Social	Acceso a internet en lugar de residencia	(Meneses, Paulina, Moraga, Ana, Puchi, 2015)

Tabla 2: Factores de Retención Estudiantil Universitaria basados en la Revisión de la Literatura (continuación)

Factor	Indicador	Fuente
Social	Interacción social y familiar	(Velázquez & González, 2017) (García, Pérez, Cavas, & Tomás, 2018)
	Adaptación social	(Esteban, Bernardo, Tuero, Cervero, & Casanova, 2017)
	Convivencia e integración	(Armijo et al., 2019)
	Satisfacción docentes y compañeros	(Verónica, 2020)
	Antecedentes familiares	(Ayala, 2018)
Personal	Gusto por la carrera	(Parada Rico, 2017).
	Motivación	(Parada Rico, 2017) (Navarro, Utreras, & Ugarte, 2019)
	Relación con la familia	(Parada Rico, 2017)
	Gusto por estudiar	(Parada Rico, 2017)
	Vocación	(Parada Rico, 2017) (Navarrete, Candia, & Puchi, 2013) (Said-Hung, 2017)
	Asistencia a clases	(Esteban, Bernardo, Tuero, Cervero, & Casanova, 2017)
	Ranking	(Bordón et al., 2015)
	Perfil estudiante	(Bordón et al., 2015)
	Satisfacción estudiante	(Sánchez, & Castilla, 2017), (Owens, 2018)
	Calidad académica,	(Sánchez, & Castilla, 2017)
	Calidad en infraestructura	(Sánchez, & Castilla, 2017)
	Resiliencia	(Verónica, 2020)
	Personalidad	(Navarro et al., 2019)
	Conocimiento disciplinario básico	(Navarrete et al., 2013)
	Sentido de pertenencia	(Velázquez & González, 2017)
	Naturaleza Psicológica	(Facultad, Ribeiro, & Betti, 2020)
Académico	Rendimiento Académico	(Bordón, Canals, & Rojas, 2015)
	Homologación	(Navarrete et al., 2013)
	Desempeño académico	(Navarrete et al., 2013)
	Motivación por parte del Docente	(Espinosa, Leydi, & Mariño, 2018.)
	Convenios institucionales externas	(Espinosa, Leydi, & Mariño, 2018.) (Espinosa, Hernández, & Mariño, 2020)
	Capacitaciones a estudiantes	(Espinosa, Leydi, & Mariño, 2018.)
	Tutorías	(Velázquez & González, 2017)

Tabla 2: Factores de Retención Estudiantil Universitaria basados en la Revisión de la Literatura (continuación)

Académico	Accesibilidad servicios	(Velázquez & González, 2017)
Académico	Infraestructura	(Velázquez & González, 2017)
	Herramientas académicas	(Velázquez & González, 2017)
	Bibliotecas	(Parada Rico, 2017)
	Carga académica	(Parada Rico, 2017).
	Relación de pares	(Velázquez & González, 2017)
	Primera preferencia a carrera de ingreso	(Meneses, Paulina, Moraga, Ana, Puchi, 2015)
	Actitud positiva académicos	(Rodríguez, Emilio Pedraja, Liliana Araneda, Carmen Plitt, María Rodríguez, 2011)
	Buena comunicación académicos	(Rodríguez, Emilio Pedraja, Liliana Araneda, Carmen Plitt, María Rodríguez Ponce, 2011)
	Rendimiento académico	(Sistemática, Munizaga, & Cifuentes, 2018) (Esteban, Bernardo, Tuero, Cervero, & Casanova, 2017) (Navarrete et al., 2013)
	Titulación oportuna	(Armijo et al., 2019)
	Índice de graduación	(Olarte Moyano, 2020)
	Apoyo de pares	(Navarro et al., 2019)
Sociodemográficos	Edad	(Verónica, 2020) (Olarte Moyano, 2020)

8.7 Técnicas de Machine Learning Aplicadas a la Retención Universitaria

Oñate Bowen a partir de la información socioeconómica del estudiante realizó un análisis descriptivo utilizando K-means para la categorización de los estudiantes de admisión, donde se agrupa en 5 grupos; el grupo 0 y 3 tienen un mejor desempeño en las pruebas, mientras que los grupos 1,2,4 no presentan un buen desempeño en todas las materias. Para otro lado, para el modelo predictivo se utiliza información socioeconómica, resultados de pruebas y expedientes académicos, además, en el proceso predictivo se utiliza dos técnicas de clasificación Árboles de Decisión y Naive Bayes para predecir la perdida de la condición académica por bajo rendimiento, se desarrolla dos modelos con diferentes factores, el primer análisis se realizó en base a la información socioeconómica y los resultados de las pruebas durante el proceso de

admisión. El segundo modelo fue analizado con la información inicial del proceso de inscripción y los registros académicos de las primeras cuatro matriculas. Naive Bayes presentó mejores resultados en la primera y segunda matricula, cuando se le agrego el historial académico mejoro la predicción, el autor describe que, los árboles de decisión clasifica correctamente los registros (Oñate, 2016).

Trstenjak & Donko en su estudio para predecir el éxito de un estudiante utiliza cuatro métodos: ganancia de información, selección secuencial hacia atrás, selección secuencial hacia adelante, para determinar la importancia demográfica del estudiante. También, se utiliza para evaluar el impacto de las características de clasificación dos técnicas de predicción como Naive Bayes y Máquina de Soporte Vectorial. La base de datos está compuesta por información demográfica personal, información demográfica sobre el padre y el tipo de residencia, información sobre los exámenes aprobados, datos sobre el rendimiento de la escuela secundaria, información sobre el estado actual del estudiante, el estado social del estudiante y becas para alimentos. Como resultados finales de predicción, la técnica de Máquina de Soporte Vectorial obtuvo un mejor resultado de precisión (Trstenjak & Donko, 2014).

Cardona y Cudney, en su investigación para predecir la retención de estudiantes utiliza la técnica de Máquina de Soporte Vectorial, que es un algoritmo de aprendizaje supervisado que permite la regresión o clasificación a variables categóricas y numéricas. El modelo evalúa la medida de precisión y recuperación en el conjunto de validación con la precisión general del modelo. Este último asegura un análisis más característico de los resultados al encontrar posibles interpretaciones erróneas, por la cual se utiliza información de 282 estudiantes en la que constan variables como edad, sexo, titulo, y promedio universitario, con esta información su puede predecir la finalización del grado dentro de tres años. La técnica SVM nos permite clasificar las variables de entrada como: clases esperadas, finalización, y no finalización, al utilizar esta técnica el autor concluye que la predicción y clasificación nos permite tener una visión más clara de cómo desarrollar un programa de apoyo para retener a los estudiantes (Cardona & Cudney, 2019).

Wolff, en el desarrollo del modelo para predecir el desempeño de los estudiantes dentro de un solo módulo, menciona que un clasificador depende tanto de los datos como del grado en que se requiera para poder inspeccionar el modelo. Se utiliza la Técnica Máquina de Soporte

Vectorial que están optimizadas para la clasificación binaria, y los árboles de decisión son adecuados para problemas, donde el usuario, quiere comprender qué características han sido más informativas para desarrollar el modelo. Así mismo, se utiliza técnicas como Redes Bayesianas para la predicción de los resultados de los estudiantes y compararlos con otros métodos. Se utiliza datos demográficos, estudios previos y evaluaciones. También, los modelos desarrollados se probaron con datos históricos de tres módulos, utilizando una validación cruzada 10 veces, se demostró que los modelos de Árbol de Decisión son precisos en la predicción, tanto en la caída del rendimiento como el resultado final. El autor concluye que se necesita diferentes modelos dependiente del módulo ya que tienen diferentes formas de evaluación (Wolff et al., 2014).

Tabla 3: Técnicas Utilizadas para la Predicción de la Retención Universitaria

Tema	Técnicas	Autor
Análisis de la Deserción y Permanencia Académica en la Educación Superior Aplicando Minería De Datos.	K-means Árboles de decisión Naive Bayes	(Oñate Bowen, 2016)
Identificación de factores para la retención de estudiantes de instituciones de alto nivel mediante el árbol de decisiones	Árboles De Decisión Naive Bayes	(Gierman et al., 2018)
Predicción temprana del éxito de los estudiantes: minería de datos de inscripción de estudiantes	Árboles De Clasificación y Regresión	(J. Kovacic, 2010)
Predecir el desempeño de los estudiantes a partir de fuentes de datos combinadas	Máquina de Soporte Vectorial Árboles De Decisión Redes Bayesianas	(Wolff et al., 2014)
Predecir la retención de estudiantes utilizando máquinas de vectores de soporte	Máquina de Soporte Vectorial	(Cardona & Cudney, 2019)
Determinar el impacto de las características demográficas en la predicción del éxito de los estudiantes en Croacia	Naive Bayes	(Trstenjak & Donko, 2014)
	Máquina de Soporte Vectorial	
Predecir la finalización de títulos mediante la minería de datos.	Árboles De Decisión	(Cardona et al., 2018).
Métodos de aprendizaje automático para predecir al estudiante	Redes Neuronales	(Babić, 2017)
motivación académica	Árboles De Decisión	

Tabla 3: Técnicas Utilizadas para la Predicción de la Retención Universitaria (continuación)

motivación académica	Máquina de Soporte Vectorial	
Modelado de retención de estudiantes: una evaluación de diferentes métodos y su impacto en los resultados de la predicción	Regresión logística	(Lin et al., 2009)
	Análisis discriminante	
	Modelado de ecuaciones estructurales	
	Redes neuronales	
Aprendizaje automático para predecir la retención de estudiantes	Regresión logística	(Kaiser et al., 2016)
	Máquina de Soporte Vectorial	
	Árboles de decisión	
	Bosques aleatorios	

Fuente: grupo de trabajo

8.8 Minería de Datos

Se define la minería de datos como el proceso de escaneo de enormes repositorios de datos para generar y descubrir conocimiento. La minería de datos puede ser utilizada para extraer y descubrir conocimientos significativos de una gran cantidad de datos, en la actualidad es conocido como una herramienta que se utiliza para analizar datos y métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de base de datos. Además, existen dos tipos de algoritmos minería de datos, descriptivos y predictivos. Donde los modelos predictivos se aplican en el aprendizaje supervisado para calcular valores desconocidos o futuros de las variables dependientes, con base a las variables independientes, y los modelos descriptivos aplican funciones de aprendizaje no supervisados para producir patrones que explique o generalicen la estructura, las relaciones y la interconexión entre los datos (Marulanda, López & Mejía. 2017).

8.9 Clasificación de Minería de Datos

Además, las técnicas de la minería de datos proceden de la estadística y la inteligencia artificial, estas técnicas son algoritmos inteligentes que aplicados a un conjunto de datos se obtienen resultados que determinen el objetivo de aplicación (EcuRed, 2018).

8.9.1 Aprendizaje Supervisado

El algoritmo aprende de las etiquetas de datos. Después de conocer los datos, el algoritmo usará el patrón y asociará a los nuevos datos sin etiquetar (Rouhiainen, 2018).

8.9.1.1 Regresión

El autor Holgado-Apaza, (2018) considera a la regresión logística en la evaluación sobre la influencia que tiene las variables independientes, sobre las variables de respuesta conocida como variable dependiente, que permite controlar el efecto del resto de datos, en el proceso se podrá tomar dos valores únicamente el valor que sea "0" si el hecho no ocurre y "1" si el hecho ocurre. El proceso llamado binominal considera dos posibles resultados que la probabilidad de cada uno de ellos constante una serie de repeticiones.

8.9.1.2 Redes Neuronales.

Se considera a las redes neuronales como método y algoritmos de aprendizaje computacional, que una vez entrenado un conjunto específico de datos se puede obtener nuevos resultados de predicción conocidas como sistemas conexionistas, su estructura está inspirada en redes neuronales biológicas a través de modelos matemáticos. Se aplica una técnica específica al modelo para esperar una respuesta cercana a las expectativas. Debido a la facilidad de la aplicación de las redes neuronales y la integración de herramientas estadísticas en la clasificación de patrones y aproximación de variables, las redes neuronales han sido ampliamente aceptadas en los últimos años (González, Gómez, Pastrana, & Hernández, 2015) (Montaño, 2002), (Villada et al., 2016).

8.9.1.2.1 Perceptrom Multilayer

La red neuronal Perceptrom Multilayer es capaz de clasificar grandes grupos de datos y utiliza el algoritmo backpropagation error, mediante el uso de este algoritmo permite que la red aprenda mediante el entrenamiento de capas conocidas como entrada oculta y de salida para que aprenda el algoritmo, generaliza la información en grupos de entrenamiento y funcionamiento para entregar un valor de salida esperado (González, 2015).

8.9.1.2.2 Voted Perceptrom

Voted Perceptrom es una Red Neuronal con un algoritmo de entrenamiento y predicción, en una función de núcleo del mismo se puede aprovechar la recurrencia, es decir el entrenamiento se genera en vectores contando el número de interacciones, los pesos utilizados son el tiempo de supervivencia del vector, el recuento cae sobre el peso del vector con voto mayoritario, es decir el vector de predicción. La técnica de predicción Voted Perceptrom identifica vectores de tipo w j, que después de cada error se almacena junto a un peso que corresponde al número de decisiones, el vector sobrevive hasta el siguiente error j conocido como vector que acumula un Perceptrom (Univaso et al., 2015)(Gomaa, 2019).

8.9.1.3 Árboles de Decisiones

Por otro parte, el árbol de decisiones es uno de los algoritmos de aprendizaje automático más populares, porque es fácil de visualizar, para que la gente pueda entender lo que está sucediendo. Es como un diagrama de flujo donde cada nivel es una pregunta con una respuesta de sí o no, en donde los nodos internos representan características o atributos, las ramas representan reglas de decisión y cada nodo u hoja representa un resultado. El nodo más alto en el árbol de decisiones se llama nodo raíz (Díaz Martínez, 2007). En la cual se pueden modelar sistemas predictivos que pueden clasificarse según las reglas de decisión, que además es una técnica estadística que puede clasificar variables según los objetivos de la investigación (Jorge et al., 2015).

8.9.1.4 Máquinas de Soporte Vectorial

Smola, (1998) describe que Máquina de Soporte Vectorial (SVM) es un algoritmo supervisado que utiliza un método basado en Kernel para la extracción de características no lineales. El Kernel es una contribución muy valiosa porque nos permite comprender propiedades de operación y generalización de SVM. Además, las máquinas de vectores de soporte combinan varias técnicas, como estadística y redes neuronales. SVM permite utilizar mapas de espacio de características para convertir algoritmos lineales en no lineales (Frutos, 2017).

8.9.1.5 Naive Bayes

Por otro lado, el modelo se basa en una técnica de clasificación estadística llamada "Teorema de Bayes" que determina la probabilidad de una clase en función de predicciones encontradas. También utiliza datos de entrenamiento para evaluar probabilidades condicionadas para predecir nuevos datos. Una vez simplificado se considera la probabilidad del valor de las variables predictores, para asumir que los predictores son independientes. Además, el supuesto de independencia puede reducir significativamente la posible complejidad del cálculo sin afectar severamente el desempeño del modelo (Romero Tutor & Martínez, 2019)

8.9.2 Aprendizaje no Supervisado

Se intenta determinar la estructura de los datos, ya que no se tiene una respuesta conocida para cada situación, por lo que es necesario que el algoritmo encuentre la relación entre las variables involucradas. Un ejemplo de esto se puede visualizar así: dado un conjunto de estudiantes, encuentre una manera de agruparlos de acuerdo con sus características similares (Vance, 2020).

8.9.2.1 Clustering o Agrupación

El Clustering consiste en la agrupación de datos con el fin de identificar semejanzas altas en los objetos, y semejanzas diferentes con los objetos de otros agrupamientos (clúster), es considerada un técnica de machine learning (aprendizaje automatizado) no supervisado (Holgado-Apaza, 2018).

8.9.2.1.1 Algoritmo K-means

K-means es un método diseñado para dividir un conjunto de n observaciones en k grupos. Cada grupo está representado por la puntuación media que lo compone. El representante de cada grupo se llama centroide. El número de grupos a descubrir es un parámetro que debe configurarse. El método de agrupamiento comienza con k centroides ubicados aleatoriamente y asigna cada observación al centroide más cercano. Después de la asignación, mueve el centroide a la posición promedio de todos los datos asignados y luego reasigna los puntos de acuerdo con la nueva posición del centroide (Cestero & Caballero, 2017).

8.9.2.1.2 Reglas de Asociación

El aprendizaje de reglas de asociación es un método de aprendizaje automático basado en reglas que se utiliza para descubrir relaciones interesantes entre variables en grandes bases de datos. Su objetivo es utilizar algunas medidas de interés para identificar reglas que se encuentran en la base de datos (Guil Reyes, 2009).

8.10 Machine Learning

Machine Learning es un campo de la inteligencia artificial, que permite a la computadora aprender a través de datos introducidos. La primera definición apareció en 1995 por Arthur Samuel mediante una partida de damas, que consistía en hacer predicciones atreves de un conjunto de datos (Aa, 2018).

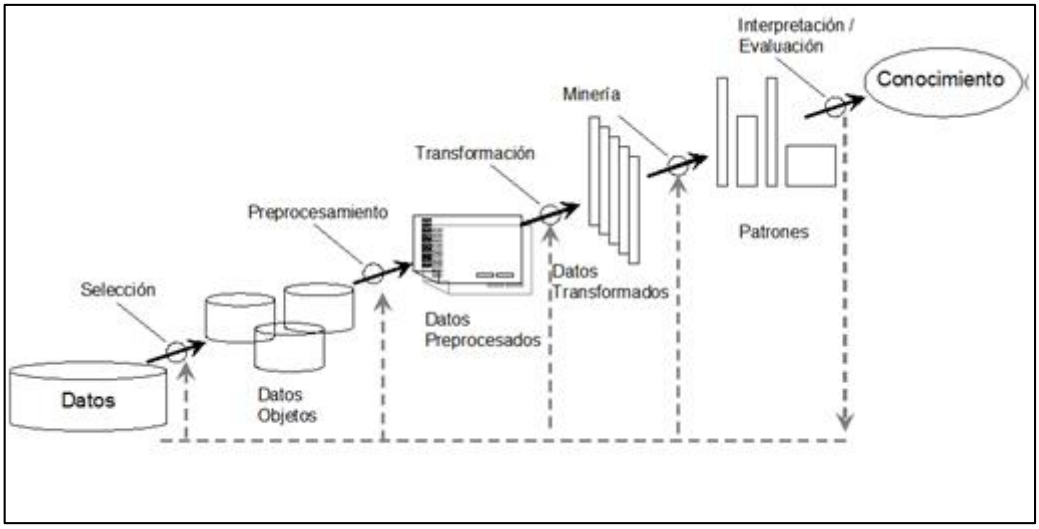
Así mismo es necesario indicar que el aprendizaje o Machine learning es la capacidad que tienen los sistemas informáticos tienen para realizar predicciones y así continuar mejorando su capacidad a partir de datos sin tener que seguir instrucciones. Entonces el aprendizaje automático implica darle a la computadora un conjunto de datos para que desarrolle una predicción, donde los primeros serán incorrectos, mientras más trabajos realiza la computadora el algoritmo de predicción ira mejorando sus predicciones (Macías et al., 2016).

8.11 Metodologías para Minería de Datos

8.11.1 Descubrimiento de Conocimiento en Base de Datos (Knowledge Discovery in Databases – KDD)

La metodología de minería de datos permite el descubrimiento de conocimiento en la base de datos constituye el primer modelo, que define como un "proceso", que consta de diferentes etapas como la preparación de los datos, hasta que se expliquen y difundan los resultados. Los datos extraídos de la base de datos a través del proceso de descubrimiento del conocimientos son los siguientes: efectivos, útiles, novedosos y comprensibles (Moine, 2013).

Figura 6: Proceso de La Metodología KDD



Fuente: Tomado de Kawano (1997)

Braulio & Josep (2015) hace mención de 5 fases para el descubrimiento de conocimiento (KDD) en donde le incluye una fase previa y otra posterior (Braulio & Josep, 2015).

- ✓ **Pre Kdd**

En esta etapa se trata comprender los atributos, limitaciones y reglas del escenario que se está estudiando para lograr los objetivos.

- ✓ **Selección**

En esta etapa se determina la fuente de datos y el tipo de información que se utilizará. En esta etapa se extrae los datos relevantes para el análisis de la base de datos.

- ✓ **Pre-procesamiento y Limpieza de Datos**

Esta etapa incluye la preparación y limpieza de datos extraídos de diferentes fuentes de datos de una manera manejable, lo cual es necesario para etapas posteriores. En esta etapa, se pueden usar varias estrategias para lidiar con datos faltantes o en blanco, datos inconsistentes o datos fuera de rango y finalmente obtener una estructura de datos adecuada para la conversión posterior.

- ✓ **Transformación**

En esta etapa, la calidad de los datos se mejora mediante conversiones que implican la reducción del número de variables en el conjunto de datos o conversiones que convierten valores numéricos en valores categóricos.

- ✓ **Minería de Datos Extracción**

Es la propia etapa de conformación, en la que se aplican métodos inteligentes para extraer patrones previamente desconocidos, efectivos, nuevos, potencialmente útiles y comprensibles que se encuentran ocultos en los datos. El resultado de la fase son los patrones/modelos de minería.

- ✓ **Interpretación y Evaluar Resultados**

En base a algunos resultados de medición se determina los patrones obtenidos y realmente interesantes, se evalúa los resultados obtenidos.

- ✓ **Pos Kdd**

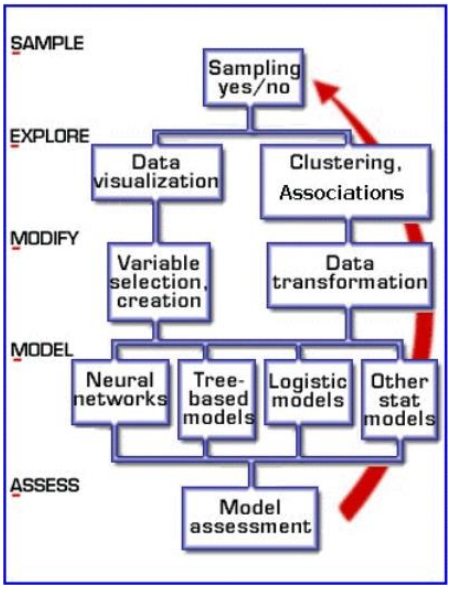
Si todos los pasos se siguen correctamente y los resultados de la evaluación son satisfactorios, entonces el último paso es aplicar el conocimiento descubierto al contexto y comenzar a resolver el problema. Si el resultado no es satisfactorio, es necesario volver a la etapa anterior para realizar algunos ajustes, desde la selección de datos hasta la etapa de evaluación para su análisis (Kawano, 1997).

8.11.2 La metodología Semma

El Instituto SAS define a SEMMA como el proceso de minería de datos que cuenta con el muestreo de datos, exploración, modificación, modelado y evaluación. Además, SEMMA cuenta con un flujo de proceso donde le permite modificar y guardar, así también, la interfaz gráfica de usuario para personas con poca experiencia en al análisis estadístico permitiéndole manejar de manera fluida el proceso de la metodología de minería de datos. Las fases de la

metodología de SEMMA conectan nodos en la herramienta de minería de datos en el diagrama para crear un proceso lógico, para que flujo de información indique las conexiones se muestran las 5 etapas (SAS Institue Inc., 2017).

Figura 7: Procesos de La Metodología Semma



Fuente: Tomado de SAS Institue Inc. (2017)

- ✓ **Sample (muestra)**

En esta etapa, toma muestras del conjunto de datos disponibles. Debe ser lo suficientemente grande para contener la información relevante y lo suficientemente pequeño para ejecutar el proceso con rapidez.

- ✓ **Explore (exploración)**

Implica explorar datos para descubrir tendencias y relaciones desconocidas. También permite conocer los datos y propone nuevas hipótesis a partir de un análisis, con el objetivo de encontrar variables explicativas como entradas del modelo.

- ✓ **Modify (modificación)**

Aquí, nos centramos en la selección y transformación de variables y datos que se utilizarán para construir el modelo.

- ✓ **Model (modelización)**

Aplicando la minería de datos para obtener el modelo, función o combinación de variables seleccionadas como variable para la predicción, lo que nos ayuda a predecir la variable objetivo. En esta etapa consiste en la creación de un modelo a partir de variables explicativas, usando algunas técnicas de pronóstico, como árboles de decisión, redes neuronales, Análisis discriminante o análisis de regresión SAS Institue Inc. (2017).

- ✓ **Assess (evaluación)**

Una vez hecho todo el trabajo, es hora de comparar los modelos. En esta etapa final, se evalúa la utilidad y confiabilidad del modelo obtenido, comparándolo a través de otros modelos con diferentes muestras de datos. (Rodríguez Montequín et al., 2005) (Virseda et al., 2019) (Hernández G & Dueñas R, 2009).

8.11.3 La metodología Crisp-Dm

Esta metodología fue creada en 1997 para el desarrollo de proyectos de minería de datos, por iniciativa de varias empresas privadas como, SPSS (empresa enfocada a software estadístico), Teradata (empresa encargada a Inteligencia de Negocios), Daimler AG (empresa automotriz que contaba con un equipo de Minería de Datos relevante), NCR (una de las mayores empresas en informática en aquella época) y Ohra (compañía aseguradora) (Espinosa Zúñiga, 2020).
La metodología está desarrollada en 6 fases las cuales pueden ser bidireccionales, la cual permite estar en una fase y volver a una anterior para revisar o realizar cambios, cada una de estas fases tiene sub etapas, los que van de una etapa general a casos más específicos (Bautista Agustin & Calderon Nepamuceno, 2019).

- ✓ **Fase de Comprensión del Negocio o Problema.**

Se trata de la compresión de requisitos desde una perspectiva institucional o empresarial para realizar el proyecto de minería de datos, esta fase es muy importante porque permitirá recolectar los datos correctos, caso contrario, las siguientes fases darán resultados erróneos. Para el desarrollo de esta fase se obtiene 4 tareas que son; Determinar los objetivos del negocio, Evaluación de la situación, Determinación de los objetivos de DM, Producción de un Plan del Proyecto.

- ✓ **Fase de Comprensión de la Data Set.**

En esta fase se trata de comprender los datos recolectados para entender y verificar su calidad en realizar relaciones entre ellos para definir las hipótesis. Esta fase está dividida en 4 tares que son; Selección de datos, Descripción de la data, Inspección de la data set, validaciones la data set.

- ✓ **Fase de Estructuración de la Data set**

Es una de las fases que más tiempo consume para su desarrollo, ya que agrupa procesos de selección, limpieza, generación de atributos adicionales, cabios de formatos, para después utilizarlos en la etapa de modelado. Se divide en 5 tareas como; Clasificación de la Data set, Integración de la Data set, Distribución de la data set, incorporación de la data se, Transformación de la data set.

- ✓ **Fase de la Determinación del modelado**

Se selecciona las técnicas de modelado más adecuadas para el proyecto de data mining que se está desarrollando, esta fase comprende los siguientes criterios; la técnica a utilizar tiene que ser adecuada para el problema, se debe disponer de los datos adecuados, la técnica debe cumplir con los requisitos planteados en el problema, tiempo adecuado para el desarrollo del modelo, conocer la técnica a utilizar. Esta etapa se desarrolla con las siguientes tareas; Selección de la técnica de modelado, Generación del plan de prueba, Construcción del modelo y Evaluación del modelo.

- ✓ **Fase de Evaluación**

Se verifica la calidad del modelo a través de análisis con métricas estadísticas, comparando los resultados obtenidos con resultados de otros proyectos realizados, o también se puede analizar los datos con expertos en el tema. En esta fase se determina seguir con el modelo o regresar a una de las fases anteriores para realizar modificaciones o incluso empezar desde cero con un nuevo proyecto. La fase está distribuida en tres tareas que son; Evaluación de los Resultados, Proceso de Revisión, Determinación de Fases Futuras.

- ✓ **Fase de Implementación**

Esta etapa consiste en transformar el conocimiento en acciones dentro del proyecto ya sea atreves de recomendaciones de un analista, para luego tomar las medidas adecuadas basadas en el modelo. El modelo obtenido en el proceso se debe documentar para tener una mejor comprensión por parte de los usuarios y hacer crecer el conocimiento en el área planteado. (Bautista Agustin & Calderon Nepamuceno, 2019), (Espinosa Zúñiga, 2020), (Galán, 2015).

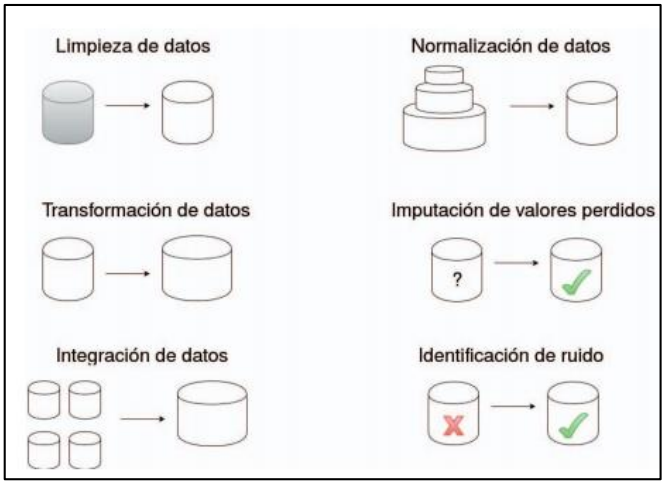
8.12 Técnicas de Pre-procesamiento

La calidad de los datos depende en gran medida en el conocimiento extraído, porque los datos se ven afectados por las inconsistencias, y los datos vacíos, lo que conduce a una baja calidad del conocimiento extraído. El pre procesamiento es la etapa principal del proceso que se considera para encontrar la información de KDD Knowledge Discovery in Databases, es la etapa que se encarga de limpieza de datos, integración, transformación y reducción de datos, para continuar con otros procesos. Además, si se presentan inconsistencias en los datos obtenidos, esto proporcionara una información deficiente en la técnica de pre procesamiento (Garcia, Sergio, Ramírez-Gallego & Herrera, 2016)**.**

Además, el pre procesamiento incluye una variedad de técnicas que distingue dos áreas: En la primera parte consta la preparación de datos consiste en una serie de técnicas que buscan inicializar correctamente los datos utilizados como entrada para los diferentes algoritmos de la minería de datos, se podrían considerar a estas técnicas como obligatorias porque sin el pre procesamiento el algoritmo de extracción no podría ejecutarse y el resultado será incorrecto.

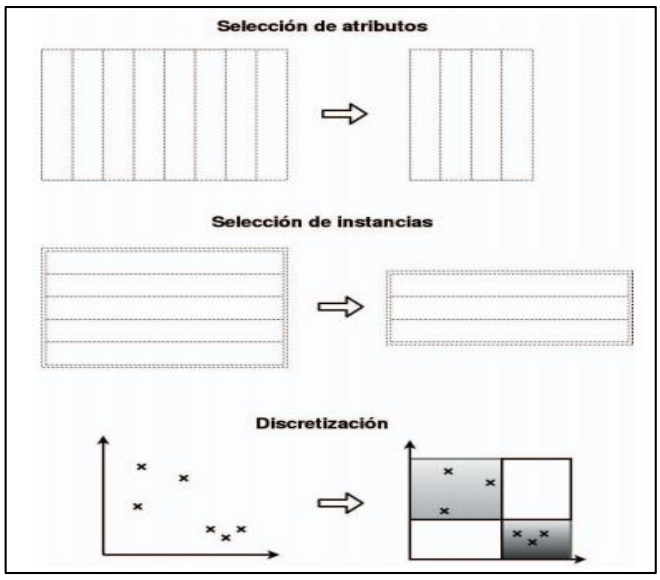
En la segunda parte trata de la reducción de los datos en donde es necesario mantener los datos originales para precautelar la integridad transmitida en los datos. Por tanto es necesario incluir la normalización, limpieza de ruido, e integración (Garcia, Sergio, Ramírez-Gallego & Herrera, 2016).

Figura 8: Desarrollo de pre-procesamiento de los datos



Fuente: Tomado de García, Sergio, Ramírez-gallego, Herrera (2016).

Figura 9: Pre-procesamiento de disminución de datos



Fuente: Tomado de García, Sergio, Ramírez-gallego, Herrera (2016).

8.12.1 Numeric To Nominal

Es un filtro que sirve para realiza transformaciones sobre los datos, este filtro convierte atributos numéricos en nominales. Además, contribuye a reducir recursos de almacenamiento para la predicción y son recomendables después de la transformación de Comma-Separated Values (CVS), para obligar a ciertos atributos que se conviertan en nominales (D. García, 2008) (Lodhi et al., 2016).

8.13 Métricas de Evaluación de Clasificadores

Las métricas utilizadas están basadas en la evaluación de rendimiento del modelo por medio de la precisión Global (Exactitud) precisión positiva (Sensibilidad) y la precisión negativa (Especificidad), datos que detalla la matriz de confusión. Así como también la técnica Área bajo la Curva de (ROC) que indica que si el resultado está cerca de 1 el modelo es considerado como excelente, pero si es mayor a 0.5 es óptimo y en el caso de ser menor a 0.5 es modelo no es apto (Haro et al., 2018). Además, se ejecuta la evaluación por el porcentaje de Error Accuracy el mismo que permite determinar si es confiable el modelo obtenido.(Grijalva Arriaga, 2018).

Precisión Accuracy se refiere a una de las métricas que indica el grado de correlación de los resultados obtenidos en la predicción, se utiliza en determinados problemas donde las variables a pronosticar no son asimétricas.(Romero Maji & Martínez, 2019).
Las métricas de evaluación para el autor (Kuhn & Johnson, 2013) se detallan en la tabla 4:

Tabla 4 Métricas de Evaluación

PRECISIÓN	**SPECIFICITY**	**SENCITIVITY**
Precisión = $\frac{TP}{TP+FP}$	Especificity = $\frac{TP}{TP+FN}$	Sencitivity = $\frac{TN}{TN+FP}$
CONDITION POSITIVE		
TP true positive (verdaderos positivos)	**FP** false positive (falsos positivos)	
CONDITION NEGATIVE		
TN true negative (verdaderos negativos)	**FN** false negative (falsos negativos)	

Fuente: Tomado de Kuhn & Johnson (2013)

- ✓ **Precisión:** se considera como el porcentaje de valores que tiene un clasificador positivo para identificar que es realmente alta debe tener pocos falsos positivos.

- ✓ **Specificity:** es considerado como el número que eventos no clasificados.

- ✓ **Sencitivity:** la sensibilidad se considera como la tasa positiva que mide eventos**,** además predice correctamente el modelo, ya que nos indica la capacidad de evaluación que obtiene los casos positivos correctamente identificados en el proceso de predicción.

- ✓ **TP true positive:** son considerados como predicciones correctas en la sensibilidad, además en base a estos valores se puede determinar que los procesos no son apropiados para dichos eventos.

- ✓ **FP false positive:** son aquellos definidos como menos en la especificidad, indicando un valor fijo de precisión para el modelo consiga remediar entre la especificidad y la sensibilidad, ya que pueden ocurrir perdidas (Kuhn & Johnson, 2013) (Haro et al., 2018).

8.14 Herramientas de Minería de datos

8.14.1 Weka.

La herramienta Weka es un software que permite el diseño de aprendizaje automático y minería de datos, apoya la ejecución de tareas o procesos de minería de datos, tales como regresión, precisión, predicción, visualización, clasificación, agrupación de tareas y permitiendo la visualización. Weka cuenta con múltiples procesos, como algoritmos, que pueden realizar análisis de datos, como modelado predictivo de casos sugeridos, y visualizar los procesos ejecutados en esta herramienta (Virseda et al., 2019).

8.14.2 R Studio.

R es el lenguaje y entorno de alto nivel más potente y profesional para análisis de datos y gráficos. Actualmente, puede realizar diversas tareas estadísticas, desde las más básicas hasta las más avanzadas (R Core Team 2016). R es un entorno de programación compuesto por un conjunto de herramientas muy flexible que se puede ampliar fácilmente a través de paquetes de

software, bibliotecas o definiendo funciones propias. También es gratuito y de código perteneciente al proyecto GNU, como Linux o Mozilla Firefox (Méndez Suárez, 2018).

8.14.3 Spss.

SPSS es un software popular entre los usuarios de Windows para capturar y analizar datos para crear tablas y gráficos con datos complejos. SPSS es conocido por su capacidad para procesar grandes cantidades de datos y puede realizar análisis de texto en otros formatos. SPSS se utiliza para una amplia gama de análisis estadísticos, como estadísticas descriptivas (como media, frecuencia), estadísticas bi-variadas (análisis de varianza, prueba t), regresión, análisis factorial y representación de gráficos de datos (Castañeda et al., 2010).

La vista principal de los datos de SPSS es similar a una hoja de cálculo, que tiene celdas para almacenar datos, que están organizadas por variables (columnas) y casos (filas). Puede ingresar o importar datos manualmente desde hojas de cálculo, archivos de texto u otros formatos. Su apariencia es similar a la hoja de cálculo de excel se diferencia por realizar actividades, a través de comandos en el menú desplegable. El usuario selecciona una prueba estadística y genera un resultado en una nueva ventana (Nell, 2014).

9 HIPÓTESIS

Si se desarrolla un modelo de retención universitaria mediante del uso de técnicas de Machine Learning, entonces se podrá contribuir a la toma oportuna de decisiones por parte de los administrados de las instituciones universitarias.

10 METODOLOGÍAS Y DISEÑO EXPERIMENTAL

10.1 Metodología Científica

La metodología científica es el conjunto de procesos relacionados con las técnicas y procedimientos utilizados para solucionar problemas de investigación por medio de pruebas, verificación e hipótesis. Además, es considerado como un método general que se utiliza en la

investigación científica por medio del método científico para conseguir los objetivos de estudio (Fadias, G, 2012), (Pulido Polo, 2015).

10.2 Tipos de Investigación

10.2.1 Investigación Bibliográfica

La investigación bibliográfica o la investigación de la literatura implican la revisión de materiales bibliográficos existentes relacionados con el tema a estudiar. Este es uno de los pasos principales de cualquier investigación, incluida la selección de fuentes de información. Se considera un paso esencial porque incluye un conjunto de etapas como observación, indagación, interpretación, reflexión y análisis, sentando así las bases necesarias para realizar cualquier investigación (Sampieri et al., 2004).

Para la revisión de la literatura se utiliza la metodología propuesto Bárbara Kitchenham llamada Revisión del estado del arte que se divide en tres etapas.

1. **Planificación de la Revisión. -** En esta etapa se define el problema de investigación a resolver, consta de las siguientes fases. Identificación de la necesidad de una revisión, Especificación de las preguntas de investigación, Desarrollo del protocolo de revisión y Evaluación del protocolo de revisión

2. **Elaboración de la Revisión. -** En esta etapa, Se realiza una revisión literaria donde se busca, selecciona y se filtra de acuerdo a los paramentos establecidos en la etapa anterior. Se divide en sub etapas que son las siguientes: Identificación de la investigación, Selección de estudios primarios, Evaluación de la calidad de los estudios, Extracción de datos y Síntesis de datos.

3. **Redacción de la Revisión. -** En etapa final consiste en la redacción y difusión de los resultados a las partes interesadas en el tema (González Martínez, 2013).

10.2.2 Investigación Tipo Mixta.

La investigación mixta es un estudio en el que los investigadores utilizan múltiples métodos para obtener resultados. En la mayoría de los casos, se trata de un desarrollo de la investigación que combina métodos cuantitativos y cualitativos para obtener resultados más amplios. Debido a su naturaleza, este método de investigación a veces se denomina metodología múltiple. En lugar de realizar dos estudios separados, es más práctico utilizar métodos cuantitativos y cualitativos para crear una sola investigación, para aclarar mejor el objetivo de investigación (Valbuena, 2017).

10.2.3 Investigación Tipo Cuantitativa.

La investigación cuantitativa trata de determinar la fuerza de asociación o relación entre variables, así como la generalización y objetivación de los resultados a través de una muestra. Usando estos métodos, el investigador puede averiguar si su hipótesis está respaldada. Los métodos de investigación cuantitativa incluyen: encuestas, experimentos y entrevistas, esta investigación se caracteriza porque requiere variables numéricas para expresar preguntas de investigación. En otras palabras, los datos analizados deben ser siempre cuantificables, es decir, pueden expresarse en números (Dominguez, 2015).

10.2.4 Investigación Tipo Cualitativa

La investigación Cualitativa es información recopilada en base a observaciones de comportamiento natural, palabras y respuestas públicas para la interpretación posterior del significado, mientras que los métodos cuantitativos proporcionan valores de encuestas, experimentos y entrevistas con respuestas específicas para realizar investigaciones estadísticas y ver cómo se comportan sus variables, sin embargo, el concepto de métodos cualitativos analiza los conjuntos de discursos entre sujetos y las relaciones de significado de los mismos con base en antecedentes culturales, ideológicos y sociológicos. Si la selección se basa en un determinado parámetro, ya no se considerará cualitativa (Strauss & Corbin, 2002).

10.3 Métodos de Investigación

10.3.1 Método Deductivo

El método deductivo permite establecer características particulares del objeto de estudio que pueden ser atributos contenidos o leyes científicas indicadas con anterioridad. De la deducción se obtiene resultados particulares o individuales de consecuencias en base a las conclusiones generalmente aceptadas (Areu, 2014).

10.4 Técnicas de Investigación

10.4.1 Encuesta

La encuesta permite recoger información sea esta verbal o escita por medio de un cuestionario sea que esté compuesta de preguntas abierta o cerradas estas se utilizan como instrumento para recolectar la información adecuada, además la encuesta tiene una característica en particular que se puede aplicar las mismas preguntas con el mismo orden a los encuestados. También la encuesta en la actualidad se puede aplicar a diferentes disciplinas en investigaciones orientadas al conocimiento de un determinado objetivo de estudio (Pardo & Rivera, 2017).

10.5 Instrumento de Investigación

10.5.1 Cuestionario

El cuestionario es una herramienta compuesta por una serie de preguntas, diseñada para generar los datos necesarios para lograr los objetivos de la investigación; es un plan formal diseñado para recolectar información de cada unidad de análisis de estudio, y constituye el centro de la investigación. El cuestionario nos permite estandarizar y unificar las informaciones recolectada. Un diseño inadecuado o inadecuado hará que recopilemos datos incompletos e inexactos, y debemos asumir que generará información no confiable (Fàbregues et al., 2016).

10.5.2 Población

La población es un conjunto de elementos de los cuales tenemos interés de conocer y alcanzar para una toma de decisiones (Graus 2017). La población del caso práctico de investigación se encuentra en la tabla 5 y está constituida por estudiantes de una universidad pública del Ecuador.

Tabla 5 Población Estudiada

Ciclo	Número de estudiantes
1	59
2	44
3	34
4	14
5	40
6	61
7	63
8	45
9	62
10	46
Total	468

10.5.3 Muestra

El cálculo de la muestra es un aspecto importante para determinar los participantes que se deben incluir en la investigación, permitiendo a los investigadores conocer cuántos individuos se necesitan para el estudio. El cálculo de la muestra parte de una función matemática, aplicando fórmulas que se especificaran más adelante (García-García et al., 2013), (Graus, 2017).

10.5.3.1 Muestreo no Probabilístico

El muestreo no probabilístico permite conocer las unidades de estudio en la muestra para obtener un número concreto de la población que va a ser encuestada (De la Cruz, 2019).

10.5.3.2 Muestra con Población Conocida

La muestra con la población conocida se aplica en estudios descriptivos en la cual se incluye una muestra a un determinado grupo de estudio (Villavicencio Caparó, 2018)**.**

Además se presenta la fórmula para calcular el tamaño de la muestra cuando se conoce la población y los valores que intervienen en la fórmula (Castillo et al., 2008), (Aguilar-Barojas, 2005).

- ✓ **N=** tamaño de la población
- ✓ **Z=** nivel de confianza 1.65
- ✓ **P=** probabilidad de éxito, o proporción esperada 0.5
- ✓ **Q=** probabilidad de fracaso 0 .5
- ✓ **D=** precisión (error máximo admisible en términos de proporción) 5% 0.05

Ecuación 1: Función para obtener la muestra

$$n = \frac{N * Z^2 * \mathrm{p} * \mathrm{q}}{d^2 * (N-1) + Z^2 * \mathrm{p} * \mathrm{q}} \quad (1)$$

10.6 Métodos de Desarrollo para la Predicción de la Retención

Para el procero de minería de datos se utilizará la metodología KDD, para determinar la precisión del modelo. La metodología se desarrolla con siguientes etapas: selección, pre-procesamiento, transformación, extracción y por último la interpretación y evaluación de los datos (Timarán, Hernández, Caicedo, Hidalgo, & Alvarado, 2016).

- ✓ **Etapa de Selección**

De acuerdo con la opinión n., Hernández (2016) una vez que se identifica el conocimiento más importante y determinados los procesos de la metodología KDD. Desde la perspectiva de usuario final, se creará una base de datos a partir de una población elegida, en la cual se efectuará los procesos de selección de datos, la misma que mantendrá relación con el objeto de estudio.

- ✓ **Pre-procesamiento y Limpieza**

La etapa data cleaning pre procesamiento es la encargada de analizar la calidad de datos aplicando estrategias para el manejo de datos nulos, y duplicados, utilizando técnicas

estadísticas para su efectividad. También, se puede determinar como una etapa de gran importancia en la interacción con el usuario analista. Para considerar datos ruidosos como errores humanos sea estos por cambios en el sistema o por fuentes mezcladas de datos, que se escapan del rango de valores esperado (Timarán-Pereira, S. R., Hernández-Arteaga, ., et. al., 2016).

- ✓ **Etapa de Transformación**

Desde el punto de vista Kawano (1997) describe a la transformación de los datos como la búsqueda de encontrar características que permitan representar los datos, dependiendo del objetivo de estudio. Para dicho proceso se usa la reducción de dimensionalidad o trans-métodos de formación para reducir el número de variables.

- ✓ **Etapa de Extracción**

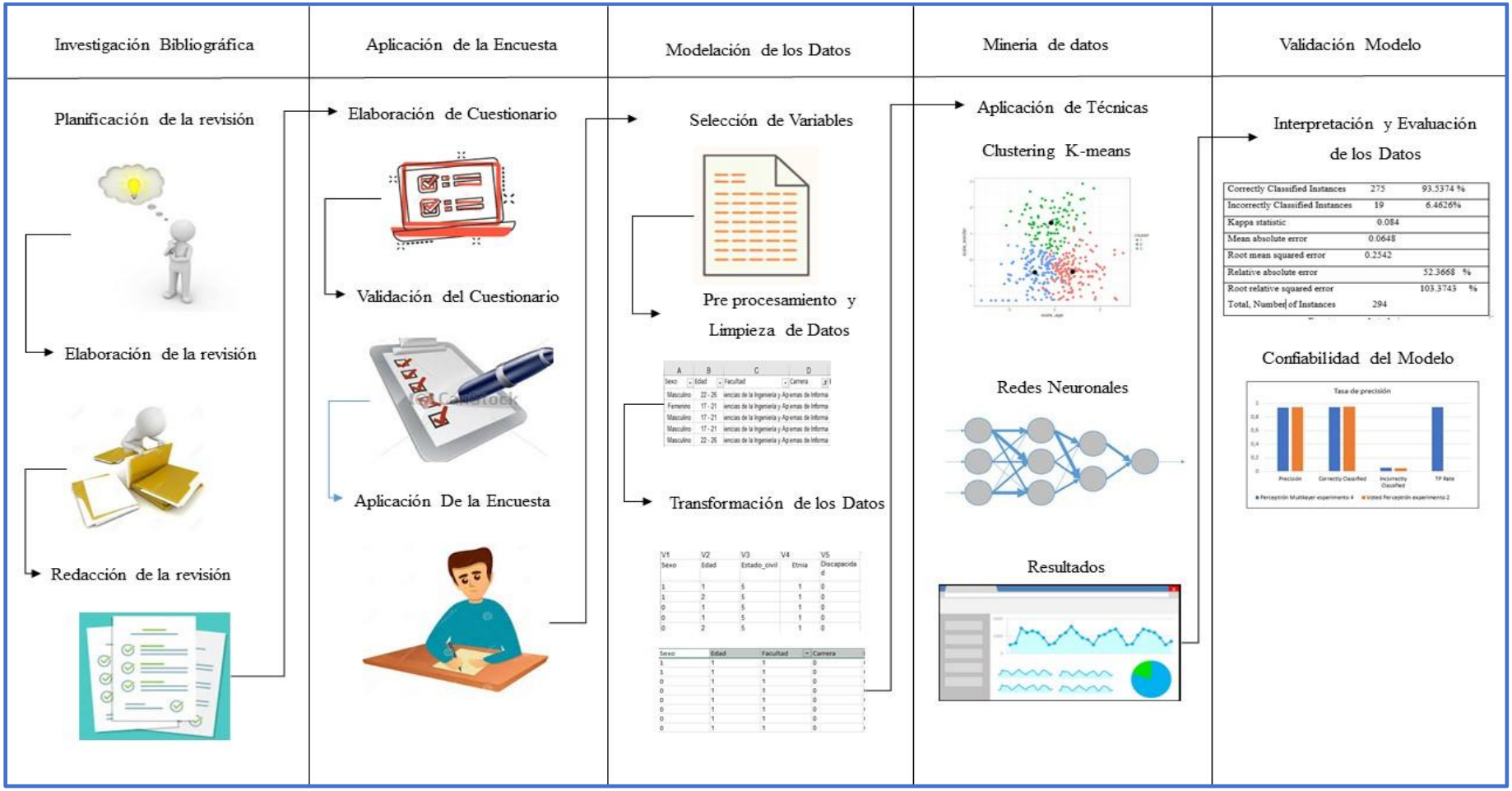
Para el autor Timarán et al., (2016) esta etapa determina la búsqueda de patrones mediante el uso de algoritmo, en la clasificación y regresión. Además, se puede crear modelos predictivos y descriptivos. Los modelos predictivos predicen valores desconocidos de la variable objetivo conocida como variable dependiente, las otras variables conocidas como independientes o predictores, también el modelo puede identificar patrones que resumen los datos que sirven para conocer ciertas características como los clústeres y sus correlaciones.

- ✓ **Etapa de Interpretación Evaluación**

El autor (Timarán et al., 2016) al respecto en esta etapa se interpretan los patrones descubiertos y es posible que retorne a las etapas anteriores para realizar posteriores interacciones. Además, se puede visualizar y eliminar los patrones innecesarios para el usuario. Por último, se potencia el conocimiento extraído en la documentación en los reportes presentados a las partes interesadas para verificar y resolver posibles problemas potenciales, con un conocimiento previamente descubierto. También, (G. García et al., 2014) considera que los patrones encontrados en el proceso se pueden usar para la toma de decisiones, clasificación o predicción, comparándolos con otros trabajos para verificar la validación del modelo. Por lo cual los patrones encontrados deben ser novedosos, comprensibles y útiles.

10.7 Diseño de la Investigación

Figura 10: Desarrollo de la Investigación



En la figura 10 se identifica el proceso de la investigación para determinar los posibles factores que influyen en la retención universitaria, el modelo propuesto determina 5 etapas. En el inicio de las etapas se procedió al análisis en artículos, tesis que tienen una estrecha relación con la retención estudiantil, utilizando gestores científicos de investigación, la segunda etapa consta de la elaboración y aplicación de la encuesta que permitió conocer los problemas que afectan al estudiante en la retención universitaria, esta encuesta se aplicó a estudiante en la carrera de ingeniería de Sistemas de Información de la universidad técnica de Cotopaxi. A partir de esta encuesta se obtuvo una data set, el misma que permitió recoger características representativas del objeto de estudio. En la tercera etapa, se realiza la determinación del modelo teórico para encontrar factores de retención. En la cuarta etapa se emplearon técnicas de minería de datos para la predicción de la retención detallados a continuación: clustering (K-means), Redes Neuronales (Perceptrom Multilayer, Voted Perceptrom). Por último, la quinta etapa, se presenta los resultados en cuanto a los factores de retención estudiantil que corresponden: valoración del esfuerzo del trabajo académico, trabajo en equipo, formación nivelación, satisfacción de la formación recibida, perspectiva inserción laboral, asignaturas y contenidos, apoyo familiar y expectativas futuras con respecto a los estudiantes de la carrera aplicada.

11 ANÁLISIS Y DISCUSIÓN DE RESULTADOS

El desarrollo de este modelo se realizó como caso de estudio en la carrera de Sistemas de Información en la Universidad Técnica de Cotopaxi en la cuidad de Latacunga del periodo académico septiembre 2019 – febrero 2020. El proceso de la investigación inicia con la determinación de factores de éxito de la retención, en donde se analizan datos a partir de la satisfacción del estudiante sobre la educación recibida. Se construye el modelo teórico, mediante regresión lineal que permite obtener las variables más significativas, analizando las variables que son influyentes en el análisis de retención, para validar modelo teórico se diseñaron modelos de predicción y clasificación con las técnicas de minería de datos con Perceptrom Multilayer y Voted Perceptrom, además, nos permite determinar la tasa de probabilidad de influencia que tiene los factores identificados en la retención universitaria.

11.1 Encuesta para determinar factores de retención en las universidades

Se aplicó una encuesta en línea a través de Google Forms, para conocer la apreciación que tienen los estudiantes sobre la permanencia y retención estudiantil, así como también, la satisfacción con la formación académica recibida. La encuesta consta de 46 preguntas entre cerradas y opción múltiple, se aplicaron en el lapso de un mes, desde 6 de enero hasta el 6 de febrero del 2020. La encuesta estuvo orientada hacia los estudiantes que pertenecen a la carrera de Sistemas Información del periodo septiembre 2019- febrero 2020, en el que participaron 294 estudiantes de los ciclos de estudio comprendidos entre primero a décimo. La encuesta tuvo como objetivo descubrir los factores de éxito en la retención estudiantil universitaria. La misma que se divide en dos grupos, primer grupo contiene 15 preguntas con información personal e información relacionada a la institución, el segundo grupo contiene 31 preguntas que contribuyen a determinar la influencia en la retención estudiantil. Los resultados del primer y segundo grupo fueron validados según indicadores propuestos por Matas (2018) por medio de indicadores que contiene valores de Si= 1 y No= 0, por otro lado, se tomó la alternativa en la escala de Licker comprendida (entre 1 al 5), que corresponde a No influye=1, Baja influencia=2, Mediana influencia=3, Alta influencia=4, Influye totalmente=5.

11.2 Analítica Descriptiva de los Datos.

Según el autor Sampedro Carmen (2019) la estadística descriptiva contribuye a la descripción de los datos a través del análisis de la información. Además, se puede seleccionar datos de todas las variables a los temas de interés en la población, y también se puede obtener las características relevantes de las variables a utilizar. El proceso incluye aplicar filtros lógicos y tecnológicos, en los que se eliminan datos del conjunto original correspondiente a 300 instancias, los datos eliminados corresponden a: datos vacíos, celdas en blanco y los datos duplicados. La tabla 6 proporciona los resultados de 294 registros y 46 atributos con sus respectivas características, prestando información confiable y consistente en el conjunto de datos obtenido.

Tabla 6 Data Set

No	Atributo	Descripción
Inr1	Sexo	Género del Estudiante
Inr2	Edad	Edad estudiante
Inr3	Estado-civil	Estado civil estudiante
Inr4	Etnia	Etnia
Inr5	Discapacidad	Posee algún tipo de discapacidad
Inr6	Ubicación_residencia_Universidad	Vive cerca de su universidad (AF)
Inr7	Lugar Origen	Lugar de origen (AF)
Inr8	Tipo Hogar	Tipo de Hogar (AF
Inr9	Num_Miemb_familia	Número de miembros de familia, incluido usted (AF)
Inr10	Tipo_Vivenda	Tipo de vivienda en que reside (AF)
Inr11	Ingresos Familiares	Ingresos familiares mensuales (AF)
Inr12	Recursos estudios	Origen recursos estudios (AF)
Inr13	N_Formacion_Padre	Nivel formación padre, (AF)
Inr14	N_Formacion_Padre	Nivel formación madre, (AF)
Inr15	Tipo Colegio	Tipo de colegio
Inr16	Satisfacción_educación_recibida	Está satisfecho con la educación que se encuentra recibiendo en la universidad, (DH)
Inr17	Perspectiva_inserción_laboral	Siente que su proceso de formación académica está preparándolo correctamente para la carrera después de graduarse, (DH)
Inr18	Formación Nivelación	Completar nivelación le ayudó a sentirse más conectado con su carrera, (EP)
Inr19	Experiencia_bachillerato	Los conocimientos que adquirió en el colegio le ayudaron a prepararse para ingresar a la Universidad Técnica de Cotopaxi, (EP)
Inr20	Rendimiento académico	Su rendimiento académico es satisfactorio, (RA)
Inr 21	Asignaturas_contenidos	Malla curricular diseñada para su carrera es la adecuada, (RA
Inr22	Tiempo_estudio_Fuera_Horario	Está de acuerdo en que se debe dedicar tiempo de estudio fuera del horario de clases, (RA)
Inr 23	Aspiración_obtener_titulo	Dentro de sus aspiraciones está obtener un título profesional, (RA
Inr 24	Valoración_esfuerzo_trabajo_académico	Sus profesores valoran su esfuerzo y su trabajo, (RA
Inr 25	Interación_profesoralumno	Es importante la interacción entre profesor-alumno, (IP
Inr 26	InfluenciaProf_actitudmateria	La figura del profesor influye en su actitud hacia la materia, (IP)
Inr 27	Motivación docente	Sus profesores motivan la autonomía y la responsabilidad de su propio aprendizaje, (IP)
Inr 28	Participación_Actividades_Extrac	La participación en clubes estudiantiles, organizaciones o juegos contribuyen en su experiencia dentro de la Universidad Técnica de Cotopaxi, (AE)
Inr 29	Interacción Compañeros	Le gusta interactuar con sus compañeros, (IP

Tabla 6 Data Set (continuación)

No	Atributo	Descripción
Inr 30	Disfruta_Univesidad	Disfruta de la Universidad Técnica de Cotopaxi, (IP)
Inr 31	Interacion_Actividades_Compañeros	Le gustan las actividades sociales que proponen sus compañeros, (IP
Inr 32	Satisfación_materia_recibida	Está satisfecho con las asignaturas que ofrece su carrera (MCI).
Inr 33	Mecanismos_titulación	La Universidad Técnica de Cotopaxi, ofrece mecanismos adecuados para que usted se mantenga en la Institución y logre titularse (MCIT)
Inr 34	Intercacón_Miembros_Comunidad_Universitaria	Conocer el personal y estudiantes de la Universidad Técnica de Cotopaxi es útil, (MT)
Inr 35	Dialogo-estudiantesprofesores	Le agrada el tipo de conversaciones que tiene con otros estudiantes y profesores, (MS)
Inr 36	Libertad_de_expresion_clases	Es capaz de realizar diferentes preguntas a sus profesores, (EF
Inr 37	Trabajo en equipo	Tiene problemas para realizar trabajos grupales, (EF)
Inr 38	Formación_amistades	Tiene tantos amigos como quisiera
Inr 39	Facilidad comunicación	Siente que puede expresarse fácilmente con sus compañeros
Inr 40	Lazos_amistades_fuera_aula	Fuera de la institución continúa con las amistades que usted tiene dentro de la Universidad Técnica de Cotopaxi, (MT)
Inr 41	Compromiso_formación_académica	El compromiso que tiene el estudiante con su formación académica
Inr 42	Importancia_realizar_trabajos	Considera usted que es importante permanecer cerca de la Universidad realizando trabajos universitarios, (MS)
Inr 43	Influencia familiar	Sus problemas familiares influyen en sus estudios, (MS)
Inr 44	Visión_profesional_futuro	Ir a la Universidad ayuda a mejorar su estilo de vida, (EF)
Inr 45	Valoración_esfuerzo_familiares	Espera obtener algún tipo de reconocimiento por otras personas fuera de la Universidad, (EF)
Inr 46	Apoyo_familiar_Expectativas_futuro	Las perspectivas que el estudiante tiene a futuro son aceptadas por su familia

Fuente: grupo de trabajo

11.3 Confiabilidad de los Datos

La confiabilidad interna del instrumento se muestra en la tabla 7, realizada mediante la prueba del coeficiente de Alfa de Cronbach y el software Spss. Se obtiene un resultado de 0,866 de un

total de 294 instancias, estos valores sugieren alta confiabilidad de la encuesta según (Ibarra, Segredo, & Juárez, 2018)

Tabla 7 Resumen del Procesamiento de Casos

		N	Tos	%
Casos	Válidos	294		100,0
	Excluidos[a]	0		0,0
	Total	294		100,0

Fuente: grupo de trabajo

11.4 Analítica Descriptiva de la Población

En la descriptiva de la población interviene las preguntas con el objetivo de recoger información que permita conocer y comprender el porcentaje de pertenencia de los estudiantes, en sus diferentes categorías (Mendoza, Rubio, & Romero, 2014). Con respecto del análisis descriptivo de la población se encuentra un resumen de las características sociodemográficas consideradas como importantes y detallada en la tabla 8 que se aplicó a los estudiantes matriculados en la carrera de Sistemas de Información en la Universidad Técnica de Cotopaxi.

Los resultados obtenidos se visualizan en la tabla 8 que corresponden a la encuesta aplicada a 294 estudiantes de la Carrera de Sistemas de Información en la Universidad Técnica de Cotopaxi, se puede evidenciar que más de la mitad de los datos se encuentran registrados a la población que corresponde el 60% en el género femenino mientras que la población registrada con menor porcentaje corresponde al 40% al género masculino, las edades con mayor porcentaje se encuentran distribuida en un rango de 22- 26 años de edad en 51%, seguido del rango, 17-21 años de edad que registra un 45% de la población de estudiantes encuestados. Con respecto al estado civil de los estudiantes se distinguen que el 88% son solteros, un 6% casados, mientras que 4% se mantiene en unión libre y con bajos registros en estudiantes divorciados 1% y separados 1%. Por otro lado, los estudiantes que viven cerca de la universidad corresponden 69%, mientras que el 31% menciona que no viven cerca de la universidad, la estructura familiar de los estudiantes en su mayoría corresponde a 4 miembros de familia en un 28%, además se indica que el 1% de los estudiantes viven en casa y departamento propio, mientras que el 16% y 19% en departamentos y casas arrendadas. Se puede señalar con respecto al origen económico de los estudiantes proviene en su gran mayoría de los padres tutores con

71%, y un 21% proviene de recursos propios, el nivel de educación que cuentan el padre corresponde a educación básica en 43%, así como también el nivel de formación de la madre prevalece en educación básica en 41%.

Tabla 8 Analítica Descriptiva de la Población

Variables		**Porcentaje**
Género	Masculino	40%
	Femenino	60%
Edad	17_21	45%
	22- 26	51%
	27-31	3%
	32-36	1%
Estado Civil	Solteros	88%
	Casados	6%
	Unión Libre	4%
	Divorciado	1%
	Separado	1%
Viven cerca de la Universidad	Si	69%
	No	31%
Estructura Familiar	1 miembros familia	1%
	2 miembros familia	4%
	3 miembros familia	15%
	4 miembros familia	28%
	5 miembros familia	23%
	6 miembros familia	16%
	7 miembros familia	7%
	8 miembros familia	4%
	9 miembros familia	2%
	15 miembros familia	1%
Tipo vivienda	Casa arrendada	8%
	Casa propia	75%
	Departamento arrendado	16%
	Departamento propio	1%
Origen Económico	Hermanos	2%
	Otros familiares	2%
	Otros miembros del hogar	1%
	Padres Tutores	71%
	Pareja sentimental	1%
	Recursos propios	21%
	No registra	2%
Educación del padre	Centro alfabetización	1%
	Educación Básica	43%
	Educación Media	33%
	Jardín de Infantes	1%
	Ninguna	1%
	Posgrado Maestría	2%
	Superior no universitario completa	8%
	Superior no universitaria incompleta	2%
	Superior universitaria completa	5%

Tabla 8 Analítica Descriptiva de la Población (continuación)

Variables		Porcentaje
Educación del padre	Superior universitaria incompleta	4%
Educación de la madre	Centro alfabetización	2%
	Educación Básica	41%
	Educación Media	35%
	Jardín de Infantes	2%
	Ninguna	2%
	Posgrado Maestría	1%
	Superior no universitario completa	6%
	Superior no universitaria incompleta	1%
	Superior universitaria completa	8%
	Superior universitaria incompleta	3%

Fuente: grupo de trabajo

11.5 Estadística Descriptiva de los Datos

La estadística descriptiva de los datos se presenta en la tabla 9, permite visualizar individualmente las características de 46 atributos en un resumen estadístico, en el cual 10 atributos son categóricos y 33 numéricos que indica la media, moda y desviación típica de cada uno de los atributos de la Data Set para entregar información concreta sobre los estudiantes encuestados (Esteban et al., 2017).

Tabla 9 Estadística Descriptiva de Datos

	N	Media	Moda	Desv. Típ	Varianza	Mínimo	Máximo
Inr 01	294	0,6	1	0,49	0,24	0	1
Inr 02	294	1,61	2	0,608	0,369	1	5
Inr 03	294	3,3	3	0,465	0,216	2	4
Inr 04	294	3,41	4	1,047	1,096	1	4
Inr 05	294	4,63	5	1,088	1,183	1	6
Inr 06	294	3,95	4	0,457	0,209	1	7
Inr 07	294	6,09	6	1,167	1,361	1	12
Inr 08	294	0,03	0	0,163	0,027	0	1
Inr 09	294	4,12	4	0,651	0,424	1	7
Inr 10	294	4,84	5	0,675	0,456	1	5
Inr 11	294	3,9	4	0,762	0,581	1	7
Inr 12	294	0,31	0	0,463	0,214	0	1
Inr 13	294	1,18	1	0,542	0,294	1	4
Inr 14	294	2,84	3	0,719	0,517	1	5

Tabla 9 Estadística Descriptiva de Datos (continuación)

	N	Media	Moda	Desv. Típ	Varianza	Mínimo	Máximo
Inr 15	294	4,94	4	1,945	3,785	1	16
Inr 16	294	2,1	2	0,52	0,217	1	4
Inr 17	294	1,37	1	0,836	0,699	1	7
Inr 18	294	6,35	6	1,119	1,253	1	9
Inr 19	294	3,6	2	2,411	5,815	1	10
Inr 20	294	3,6	2	2,399	5,756	1	10
Inr 21	294	1,55	1	1,131	1,279	1	5
Inr 22	294	3,21	3	0,718	0,515	1	4
Inr 23	294	3,88	4	0,858	0,736	1	5
Inr 24	294	3,82	4	0,9	0,81	1	5
Inr 25	294	3,39	4	1,264	1,597	1	5
Inr 26	294	3,7	4	1.166	1.359	1	5
Inr 27	294	3,69	4	0,713	0,508	1	5
Inr 28	294	3,66	4	0,934	0,873	1	5
Inr 29	294	3,94	4	0,991	0,982	1	5
Inr 30	294	4,72	5	0,659	0,434	1	5
Inr 31	294	3,55	4	0,972	0,944	1	5
Inr 32	294	4,44	5	0,776	0,602	1	5
Inr 33	294	4,19	5	0,849	0,721	1	5
Inr 34	294	3,8	4	0,916	0,839	1	5
Inr 35	294	3,88	4	1,023	1,047	1	5
Inr 36	294	3,94	4	1,052	1,106	1	5
Inr 37	294	4,14	4	0,925	0,855	1	5
Inr 38	294	3,75	4	1,01	1,02	1	5
Inr 39	294	3,94	4	0,838	0,702	1	5
Inr 40	294	3,96	4	0,885	0,783	1	5
Inr 41	294	4,08	4	0,855	0,731	1	5
Inr 42	294	3,99	4	0,805	0,648	1	5
Inr 43	294	3,7	4	0,981	0,963	1	5
Inr 44	294	3,36	4	1,153	1,33	1	5
Inr 45	294	3,61	4	1,142	1,305	1	5
Inr 46	294	3,78	4	1,035	1,072	1	5
Inr 47	294	4,01	5	1,029	1,058	1	5
Inr 48	294	4,19	5	0,948	0,899	1	5
Inr 49	294	3,91	4	0,939	0,882	1	5
Inr 50	294	3,17	4	1,372	1,882	1	5
Inr 51	294	4,21	5	0,912	0,831	1	5
Inr 52	294	3,87	4	1,122	1,26	1	5
Inr 53	294	3,88	4	1,004	1,009	1	5

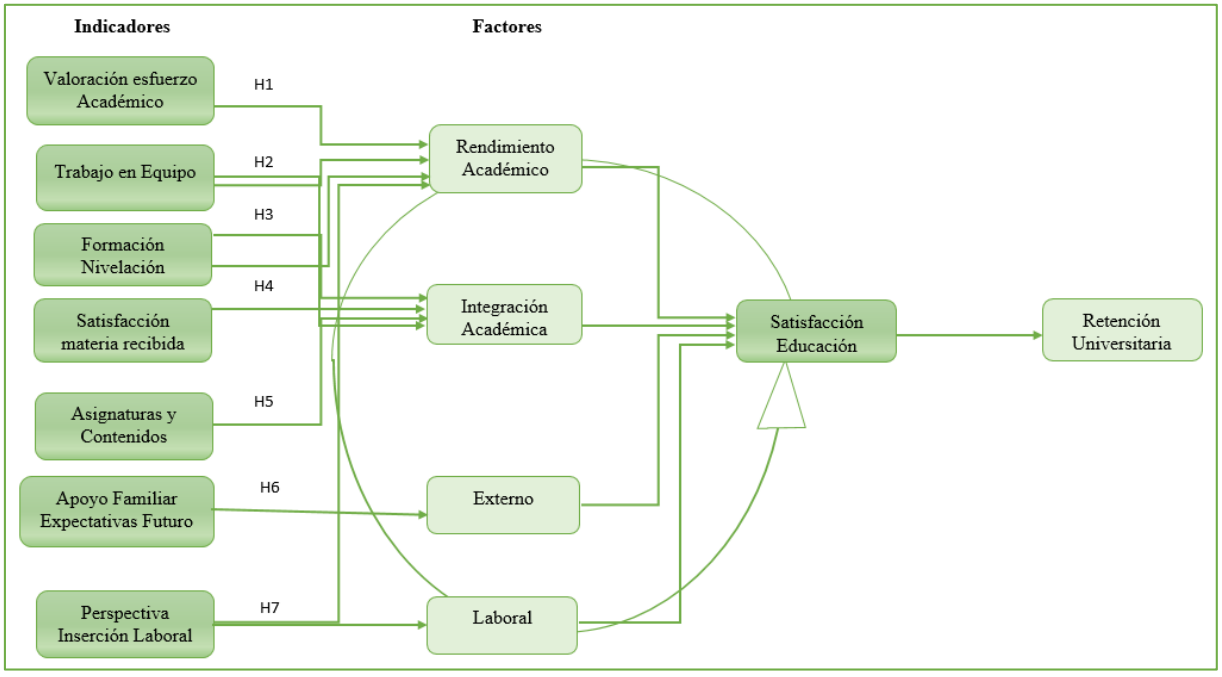
Tabla 9 Analítica Descriptiva de la Población (continuación)

	N	Media	Moda	Desv. Típ	Varianza	Mínimo	Máximo
Inr 54	294	4,41	5	0,892	0,795	1	5

11.6 Modelo Teórico

El modelo teórico de retención estudiantil propuesto, contiene componentes para análisis del comportamiento del sujeto mediante el cálculo de ocurrencia de eventos. La figura 11 presenta factores obtenidos en el modelo en relación al factor satisfacción en la educación recibida por los estudiantes, que se encuentra agrupada en cuatro factores importantes denominados: académico, integración académica, externo, laboral que influyen en la retención estudiantil universitaria.

Figura 11: Modelo de Retención Universitaria



Fuente: Estudiantes de tesis

La selección de variables determina el proceso que se realiza a partir de la correlación que encuentra en las variables entrenadas (Bach et al., 2017). El modelo teórico está basado en la satisfacción académica que tiene el estudiante determinado los siguientes factores que influyen en la retención estudiantil universitaria.

Perspectiva en la inserción laboral Inr 16: se refiere al interés, vocación y habilidades requeridas que demanda la carrera en su área para una mejor proyección laboral, ya que los estudiantes observan alto grado de empleabilidad que prestan mejores condiciones salarial en la carrera de su elección lo que permite la permanencia estudiantil en la universidad (Cassiano, Angela, Cipaguata, Patricia & Reyes, 2016). Además, se puede determinar la perspectiva inserción laboral como la motivación que presta el estudiante en la formación de competencias académicas, identificando el aprendizaje autónomo, responsabilidad que permitan el rendimiento académico en la educación superior (Martinez, 2011).

Formación de nivelación Inr 18: representa un importante proceso que se podría llevar a cabo con tutorías individuales para procesos de actualización de conocimientos, que permitan elaborar estrategias para un mejor rendimiento académico impulsando hábitos de estudio y la calidad de acceso y retención de estudiantes en la Institución de Educación Superior (Pellerano & Matus, 2013). Así mismo, la nivelación de conocimientos permite un mejor rendimiento en determinadas asignaturas a través de programas utilizando técnicas de estudio especializado, que contribuyen a la permanencia estudiantil universitaria (Navarrete, Candia, & Puchi, 2013)

Asignaturas y contenidos Inr 21: está relacionada con el área académica del programa que por medio de la adquisición de conocimientos que contribuye a un mejor desempeño curricular que permite mejorar el rendimiento académico y la permanencia del estudiante en la institución universitaria (Navarrete et al., 2013)

Valoración y esfuerzo del trabajo Inr 24: está ligado al factor rendimiento académico que demuestra las capacidades de estudio, vocación, disposición e interés en estudiar determinada carrera comprometiendo a la institución en ofrecer más y mejores oportunidades con respecto a incentivar el desempeño de los estudiantes. Así como también, la motivación por parte de del docente y su compromiso que tiene con respecto al estudiante en su proceso de formación académico (Roberti, Miranda, & Roberti, 2010).

Satisfacción materia recibida Inr 32: se encuentra ligada al rendimiento académico que permite el éxito de los estudiantes en la que se considera el porcentaje de las asignaturas aprobadas para el avance curricular y la consecución de la carrera, así como también de la permanencia estudiantil en los programas que brinde la institución educación superior

(Navarrete et al., 2013). Por otro lado, podría ser considerada dentro de la relación que existe en la formación académica y la finalización de la preparación profesional del estudiante, además, indica que la satisfacción se encuentra relacionada con la carrera y la vocación que permite mayor satisfacción académica en los estudiantes (Tobon, Durán, & Áñez, 2016).

Trabajo en equipo Inr 37: es considerada como una estrategia de aprendizaje puesto que tiene una apreciación muy favorable tanto en estudiantes como en docentes ya que está orientada a trabajos independientes o grupales que permiten mayor captación de contenidos en cuanto al progreso de las asignaturas y la interacción entre docente y estudiante, considerando elementos muy importantes en la retención estudiantil universitaria (Vélez, 2017).

Apoyo_familiar_Expectativas_futuro Inr 46: la literatura lo encuentra importante, se identifica como un factor externo puesto que el estudiante depende de entre otras cosas de ingresos económicos, que podría convertirse en una opción para tomar o no la educación superior, el apoyo que recibe de su entorno familiar tanto económico como emocional podría influir en su decisión de permanencia en la universidad (Donoso, Donoso, & Arias, 2010). Además, el entorno familiar influye en los valores, actitudes y hábitos que inculcan al trabajo, permitiendo identificar la madurez que se necesita en las decisiones en cuanto a las expectativas hacia un futuro en base a la formación de un ambiente académico que permiten generar las aspiraciones de la obtención de un título académico (Montiel, Osorio, Valcárcel, & Tejedor, 2019).

11.7 Estimación del modelo de Regresión Lineal

La regresión lineal simple es una técnica estadística, que se utiliza para describir la relación entre sus variables, lo cual le permite determinar si existe relación entre la variable dependiente y las variables independientes. Si existe más de dos variables se conoce como regresión múltiple (Támara, 2019).

11.7.1 Mínimos Cuadrados Ordinarios (MCO)

El método de mínimos cuadrados ordinarios (MCO), es el método de evaluación más usado cuando se efectúa un ajuste en un modelo de regresión lineal utilizando parámetros. Permite

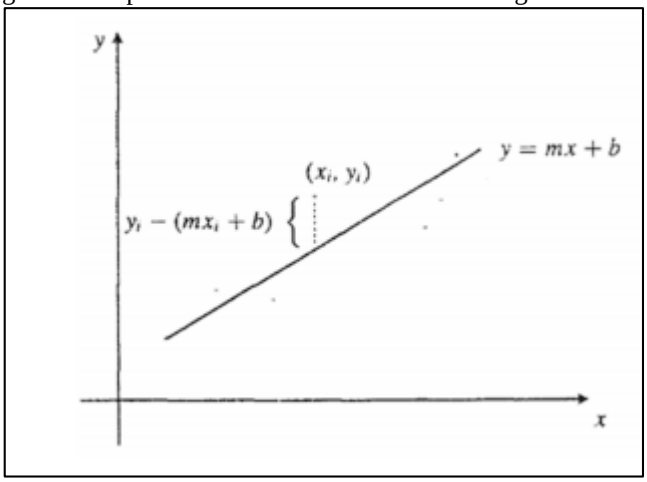
minimizar al cuadrado de algo, y suele propone dos tipos de mínimos cuadrados. El "error" corresponde a lo que se puede cuantificar observando la diferencia entre el valor real y el valor esperado. Esta diferencia se conoce como residuo, y se puede decir que minimiza la suma de cuadrados de los residuos para estimar los parámetros del modelo (Chirivella, 2019).

El método de mínimos cuadrados nos proporciona un criterio por el cual podemos conseguir la mejor recta que representa los puntos dados, en las ecuación de la tabla 10 se presenta la especificación de la fórmula de mínimos cuadrados propuesta por Hurtado (2016).

Tabla 10 Ecuaciones de Mínimos Cuadrados

Especificación	**Ecuación**
Se desearía tener	$yi = \mathrm{m}x_i + \mathrm{b}$
Para todos los puntos (X_i, Y_i) de i=1, …, n, Sin embargo, como en general	$yi \neq \mathrm{m}x_i + \mathrm{b}$
Las desviaciones de las sumas de los cuadrados	$yi - (\mathrm{m}x_i + \mathrm{b})$

Figura 12: Representación de la ecuación de la Regresión Lineal



Fuente Tomado de Hurtado (2016).

11.7.2 Estimador de Mínimos Cuadrados Ecuaciones Lineales

El autor Chirivella (2019) plantea ecuaciones de la regresión lineal indicando que la suma de cuadrados de residuo depende de las estimaciones de parámetros β que permite reducir al mínimo posible, ecuaciones (2), (3), (4).

Ecuación 2 Suma de cuadrados de residuos

$$SCR = \sum_{j=1}^{n} ej2 = \sum_{j=1}^{n} (Y_j + \bar{y}_j) = \sum_{j=1}^{n} (Y_j - (b_o + b_1 X_{1j} + b_2 X_{2j} + \cdots + b_k X_{kjj}))2 \quad (2)$$

Ecuación 3 Estimador de parámetros

$$b=(X'X)^{-1} X'Y \quad (3)$$

Ecuación 4 Estimador de varianza de error

$$\hat{o}^2 = \frac{SCR}{n-k-1} \quad (4)$$

La tabla 11 muestra el modelo original de regresión lineal propuesto mediante el uso de la herramienta Spss en el resultado de una expresión algebraica en la que se ha estimado los coeficientes y ecuaciones con sus respectivos resultados, además los pesos resultantes de cada una de las variables y los valores significativos permiten identificar los principales factores de la retención estudiantil, por otra parte, se identifica con el modelo de regresión lineal que existe correlación entre las variables del conjunto de datos (Robalino,. et.all., 2020), (Hurtado, 2016).

11.7.3 Modelo original

Tabla 11: Modelo de Regresión Lineal Original

Variable dependiente: Satisfacción_educación_recibida							
Variables	**B**	**Error típ.**	**Beta**	**t**	**Sig.**	**Intervalo de confianza de 95,0% para B**	
						Inferior	**Superior**
Inr 1	0,012	0,022	0,024	0,546	0,586	-0,031	0,055
Inr 2	0,35	0,136	0,117	2,579	0,01	0,083	0,618
Inr 3	-0,064	0,036	-0,081	-1,805	0,072	-0,134	0,006
Inr 4	0,046	0,06	0,034	0,76	0,448	-0,073	0,164
Inr 5	0,006	0,044	0,006	0,144	0,885	-0,081	0,094
Inr 6	-0,038	0,023	-0,072	-1,675	0,095	-0,083	0,007
Inr 7	0,05	0,042	0,051	1,204	0,23	-0,032	0,132
Inr 8	0,022	0,026	0,038	0,844	0,399	-0,029	0,073
Inr 9	-0,061	0,062	-0,047	-0,996	0,32	-0,183	0,06
Inr 10	-0,013	0,027	-0,022	-0,488	0,626	-0,065	0,039
Inr 11	0,058	0,063	0,041	0,918	0,359	-0,067	0,183
Inr 12	0,019	0,043	0,02	0,445	0,657	-0,065	0,104

Tabla 11 Modelo de Regresión Lineal Original (continuación)

Inr 13	0,032	0,031	0,054	1,03	0,304	-0,029	0,094
Inr 14	-0,002	0,033	-0,003	-0,065	0,948	-0,066	0,062
Inr 15	0,018	0,05	0,016	0,364	0,716	-0,08	0,117
Inr 17	0,343	0,053	0,367	6,505	0	0,239	0,447
Inr 18	0,082	0,029	0,139	2,862	0,005	0,026	0,138
Inr 19	-0,026	0,031	-0,038	-0,841	0,401	-0,088	0,035
Inr 20	0,121	0,069	0,097	1,762	0,079	-0,014	0,256
Inr 21	0,099	0,045	0,12	2,206	0,028	0,011	0,187
Inr 22	-0,043	0,047	-0,045	-0,913	0,362	-0,136	0,05
Inr 23	-0,442	0,172	-0,232	-2,572	0,011	-0,78	-0,104
Inr 24	0,016	0,042	0,021	0,391	0,696	-0,066	0,099
Inr 25	0,464	0,132	0,267	3,519	0,001	0,204	0,724
Inr 26	-0,007	0,022	-0,014	-0,323	0,747	-0,05	0,036
Inr 27	0,047	0,051	0,053	0,911	0,363	-0,054	0,148
Inr 28	0,038	0,042	0,044	0,897	0,371	-0,045	0,12
Inr 29	-0,14	0,045	-0,175	-3,131	0,002	-0,228	-0,052
Inr 30	-0,084	0,064	-0,084	-1,308	0,192	-0,211	0,042
Inr 31	0,026	0,044	0,031	0,587	0,558	-0,061	0,113
Inr 32	0,259	0,057	0,265	4,546	0	0,147	0,371
Inr 33	0,013	0,068	0,013	0,189	0,85	-0,122	0,148
Inr 34	-0,052	0,076	-0,04	-0,686	0,493	-0,202	0,098
Inr 35	-0,179	0,092	-0,138	-1,951	0,052	-0,36	0,002
Inr 36	0,041	0,041	0,054	0,991	0,323	-0,04	0,121
Inr 37	0,048	0,028	0,079	1,724	0,086	-0,007	0,102
Inr 38	0,012	0,038	0,018	0,327	0,744	-0,063	0,088
Inr 39	0,037	0,044	0,049	0,822	0,412	-0,051	0,124
Inr 40	-0,084	0,048	-0,095	-1,765	0,079	-0,178	0,01
Inr 41	0,052	0,032	0,083	1,639	0,103	-0,011	0,115
Inr 42	-0,048	0,053	-0,047	-0,916	0,36	-0,153	0,056
Inr 43	0,01	0,024	0,02	0,444	0,658	-0,036	0,057
Inr 44	-0,106	0,067	-0,092	-1,577	0,116	-0,238	0,026
Inr 45	-0,009	0,037	-0,011	-0,231	0,818	-0,081	0,064
Inr 46	0,126	0,067	0,109	1,891	0,06	-0,005	0,257

Una vez eliminadas las variables que estadísticamente no son significativas a través del indicador estadístico *p-value* con un nivel de confianza de la técnica del 95% propuesta. La

estimación muestra los resultados obtenidos en relación a la evaluación de significancia con respecto a 5 variables predictores con el valor de coeficiente (β) que identifica el nivel de significancia sea esta positiva o negativa para cada variable del conjunto de datos Carrera (2019).

11.7.4 Modelo Ajustado

Tabla 12: Modelo Ajustado de Retención Estudiantil

Variable dependiente: Satisfacción_educación_recibida							
Variables	**B**	**Error típ.**	**Beta**	**T**	**Sig.**	**Intervalo de confianza de 95,0% para B**	
						Límite inferior	**Límite superior**
Inr 17	0,343	0,053	0,367	6,505	0	0,239	0,447
Inr 18	0,082	0,029	0,139	2,862	0,005	0,026	0,138
Inr 25	0,464	0,132	0,267	3,519	0,001	0,204	0,724
Inr 29	-0,14	0,045	-0,175	-3,131	0,002	-0,228	-0,052
Inr 32	0,259	0,057	0,265	4,546	0	0,147	0,371

11.8 Predicción de la Retención Estudiantil Universitaria

Para realizar la predicción de la retención se utilizó técnicas de aprendizaje supervisado y no supervisado, en el aprendizaje no supervisado se utiliza la técnica Clúster con el algoritmo K-means y en el aprendizaje supervisado se utiliza las técnicas de redes neuronales a través de algoritmo Perceptrom Multilayer y Voted Perceptrom.

11.8.1 Aprendizaje No Supervisado

11.8.1.1 Clúster

Con el fin de identificar estudiantes con características similares, se optó por utilizar el agrupamiento en clústeres mediante K-means, el cual utiliza parámetros de entrada que permiten identificar la correlación que existe en los diferentes grupos de estudiantes. En la tabla 13 se muestra el resultado de la distribución de registros de manera porcentual, la misma que agrupa los datos en tres grupos. El grupo 1 conformado por 88 estudiantes obtuvo un porcentaje 29.93%, así mismo el grupo 2 que tiene un total de 73 estudiantes con un porcentaje de 24.83%,

el grupo 3 estuvo determinado por un total de 133 estudiantes con un porcentaje alto de 45.24% que muestra el mayor grado de correlación de atributos que pueden ser utilizados en procesos posteriores (Oñate, 2016) (Holgado-Apaza, 2018).

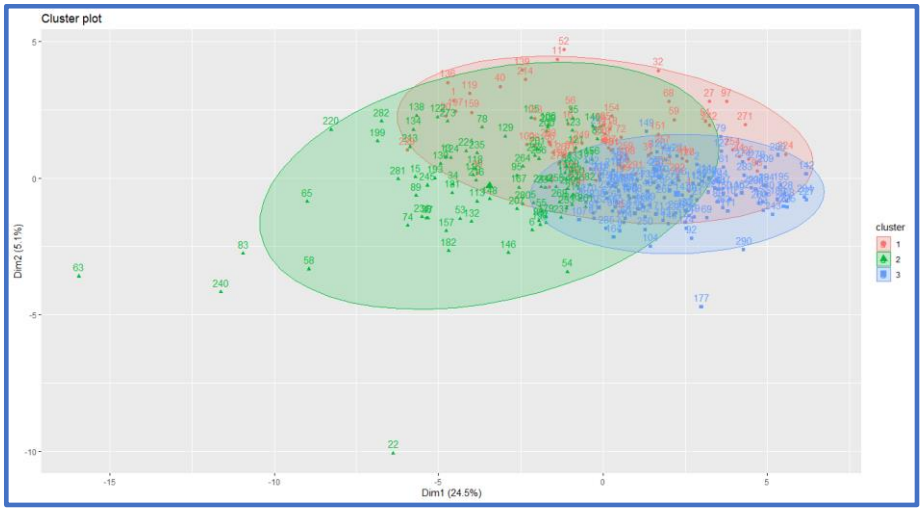
Tabla 13 Resultados de Aplicación de Clúster

# Clúster	***1***	***2***	***3***	
Estudiantes	88	73	133	294
%	29.93%	24.83%	45.24%	100%

Fuente: grupo de trabajo

EL primer clúster conformado por 88 estudiantes equivalentes al 29.93% como se muestra en la Tabla 13, de los encuestados se encuentra indicadores como estructura familiar conformada entre 4 y 5 miembros, en su mayoría vive en casa propia, reciben ayuda de sus padres tutores, sus ingresos familiares comprenden entre 386 y 772 (dólares), terminaron su secundaria en colegios fiscales y fisco misionales, la formación de sus padres es educción básica y media. Se encontraron factores con mayor relevancia como la aspiración de obtener un título, la interacción entre profesor y alumno, la influencia de la actitud de profesor en la materia, interacción entre compañeros en actividades, facilidad de comunicación, lazos de amistad fuera del aula, compromiso con la formación académica, visión profesional a futuro y El apoyo familiar en las expectativas del futuro.

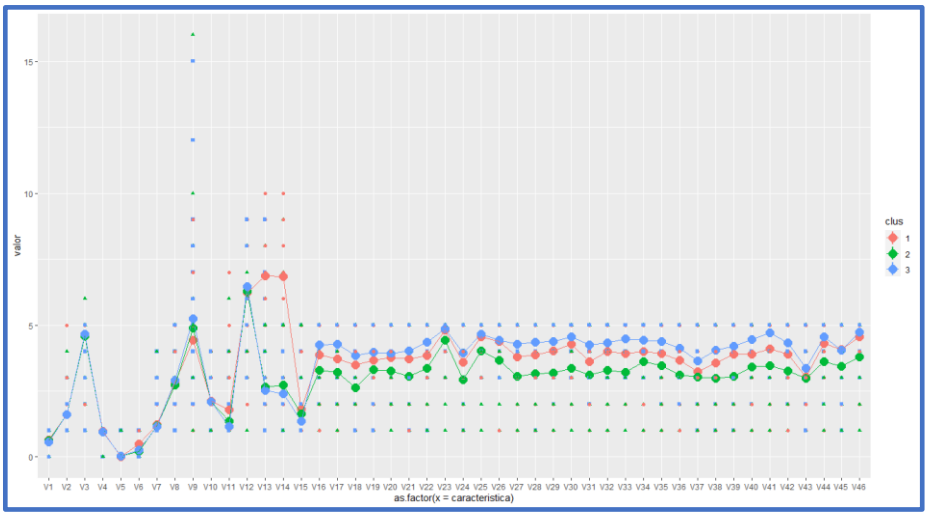
Figura 13: Visualización de datos agrupados con K-Means 3 Clúster



Fuente: R studio

En el clúster 2 como se muestra en la Figura 13 se agrupan 73 estudiantes equivalente a un 24.83% de total, se encuentra indicadores con mayor relevancia como; padres con una formación superior universitaria no completa, aspiración por obtener un título universitario, interacción entre profesor alumno, influencia de la actitud de profesor en la materia, interacciona con la comunidad universitaria, satisfacción con la materia recibida, dialogo entre profesor alumno, los lazos de amistad fuera del aula, compromiso con la formación académica, visión profesional a futuro, y el apoyo familiar en la expectativas a futuro.

Figura 14: Estadística Con Todas Las Características



Fuente: R studio

Por último, el grupo tres conformado por 133 estudiantes equivalente a un 45.24% de la población encuestada, en esta agrupación encontramos características con un alto índice de relevancia como son la aspiración a obtener un título, interacción entre profesor alumno, interacción en actividades con compañeros, compromiso con la formación académica, visión profesional a futuro, apoyo familiar en las expectativas a futuro.

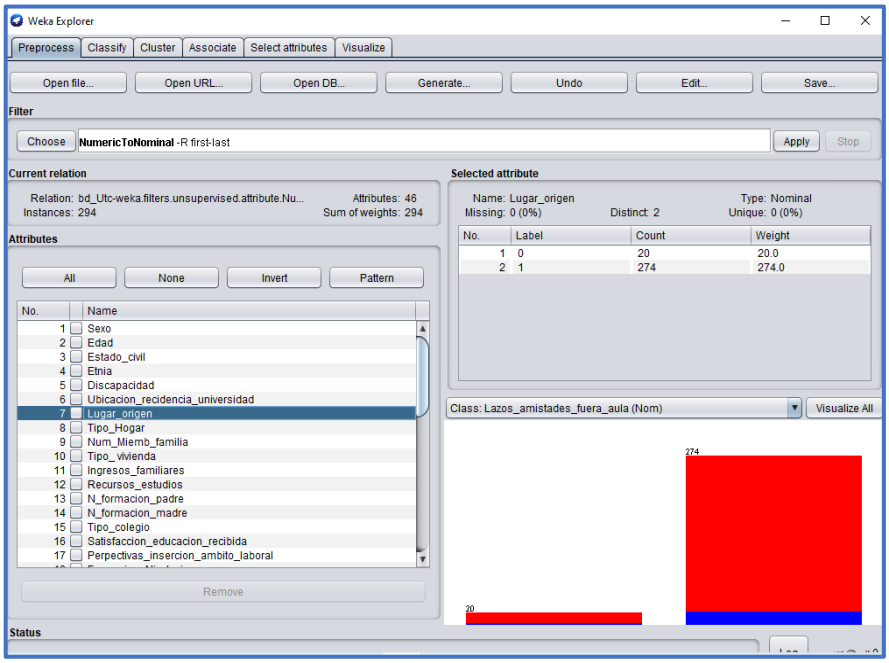
11.8.2 Aprendizaje Supervisado

En la predicción del aprendizaje supervisado se utilizó la técnica del algoritmo de Perceptrom Multilayer y Voted Perceptrom, técnicas de inteligencia artificial utilizadas para determinar la tasa de precisión de las variables identificadas.

11.8.3 Fase del Pre-procesamiento

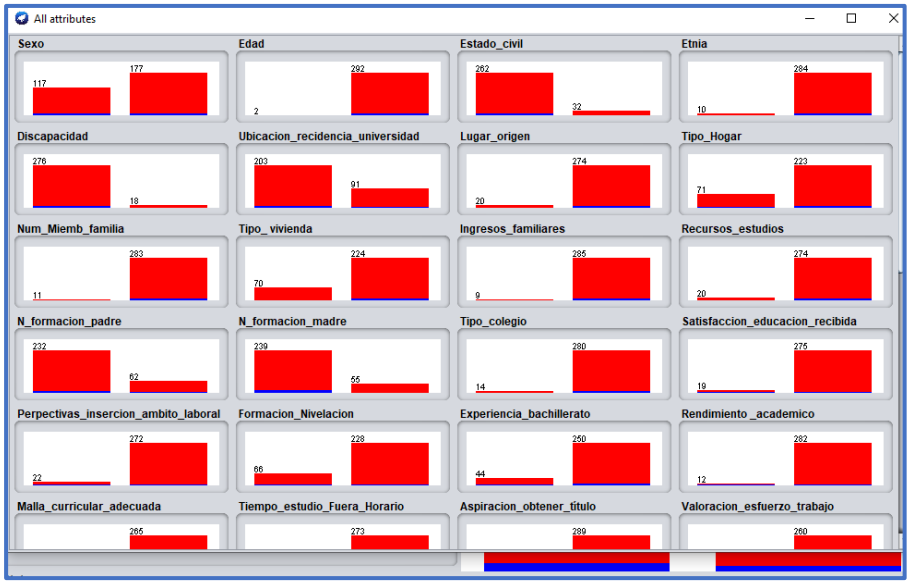
El pre-procesamiento incluye la recopilación de datos que son propensos a varios errores a través de diferentes métodos. Los datos pueden estar ruidosos y con entradas vacías o nulas. Una vez que se obtiene el conjunto de datos correcto, el proceso puede mejorar los resultados de la predicción (Chaudhury et al., 2016). Es importante mencionar que el pre procesamiento de datos es parte importante en los procesos de predicción, la Figura 15 muestra en el sistema 294 instancias y 46 atributos; el sistema muestra una lista de datos estadísticos, indicando la distribución por el tipo de atributo. En tercer lugar, la Figura 16 muestra la distribución de histograma que contiene la información correspondiente para cada atributo, categoría y cada registro por colores.

Figura 15: Pre procesamiento



Fuente: Weka

Figura 16: Histograma de Distribución



11.8.2.1 Selección de Atributos

El método selección de variables se realiza a través de la herramienta Weka, el cual permite seleccionar un subconjunto de atributos de manera automática para determinar las variables más relevantes que buscan identificar el atributo de objeto de la Data Set ingresado, permitiendo la construcción del modelo (Gutiérrez García, 2016). Se aplicó el método de evaluación CfsSubsetEval el cual permite deducir la correlación existe en la clase de cada atributo, los atributos con alta correlación alta son eliminados por ser considerados como atributos redundantes. Así también, tenemos el método de búsqueda Best First que predice los cambios buscando coincidencias en las variables (Bach et al., 2017). En la tabla 14,15 se muestra las variables que permitieron la construcción del modelo.

Tabla 14 Método: BestFirst Atributo CfsSubsetEval

Search Method:	
	Best first.
	Start set: no attributes
	Search direction: forward
	Stale search after 5 node expansions
	Total number of subsets evaluated: 449
	Merit of best subset found: 0.376
Attribute Subset Evaluator	(supervised, Class (nominal): 16 Satisfacción_educación_recibida):
	CFS Subset Evaluator
	Including locally predictive attributes

Fuente: grupo de trabajo

Tabla 14 Método: BestFirst Atributo CfsSubsetEval (continuación)

Selected attributes:	17,18,21,24,32,37,46: 7
	Perpectivas_inserción_ambito_laboral
	Formación_Nivelación
	Asignaturas_contenidos
	Valoración_esfuerzo_trabajo
	Satisfación_materia_recibida
	Trabajos_equipo
	Apoyo_familiar_Expectativas _futuro

Tabla 15 Selección de Atributos

Number of folds (%)	Attribute
10 (100 %)	Inr 17 Perpectivas_inserción_ambito_laboral
6 (60 %)	Inr 18 Formación Nivelación
10 (100 %)	Inr 21 Asignatura_contenidos
7 (70 %)	Inr 24 Valoración_esfuerzo_trabajo
10 (100 %)	Inr 32 Satisfación_materia_recibida
3 (30 %)	Inr 37 Trabajos_equipo
6 (60 %)	Inr 46 Apoyo_familiar_Expectativas _futuro

Fuente: grupo de trabajo

11.8.4 Fase de Extracción del Conocimiento

En la fase de extracción para la construcción del modelo de retención estudiantil universitaria, se realiza en la herramienta Weka mediante la técnica Perceptrom Multilayer (PML), y Voted Perceptrom (VP) que constara de la intervención de 4 experimentos que guardaran relación con la variable dependiente satisfacción educación recibida (Inr 16), además se utiliza la métrica Cross Validation, conocida como validación cruzada, y accuracy que permite identificar el modelo con mayor valor pronosticado, por otro lado, los resultados obtenidos contribuirá a la tomar decisiones para la Carrera de Sistemas de Información.

11.8.4.1 Red Neuronal Perceptrom Multilayer

- **Experimento 1**

El primer experimento se realiza con respecto de la satisfacción educación recibida (Inr 16), para la clasificación se eligió las técnicas Perceptrom Multilayer, con Validación Cruzada que permite deducir los resultados de acuerdo al problema establecido. El modelo toma el total de

variables en fase de aprendizaje, validación y pruebas para probar la red y entregar resultados óptimos (Arlot & Celisse, 2010). El algoritmo se configura con parámetros HiddenLayer a, momentum de 0.2, y LearnigRate de 0.3 para procesar la predicción, el porcentaje de entrenamiento se emplea 66% y un 44% de los datos tomados para el test. Además, la tabla 16 muestra el resultado del clasificador elegido que proporciona la información del resumen del experimento. En el modelo de la red neuronal resultante obtenemos el test de 294 instancias, de las cuales 271 fueron clasificadas correctamente en 92.17% y 23 instancia clasificadas como incorrectas que corresponde a un 7.82%. En la tabla 17 se muestra la confiabilidad en el modelo sobre los verdaderos positivos en 93.2%, con un valor en falso positivos de 0,593 indicando valores predictivos altos para la aplicación del modelo

Tabla 16 Validación Cruzada Estratificada

Stratified cross-validation

Correctly Classified Instances	271	92.1769 %
Incorrectly Classified Instances	23	7.8231 %
Kappa statistic	0.3367	
Mean absolute error	0.0767	
Root mean squared error	0.2617	
Relative absolute error		62.0052 %
Root relative squared error		106.3948 %
Total, Number of Instances	294	

Fuente: grupo de trabajo

Tabla 17 Precisión Detallada por Clase Perceptrom Multilayer Experimento 1

Precisión detallada por clase capa a

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Área	PRC Área	Class
0,368	0,040	0,389	0,368	0,378	0,337	0,803	0,343	0
0,960	0,632	0,957	0,960	0,958	0,337	0,803	0,978	1
0,932	0,593	0,920	0,922	0,921	0,337	0,803	0,937	

Fuente: grupo de trabajo

- **Experimento 2**

El experimento 2 consiste en seguir con el proceso de clasificación con la técnica Perceptrom Multilayer considerado por su rapidez de aplicación, se utiliza el algoritmo de Backpropagation en la experimentación en las redes neuronales, ya que sus parámetros pueden ser configurados para obtener un óptimo resultado tanto en tasas de aprendizaje, numero de épocas, mínimo,

velocidad del aprendizaje (Garzón & Landin, 2013), (Gamarra et al., 2018). El proceso se realiza con la configuramos en la función del algoritmo sobre la capa HiddenLayers (i), momentum de 0.03, y la velocidad del aprendizaje (learningRate) de 0.3, para obtener la predicción del entrenamiento. La tabla 18 muestra la información de la red neuronal de los experimentos realizados en cuanto a la correcta clasificación de sus instancias tiene un 91.83% y la incorrecta clasificación representa 24 instancias con un valor en 8.16%. Además, la tabla 19 indica la precisión en verdaderos positivos con 0,918 y con falso positivos de 0.643 indicando una confiabilidad de 91% de precisión sobre el modelo elaborado mostrando una capacidad de clasificación calificada como buena.

Tabla 18: Validación Cruzada Estratificada

Stratified cross-validation

Correctly Classified Instances	270	91.8367 %
Incorrectly Classified Instances	24	8.1633 %
Kappa statistic	0,29	
Mean absolute error	0.0748	
Root mean squared error	0.2499	
Relative absolute error		62.4522 %
Root relative squared error		101.6237%
Total, Number of Instances	294	

Fuente: grupo de trabajo

Tabla 19: Precisión de Clase Perceptrom Multilayer Experimento2

Precisión detallada por clase capa i

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Área	PRC Área	Class
0,316	0,040	0,353	0,316	0,333	0,291	0,888	0,4	0
0,960	0,684	0,953	0,960	0,957	0,291	0,888	0,989	1
0,918	0,643	0,914	0,918	0,916	0,291	0,888	0,951	

Fuente: grupo de trabajo

- **Experimento 3**

De acuerdo al orden del entrenamiento se realizó el experimento 3 que incluye la clasificación en la técnica Perceptrom Multilayer que contribuye a la predicción como a la comparación de modelos para comprender la precisión y la tasa de error que puede proporcionar el algoritmo (Widyahastuti & Tjhin, 2017). Con respecto de la satisfacción educación recibida (Inr 16). Se utiliza la validación cruzada, que muestra de 294 instancias que tiene la data set, de las cuales

276 fueron clasificadas correctamente y 18 instancias incorrectas, también se realiza la configuración en la función del algoritmo en capa HiddenLayers o, momentum de 0.4, y la velocidad del aprendizaje (learningRate) de 0.003, para obtener la predicción del entrenamiento. En la tabla 20 se muestra la información de la red neuronal del experimento realizado. Además, el modelo de red neuronal, indica valores de 0.939 en verdaderos positivos, mientras que en falsos positivos muestra 0,739 dando una confiabilidad de 93.8% en el Modelo.

Tabla 20: Validación Cruzada Estratificada

Stratified cross-validation

Correctly Classified Instances	276	93.8776 %
Incorrectly Classified Instances	18	6.1224 %
Kappa statistic	0,2827	
Mean absolute error	0.1071	
Root mean squared error	0.211	
Relative absolute error		86.5609 %
Root relative squared error		85.7957 %
Total, Number of Instances	294	

Fuente: grupo de trabajo

Tabla 21: Precisión por Clase Perceptrom Multilayer Experimento 3

Precisión detallada por clase capa o

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Área	PRC Área	Class
0,211	0,011	0,571	0,211	0,308	0,322	0,291	0,455	0
0,989	0,789	0,948	0,989	0,968	0,322	0,291	0,994	1
0,939	0,739	0,923	0,939	0,925	0,322	0,921	0,959	

Fuente: grupo de trabajo

- **Experimento 4**

Por último, tenemos el experimento 4 que consiste en la aplicación del Perceptrom Multilayer que se realiza con la aplicación de Cross-validation por 10 veces, los investigadores indican que si la fiabilidad de los resultados del Perceptrom muestran un valor superior al 0,7 el modelo predictivo puede ser considerado como confiable (Malhotra, 2016), (Owens, 2013). La tabla 22 presenta los resultaos del proceso de la validación del modelo propuesto. También, permite identificar 277 instancias que fueron clasificadas correctamente en 94.21% y para la clasificación incorrecta se muestran 18 instancias que corresponden a 5.78%. Además, se configura los cambios para la experimentación que se efectúan sobre capas HiddenLayers (t),

momentum de 0.5, y la velocidad del aprendizaje (learningRate) de 0.0003, para obtener la predicción del entrenamiento. En la tabla 22 se puede visualizar el desempeño del modelo tanto en el entrenamiento como la validación, indicando que el resultado aumenta en relación a de los demás experimentos obteniendo una confiabilidad alta en el modelo de 94.2% según (Garzón & Landin, 2013).

Tabla 22: Validación Cruzada Estratificada

Stratified cross-validation

Correctly Classified Instances	277	94.2177 %
Incorrectly Classified Instances	17	5.7823 %
Kappa statistic	0.455	
Mean absolute error	0.0896	
Root mean squared error	0.2157	
Relative absolute error		72.4245 %
Root relative squared error		87.692 %
Total, Number of Instances	294	

Fuente: grupo de trabajo

Tabla 23: Precisión por Clase Perceptrom Multilayer Experimento 4

Precisión detallada por clase capa t

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Área	PRC Área	Class
0,421	0,022	0,571	0,421	0,485	0,461	0,911	0,439	0
0,978	0,579	0,961	0,978	0,969	0,461	0,911	0,993	1
0,942	0,543	0,936	0,942	0,938	0,461	0,911	0,957	

Fuente: grupo de trabajo

11.8.4.2 Red Neuronal Voted Perceptrom

El segundo proceso de predicción incluirá la intervención del algoritmo Voted Perceptrom. Al igual que el proceso anterior, consta de la intervención de 4 experimentos que estará relacionado con la satisfacción educación recibida (Inr 16), se utiliza también para los procedimientos experimentales mediante validación cruzada por 10 veces, estos procesos se realizarán para determinar una mayor precisión del modelo de predicción, por otro lado, los resultados obtenidos contribuirán a la tomar decisiones en la comunidad universitaria.

- **Experimento 1**

Realizamos el primer experimento con respecto de la satisfacción en la educación recibida (Inr 16), para la clasificación elegimos la técnica Voted Perceptrom, puesto que el algoritmo del perceptrom eventualmente convergerá una hipótesis correcta, que predice utilizando el vector de predicción que mayor voto obtenido. También, efectúa cambios que contribuye a encontrar una regla de predicción consistente para el resultado esperado (Nason et al., 1996). La configuración de algoritmo se realiza sobre los parámetros exponente 1.0, y seed en 1, una vez seleccionada la configuración, se realiza el entrenamiento del modelo. Cabe señalar, que en la tabla 24 existe una correcta clasificación de 275 instancias en 93.53% y una incorrecta clasificación de 19 instancias que cuentan con el 6.46 %. Así mismo, en el resumen de precisión en la tabla 24 y 25 se identifica valores en verdaderos positivos en 0,935 y en falsos positivos en 0,886 dando una confiabilidad de predicción en 93.5% en el modelo.

Tabla 24: Validación Cruzada Estratificada

Stratified cross-validation

Correctly Classified Instances	275	93.5374 %
Incorrectly Classified Instances	19	6.4626%
Kappa statistic	0.084	
Mean absolute error	0.0648	
Root mean squared error	0.2542	
Relative absolute error		52.3668 %
Root relative squared error		103.3743 %
Total, Number of Instances	294	

Fuente: grupo de trabajo

Tabla 25: Precisión de Clase Experimento 1 Voted Perceptrom

Precisión detallada por clase 1

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Área	PRC Área	Class
0,053	0,004	0,500	0,053	0,095	0,147	0,547	0,110	0
0,996	0,947	0,938	0,996	0,966	0,147	0,547	0,941	1
0,935	0,886	0,91	0,935	0,910	0,147	0,547	0,887	

Fuente: grupo de trabajo

- **Experimento 2**

El experimento dos consiste en la secuencia de la técnica Voted Perceptrom a través del entrenamiento y las pruebas se pueden superar los procesos, manipulando el algoritmo para conseguir el resultado esperado. Además la técnica también determina el número de interacciones durante el entrenamiento y el número de exponente con números aleatorios para generar el entrenamiento el que mayor peso obtenga será determinado en la predicción del modelo (Tolegen et al., 2020). En la experimentación se configura la función en el exponente 2.0, y seed 1. Así mismo en tabla 26 se identifica que del total de las instancias de la data set, 279 fueron clasificadas correctamente y 15 instancias incorrectas. Además, en la tabla 27 se muestra la información de la red neuronal obtenida del experimento dos. Como resultado se muestran un resumen porcentual con valores en verdaderos positivo 0,949% y los falsos positivos a 0.689%. Determinando que la predicción proporciona una confiabilidad de 94.8%.

Tabla 26 Validación Cruzada Estratificada

Stratified cross-validation

Correctly Classified Instances	279	94.898 %
Incorrectly Classified Instances	15	5.102 %
Kappa statistic	0.3808	
Mean absolute error	0.051	
Root mean squared error	0.2259	
Relative absolute error		41.2248 %
Root relative squared error		91.8449 %
Total Number of Instances	294	

Fuente: grupo de trabajo

Tabla 27 Precisión por Clase Voted Perceptrom Experimento 2

Precisión detallada por clase 2

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Área	PRC Área	Class
0,263	0,004	0,833	0,263	0,400	0,451	0,630	0,252	0
0,996	0,737	0,951	0,996	0,973	0,451	0,629	0,951	1
0,949	0,689	0,944	0,949	0,936	0,451	0,629	0,906	

Fuente: grupo de trabajo

- **Experimento 3**

El experimento 3 consiste en la secuencia del experimento Voted Perceptrom y se utilizara la validación cruzada 10 veces, permite distribuir los datos en 10 particiones hasta que concluya el proceso, la característica de esta técnica es que mientras entrena se almacena la información para entregar mejores predicciones, además se determina que la técnica es comparable con otras técnicas de estudio en términos de exactitud tiempo velocidad y conocer su utilidad (Sassano, 2008). La configuración del algoritmo se realizó en el exponente 3.0 y sed semilla de 1. El entrenamiento proporciona un modelo de red neuronal. El conjunto de datos muestra en la tabla 28 de un total 278 fueron clasificadas correctamente y 16 instancias incorrecta. Además, la tabla 29 identifica el resultado del modelo en un resumen porcentual con una alta precisión en el modelo de la red neuronal 94.6%.

Tabla 28: Validación Cruzada Estratificada

Stratified cross-validation

Correctly Classified Instances	278	94.5578 %
Incorrectly Classified Instances	16	5.4422 %
Kappa statistic	0.3624	
Mean absolute error	0.0543	
Root mean squared error	0.2326	
Relative absolute error		43.8452 %
Root relative squared error		94.5821 %
Total, Number of Instances	294	

Fuente: grupo de trabajo

Tabla 29: Precisión por Clase Voted Perceptrom Experimento 3

Precisión detallada por clase 3

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Área	PRC Área	Class
0,263	0,007	0,714	0,263	0,385	0,413	0,652	0,288	0
0,993	0,737	0,951	0,993	0,972	0,413	0,651	0,954	1
0,946	0,69	0,936	0,946	0,934	0,413	0,651	0,911	

Fuente: grupo de Trabajo

- **Experimento 4**

El cuarto experimento de Voted Perceptrom utilizado por ser un algoritmo de entrenamiento que se puede utilizar en grandes cantidades de datos, y permite reducir datos y el tiempo de

entrenamiento (Martišius et al., 2013). También, debido a que determina la capacidad del modelo (Malhotra, 2016). Así también, se menciona que el experimento está relacionado con la variable dependiente de estudio la satisfacción educación recibida (Inr 16). El número de particiones utilizadas fue validación cruzada 10 veces, al igual que en el resto de experimentos. Se configura esta función en el exponente 4.0, y seed 1. Como resultado se muestran en la tabla 30, 279 fueron clasificadas correctamente y 15 instancias incorrecta, lo que proporciona en la tabla 31 una confiabilidad de 94.9% en el resumen porcentual del cuarto experimento, con un valor en verdaderos positivos 0,949% y los falsos positivos a 0.640%.

Tabla 30 Validación Cruzada Estratificada

Stratified cross-validation			
Correctly Classified Instances	279	94.898	%
Incorrectly Classified Instances	15	5.102	%
Kappa statistic	0.4223		
Mean absolute error	0.051		
Root mean squared error	0.2256		
Relative absolute error		41.1757 %	
Root relative squared error		91.7357 %	
Total Number of Instances	294		

Fuente: grupo de trabajo

Tabla 31 Precisión por Clase Voted Perceptrom Experimento 4

Precisión detallada por clase 4

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Área	PRC Área	Class
0,316	0,007	0,750	0,316	0,444	0,466	0,680	0,352	0
0,993	0,684	0,955	0,993	0,973	0,466	0,680	0,958	1
0,949	0,640	0,941	0,949	0,939	0,466	0,680	0,919	

Fuente: grupo de trabajo

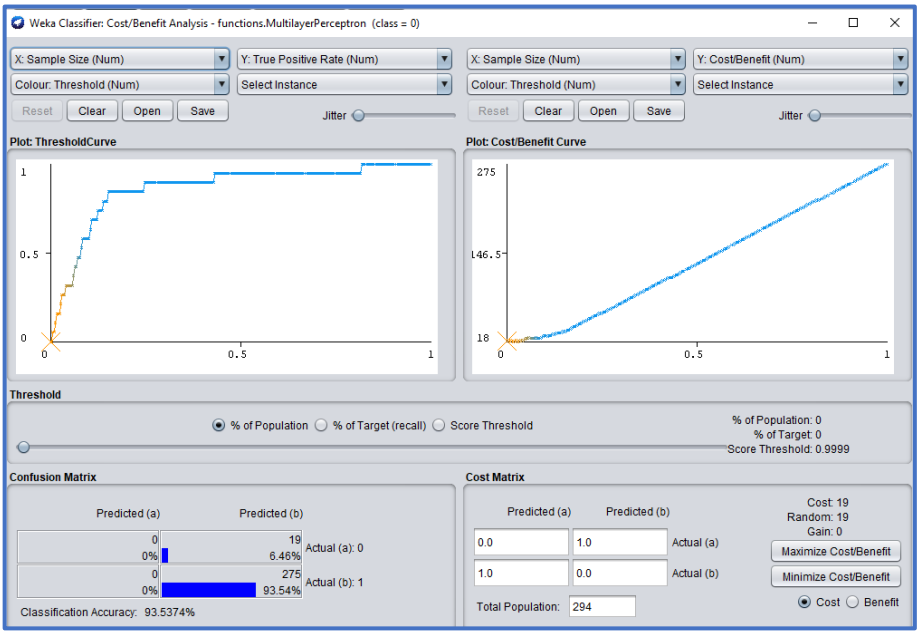
11.8.5 Validación de los Modelos de Predicción

En la etapa de interpretación y evaluación del modelo se detallarán las métricas que se utilizaron en las diferentes técnicas aplicadas que pertenecen a precisión (accuracy) y análisis de costo beneficio.

11.8.5.1 Validación del Red Perceptrom Multicapa

En la experimentación del algoritmo Perceptrom Multilayer se aplicó la validación del modelo sobre el análisis de los indicadores accuracy, costo- beneficio y thresold curve. Para los autores Romero & Martínez (2019) la técnica accuracy es la métrica más simple que indica el grado de correlación entre los resultados y las predicciones observadas. Además, el modelo mejorara a medida que aumente el margen error estos procesos se realizaran en la herramienta Weka (Kuhn & Johnson, 2013). Los resultados de la experimentación se muestran en la siguiente Figura 17, que entrega una predicción de 93.54%, y un error del 6. 46 % lo que indica que el modelo es apto para su aplicación.

Figura 17: Validación Modelo Perceptrom Multilayer



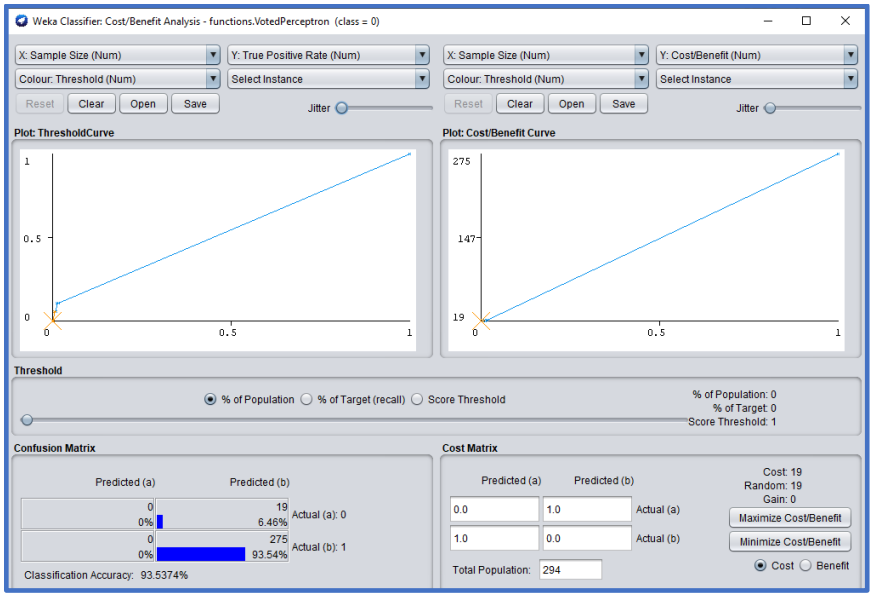
Fuente: Weka

11.8.5.2 Validación de la Red Neuronal Voted Perceptrom

El análisis de la precisión accuracy según los autores (Lehr et al., 2016) es una forma de medir el rendimiento de una red que contenga un alto valor predictivo en sus resultados, además, los resultados de la validación cruzada recopilan y promedian la estimación de la precisión accuracy que se realiza en la herramienta Weka ya que es la más usada para estos procesos (Bouckaert, Frank, Hall, & Kirkby, 2013). En la Figura 18 se visualiza los resultados obtenidos

del análisis de costo beneficio y acurracy que muestra que la predicción de datos positivos en 93.54%, y una clasificación de Accuracy es de 6.4626%.

Figura 18: Validación del Modelo Voted Perceptrom

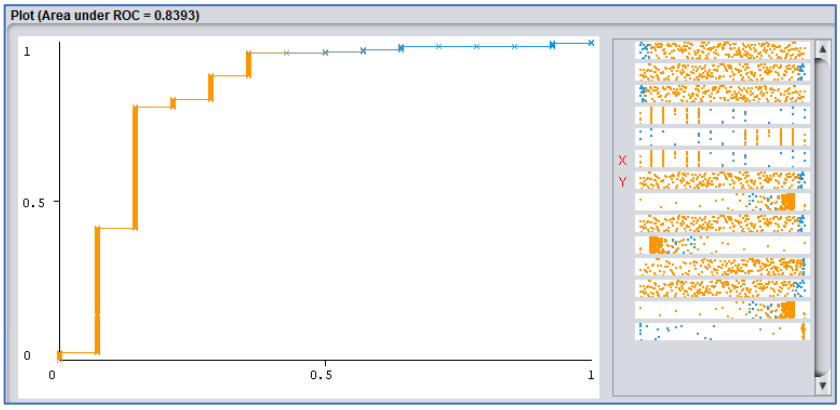


Fuente: Weka

11.8.5.3 Thresold Curve Red Neuronal Perceptrom Multilayer

Los investigadores Bouckaert, Frank, Hall, & Kirkby (2013) indica que para la obtención de una Curva ROC se busca ubicar en el eje de la X los coeficientes de falsos positivos, y en eje de las Y los coeficientes de verdaderos positivos, lo cual permitirá la visualización de la gráfica y el valor del Área correspondiente. La figura 19 identifica los resultados del proceso experimental del área ROC en el algoritmo Perceptrom Multicapa con un 0.8393%, de predicción lo que permite determinar que los factores inciden en su mayoría en la retención universitaria.

Figura 19: CostCurve 1-Perceptrón Multicapas

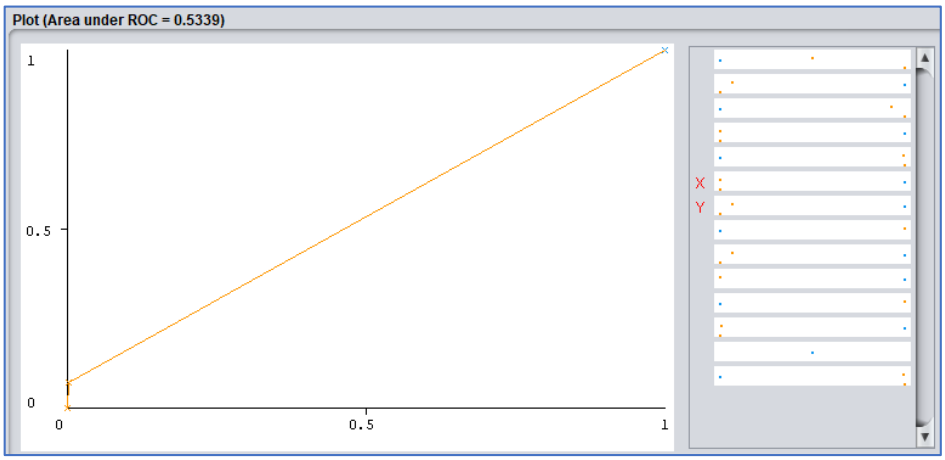


Fuente: Weka

11.8.5.4 Thresold Curve Red Neuronal Voted Perceptrom

El thresold curve según los autores Bouckaert, et,…, al. (2013) indica la variación de las proporciones que tiene cada clase. Se busca ubicar coeficientes con falso positivos en X, y verdaderos positivos en Y, para obtener una gráfica y su valor correspondiente sobre la curva ROC. La figura 20 muestra resultados en el Área ROC de 0.539%, lo que indica un valor bajo de predicción sobre los factores que inciden medianamente en la retención universitaria.

Figura 20: CostCurve

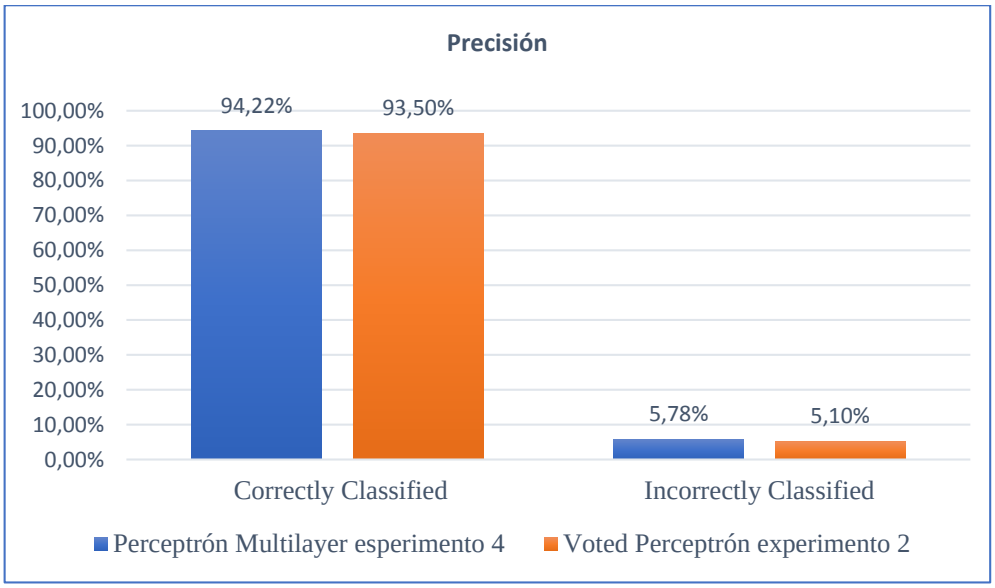


Fuente: Weka

11.9 Precisión de los Modelos de Predicción

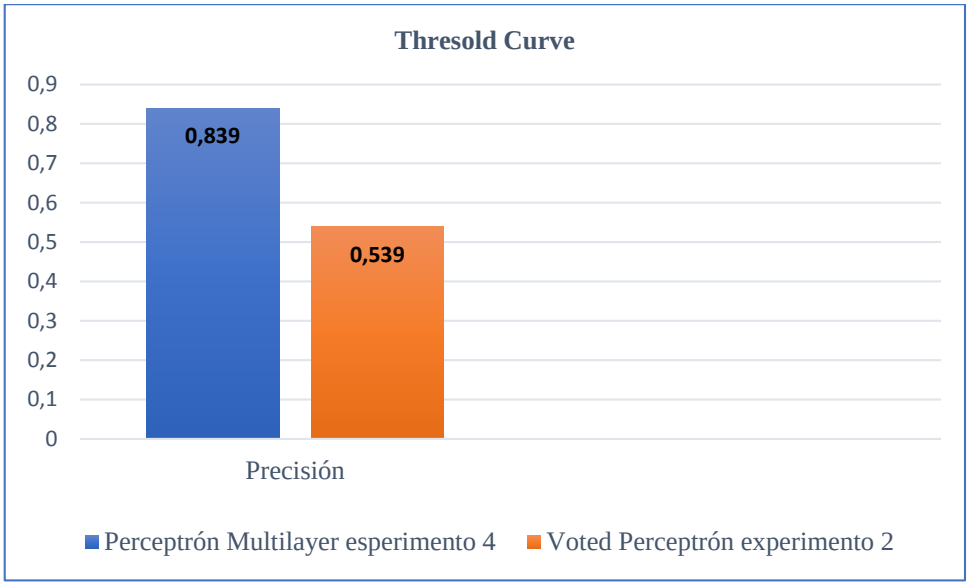
La precisión con las métricas establecidas pueden determinar la relación con los datos encontrados a través de la clasificación correcta y la reducción de errores en la clasificación al mínimo como se observa la curva de ROC (Castro & Sulla, 2016).Una vez pronosticada la satisfacción en la educación recibida (Inr 16), se puede determinar en las figuras 21 y 22 la tasa de precisión cambia con los indicadores de las métricas de evaluación, especificidad y sensibilidad. La comparación de los resultados nos permite confirmar la capacidad predictiva de la técnica Perceptrom Multilayer que indica una predicción alta en 94.20% considerada como el mejor modelo a ser aplicado en la retención universitaria.

Figura 21: Precisión de los Modelos de Predicción



Fuente Weka

Figura 22 Precisión Área COR



Fuente Weka

11.10 Discusión de los Resultados Obtenidos

El desarrollo de la investigación inicia con la determinación de factores de éxito de la retención estudiantil universitaria. Con el diseño del modelo teórico y mediante procedimiento estadístico con regresión lineal se obtienen las variables más significativas que permiten la construcción del modelo propuesto. Este proceso estadístico experimental permite determinar las variables que son influyentes en el análisis de la retención, para validar el modelo teórico se diseña modelos de predicción y clasificación que permiten conocer si el modelo obtenido es óptimo en términos de confiabilidad, mediante la precisión de fiabilidad que determina la probabilidad de influencia que tienes estos factores en la retención universitaria.

Para determinar los factores y la predicción del modelo de retención universitaria se definió la variable satisfacción en la educación recibida (Inr 16) como variable dependiente y el restante de variables como independientes o predictores. Los procesos estadísticos, así como, la predicción del modelo en su experimentación se realizó en base a estas variables identificadas mediante la encuesta aplicada a los estudiantes universitarios. El análisis de predicción de los datos se llevó a cabo a partir de la minería de datos mediante 4 fases que contempla la metodología del Descubrimiento del conocimiento en base de datos (Kdd), que permitirá la validación del modelo teórico propuesto.

En la determinación de los factores de éxito se procede a realizar el proceso estadístico experimental, mediante la regresión lineal con el modelo ajustado con un coeficiente del 95% de confianza de la técnica, del procedimiento experimental se obtienen 7 predictores como la valoración de esfuerzo y trabajo del estudiante, trabajo ene quipo, la formación de nivelación, satisfacción académica, asignaturas y contenidos curriculares adecuados, el apoyo familiar en las expectativas futuras de estudiante, y la perspectiva de la inserción laboral.

Por otra parte, en la predicción de la retención se utilizó la técnica no supervisada Clúster con el algoritmo K-means del cual se obtuvo tres grupos (Grijalva Arriaga, 2018). En el análisis de datos de los estudiantes se pude evidenciar que el mayor grupo se concentró en el clúster 3, en donde se encuentran estudiantes con indicadores relevantes como la aspiración a obtener un título, interacción entre profesor alumno, interacción en actividades con compañeros, compromiso con la formación.

En la etapa de extracción del conocimiento, podemos mencionar que el pre-procesamiento es un determinante en el desarrollo de la minería de datos ya que permiten encontrar características representativas y la reducción de variables. Además, la aplicación de filtros y métodos nos permite obtener una alta correlación sobre 7 variables de predicción que contribuirán al modelo propuesto (Bach et al., 2017). Podemos incluir la fase de extracción mediante los procesos de predicción que se realizaron entre dos algoritmos, el primer proceso de predicción se realiza mediante la técnica Perceptrom Multilayer, y el segundo proceso se realiza con la técnica Voted Perceptrom. Las dos experimentaciones constan de la intervención de 4 experimentos que guardaran relación con la variable dependiente Satisfacción educación recibida, además se utiliza la métrica validación cruzada 10 veces que permite identificar el modelo con mayor valor pronosticado, estos resultados contribuyen a la toma de decisiones para determinar los modelos propuestos.

El resultado de la variación de los experimentos propuestos mediante la técnica Perceptrom Multilayer, proporciona un resumen porcentual de aciertos desplegados por clase que indican valores en verdaderos positivos 0,942 y los falsos positivos se identifica un 0.543, la predicción de estos experimentos muestran un valor alto sobre el último experimento catalogándolo como el mejor para su aplicación según (Mahajan & Zaveri, 2017) lo que confirma un modelo apto

para predecir la retención estudiantil universitaria. Además, se presenta varios experimentos que se realizaron mediante el algoritmo Voted Perceptrom. El modelo muestra que el experimento 4 presenta mejores resultados sobre los demás, en donde se muestra el resumen porcentual en verdaderos positivos 0,949% y los falsos positivos a 0.640%, que representa una predicción alta en 94.9%, lo que indica un modelo predictivo adecuado (Mahajan & Zaveri, 2017).

Podemos incluir la evaluación de los modelos como una fase importante para determinar el rendimiento que tiene la red neuronal, se valida la experimentación mediante la evaluación accuracy y coste curve que muestra resultados de confiabilidad en el modelo Perceptrom Multilayer en 93.54% de precisión y 0,8393 % de sensibilidad, lo que indica que el modelo teórico tiene una probabilidad de influencia en los factores de retención estudiantil universitaria (Sassano, 2008),(Lehr et al., 2016).

12 IMPACTOS

12.1 Impacto Institucional

Usando modelos basados en minería de datos para determinar los factores que influyen en la retención estudiantil universitaria, se puede formular estrategias adecuadas para tomar dediciones oportunas en la carrera de sistemas de información, asegurando así la contribución permanente de los estudiantes a la comunidad universitaria.

12.2 Impacto Económico

Se considera importante puesto que permite conocer el efecto que causa la retención estudiantil a nivel económico, también la tasa de retención estudiantil es considera un indicador de calidad en las instituciones. Además, el incremento de la retención en los estudiantes permitirá el aumento del presupuesto asignado a las instituciones universitarias.

12.3 Impacto Social

La comunidad universitaria seria la beneficiada porque se puede identificar factores similares previo un seguimiento estudiantil y retribuir con estrategias de retención, lo que permite a los estudiantes completar su carrera y lograr titularse dentro del tiempo establecido para su preparación profesional.

13 PRESUPUESTO

En el siguiente aporte se muestran en la tabla 32 los gastos directos e indirectos del proyecto de investigación, también consta el detalle exhaustivo de cada uno de los gastos.

Tabla 32 Presupuesto

PRESUPUESTO					
Recursos	**Detalle**	**Personas**	**Cantidad**	**V. Unitario**	**V. Total**
Gastos Directos					
Tecnológicos	Internet		320hrs	0,60	190,00
Materiales y suministros	Resma de papel	-	1	2,50	2,50
	Impresiones	-	450	0,08	36,00
	Esferos	-	2	1	0,70
	Copias	-	40	0,05	2,00
	Cuaderno	-	1	1,00	1,00
	Empastados		2	18,00	36.00
Gastos Indirectos					
Económico	Alimentación	2	5 meses	75	150.00
				Sub Total	418.20
				Imprevistos 10%	41.82
				Total	420,42

14 CONCLUSIONES Y RECOMENDACIONES

14.1 Conclusiones

- ✓ La revisión de la literatura permitió encontrar 130 fuentes de investigación que permitieron dar sustento al marco teórico para el desarrollo de la investigación, así como

también, a la pregunta de investigación e hipótesis para la construcción del modelo teórico de retención universitaria.

- ✓ Se establece un modelo teórico de retención universitaria basada en 46 factores que a partir de un proceso experimental utilizando la regresión lineal, se determinaron 7 factores que influyen en la decisión de los estudiantes en permanecer en la institución.

- ✓ Se validad el modelo de predicción propuesto a través del uso de técnicas de Machine Learning mediante las métricas accuracy y CostCurve, que permite establecer que la técnica Perceptrom Multilayer presenta una mayor tasa de precisión en 94.21%, señalando que el modelo propuesto es confiable en términos de fiabilidad para la aplicación de la retención estudiantil universitaria.

14.2 Recomendaciones

- ✓ Se recomienda la comprobación del modelo teórico en otras instituciones de educación superior para establecer la incidencia de los factores de retención propuestos.

- ✓ Se puede usar otros algoritmos como la red neuronal Deep Learning debido a su relevancia en los últimos años y la referencia en los diferentes estudios para determinar si la tasa de predicción propuesta puede ser mejorada.

- ✓ Se recomienda que el proyecto sea aplicado a otras carreras de la institución, además se debería tomar en cuenta específicamente a los estudiantes de semestres iniciales, pues en la referencia teórica, muestra que están expuestos a presentar algún problema relacionado con la investigación, por lo cual se debería actuar a tiempo para contribuir a la retención en la Educación Superior.

15 BIBLIOGRAFÍA

Aa, V. (2018). *XIX Foro de investigación en comunicación. La gestión de los contenidos en ... - Aa. Vv. - Google Libros*. https://books.google.com.ec/books?id=n5yFDwAAQBAJ&lpg=PT223&dq=aprendizaje automático&hl=es&pg=PT223#v=onepage&q=aprendizaje automático&f=false

Aguilar-Barojas, S. (2005). Fórmulas para el cálculo de la muestra en investigaciones de salud. *Salud En Tabasco*, 2–7. https://doi.org/ISSN:1405-2091

Arellano, N., Fernández, J., Rosas, M., & Zuñiga, M. (2014). Estrategia metodológica de la enseñanza de la programación para la permanencia de los alumnos de primer año de Ingenieria Electrónica. *TE & ET: Revista Iberoamericana de Tecnología En Educación y Educación En Tecnología*, *13*, 55–60.

Areu, J. L. (2014). El Método de la Investigación. *Daena: International Journal of Good Conscience*, *9*(3), 195–204.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40–79. https://doi.org/10.1214/09-SS054

Armijo, P. C., Zárate, T. M., & Carvajal, C. M. (2019). Evaluación de un programa de apoyo psico-social en torno a los conceptos de persistencia y retención universitaria. *Revista Brasileira de Educação*, *24*, 1–24. https://doi.org/10.1590/s1413-24782019240058

Ayala, M., & Atencio, I. (2018). Retención en la educación universitaria en Chile. Aplicación de un modelo de ecuaciones estructurales. *Revista de La Educación Superior*, *47*(186), 93–118. https://doi.org/10.36857/resu.2018.186.350

Ayala Reyes, M. C., & Atencio Abarca, I. J. (2018). Retención en la educación universitaria en Chile. Aplicación de un modelo de ecuaciones estructurales. *Revista de La Educación Superior*, *47*(186), 93–118. https://doi.org/10.36857/resu.2018.186.350

Babić, I. D. (2017). Machine learning methods in predicting the student academic motivation. *Croatian Operational Research Review*, *8*(2), 443–461. https://doi.org/10.17535/crorr.2017.0028

Bach, M. P., Zoroja, J., Jakovic, B., & Sarlija, N. (2017). Selection of variables for credit risk data mining models: Preliminary research. *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2017 - Proceedings*, 1367–1372. https://doi.org/10.23919/MIPRO.2017.7973635

Bautista Agustin, H., & Calderon Nepamuceno, D. M. (2019). *Aplicación de la metodología*

CRISP DM en la Minería de Datos en la clasificación de empresas.

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education*, *12*(2), 155–187. https://doi.org/10.1007/BF00976194

Bouckaert, R., Frank, E., Hall, M., & Kirkby, R. (2013). *WEKA Manual for Version 3-7-10*.

Braulio, N., & Josep, C. (2015). *Customer analytics - Núria Braulio Gil, Josep Curto Díaz - Google Libros*. https://books.google.com.ec/books?id=GqenDAAAQBAJ&pg=PT20&dq=fases+de+la+metodología+KDD&hl=es&sa=X&ved=2ahUKEwizwb6ym7vrAhXLwVkKHQ4_Dl0Q6AEwA3oECAYQAg#v=onepage&q=fases de la metodología KDD&f=false

Cardona, T. A., & Cudney, E. A. (2019). Predicting student retention using support vector machines. *Procedia Manufacturing*, *39*(2019), 1827–1833. https://doi.org/10.1016/j.promfg.2020.01.256

Carrera, R. (2019). USO DEL ESTIMADOR DE MÍNIMOS CUADRADOS ORDINARIOS EN LA INFERENCIA CON DATOS DE SERIES DE TIEMPO EN MODELOS LINEALES USING ORDINARY LEAST SQUARES ESTIMATOR IN INFERENCE WITH TIME SERIES DATA IN LINEAR MODELS. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Cassiano, Angela, Cipaguata, Patricia, Reyes, N. (2016). *IDENTIDAD PROFESIONAL COMO FACTOR EXPLICATIVO DE LA PERMANENCIA ESTUDIANTIL*.

Castañeda, M., Cabrera, A., Navarro, Y., & De Cirie, W. (2010). *Procesamiento de datos y análisis estadísticos utilizando SPSS : un libro ...* - *Google Libros*. Editora Universitaria da PUCRS. https://books.google.com.ec/books?id=XsxqTVs9-2QC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Castiello-Gutiérrez, S. (2019). Educación superior: ¿De masiva a universal... a obligatoria? *Revista de Educación Superior En América Latina*, *6*, 10–13. https://doi.org/10.14482/esal.6.378.1

Castillo, J., Guillen, A., & Badii, M. H. (2008). Tamano optimo de la muestra (Optimum sample size). *Innovaciones de Negocios*, *5*(1), 53–65.

Castro, A. C., & Sulla-Torres, J. (2016). *Comparación De Algoritmos De Clasificación Para La Predicción De Casos De Obesidad Infantil*. *April*.

Celada, V. L., & Lattuada, M. (2018). *La evaluación en la universidad. Algunas experiencias*

internacionales. *6*(Junio), 41–70.

Cestero, E. V., & Caballero, A. M. (2017). *Data science y redes complejas: Métodos y aplicaciones*. https://books.google.com.ec/books?id=rQNGDwAAQBAJ&pg=PA131&dq=algoritmo+k-means+(Cestero+y+Caballero,+2018).&hl=es&sa=X&ved=2ahUKEwjg9NSRlLfrAhUGpFkKHcxSA5sQ6AEwAHoECAAQAg#v=onepage&q=k-means&f=false

Chanchí, G., Gómez, M., & Campo, W. (2019). Criterios de usabilidad para el diseño e implementación de videojuegos. *Iberian Journal of Information Systems and Technologies*, 461–474.

Chaudhury, P., Mishra, S., Tripathy, H. K., & Kishore, B. (2016). Enhancing the capabilities of student result prediction system. *ACM International Conference Proceeding Series*, *04-05-Marc*. https://doi.org/10.1145/2905055.2905150

Chirivella, V. (2019). Hipótesis en el modelo de regresión lineal por Mínimos Cuadrados Ordinarios. *Univeridad Politécnica de Valencia*, 8.

Chong, E. (2017). Factores que inciden en el rendimiento académico de los estudiantes de la Universidad Politécnica del Valle de Toluca Factors affecting the academic performance of students of the Universidad Politécnica del Valle de Toluca. *Revista Latinoamericana de Estudios Educativos*, *47*(1), 91–108.

De la Cruz, K. (2019). *COMUNICACIÓN FAMILIAR EN ESTUDIANTES DE UNA INSTITUCIÓN EDUCATIVA, CHIMBOTE, 2018*.

Dominguez, J. (2015). *Manual de La Metodologia De la Investigacion Cientifica* (pp. 8–121).

Donoso, S., Donoso, G., & Arias, Ó. (2010). Iniciativas de retención de estudiantes de educación superior. *Calidad En La Educación*, *33*, 15. https://doi.org/10.31619/caledu.n33.138

Dutt, A. (2015). Clustering Algorithms Applied in Educational Data Mining. *International Journal of Information and Electronics Engineering*, *5*(2), 112–116. https://doi.org/10.7763/ijiee.2015.v5.513

EcuRed. (2018). *Minería de Datos - EcuRed*.

Espinosa Zúñiga, J. J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. *Ingeniería Investigación y Tecnología*, *21*(1), 1–13. https://doi.org/10.22201/fi.25940732e.2020.21n1.008

Esteban, M., Bernardo, A., Tuero, E., Cervero, A., & Casanova, J. (2017). Variables

influyentes en progreso académico y permanencia en la universidad. *European Journal of Education and Psychology*, *10*(2), 75–81. https://doi.org/10.1016/j.ejeps.2017.07.003

Estudio para identificar y evaluar la permanencia y retención de los estudiantes del programa administración de empresas de la universidad de Cundinamarca Seccional Girardot.pdf. (n.d.).

Ethington, C. A. (1990). A psychological model of student persistence. *Research in Higher Education*, *31*(3), 279–293. https://doi.org/10.1007/BF00992313

Fabre, I. Z., Eugenia, A. M., & Ph, D. (2019). *TEMA : Análisis de retención de estudiantes en la Carrera de Negocios Internacionales de la Facultad de Ciencias Económicas y Administrativas de la UCSG , durante los dos primeros años de la carrera , mediante técnica de Minería de Datos AUTOR : Zambrano* .

Fàbregues, S., Meneses, J., Rodríguez, D., & Paré, M. (2016). *Técnicas de investigación social y educativa - Fàbregues Feijóo, Sergi, Meneses Naranjo, Julio, Rodríguez Gómez, David, Paré, Marie-Hélène - Google Libros*. Editorial UOC. https://books.google.com.ec/books?hl=es&lr=&id=ZT_qDQAAQBAJ&oi=fnd&pg=PT8&dq=técnicas+de+investigación&ots=_iZGHNGj4Z&sig=15kWsa60zrzpVmBB0C8m8-tb3bs#v=onepage&q=encuesta&f=false

Fadias, G, A. (2012). *El proyecto de investigación*.

Ferrão, M. E., & Almeida, L. S. (2018). Multilevel modeling of persistence in higher education. *Ensaio*, *26*(100), 664–683. https://doi.org/10.1590/S0104-40362018002601610

Fishbein & Ajzen (1975). (n.d.).

Flores, J. A. L. y B. (2010). *LÍDERES DE APRENDIZAJE Y RETENCIÓN*. 1–10.

Franco Ortiz, M., Maldonado Méndez, Y., Berríos Negrón, A., Ortiz, L., España, A., & Torres Rivera, C. (2016). Educacion subgraduada y pobreza: Investigacion y reflexiones en torno al acceso y retencion de estudiantes con desventaja socioeconomica. *Revista Puertorriqueña de Psicología*, *27*(2), 242–258.

Frutos, C. (2017). Análisis de la tasa de retención y su incidencia en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato. *Repo.Uta.Edu.Ec*, 148.

Galán, V. (2015). *Aplicación de la Metodología CRISP-DM a un Proyecto de Minería de Datos en el Entorno Universitario*.

Gamarra, D., Matos, R., & Yupanqui, M. (2018). Detección de patrones de éxito en estudios universitarios de la Universidad Continental. *Apuntes de Ciencia & Sociedad*, *08*(01).

https://doi.org/10.18259/acs.2018005

Gambäck, B., & Sikdar, U. K. (2017). *Using Convolutional Neural Networks to Classify Hate-Speech*. *7491*, 85–90. https://doi.org/10.18653/v1/w17-3013

García-García, J. A., Reding-Bernal, A., & López-Alvarenga, J. C. (2013). Cálculo del tamaño de la muestra en investigación en educación médica. *Investigación En Educación Médica*, *2*(8), 217–224. https://doi.org/10.1016/s2007-5057(13)72715-7

Garcia, Sergio, Ramírez-Gallego, Herrera, F. (2016). Big Data: Preprocesamiento y calidad de datos. *Novática*, *237*, 17.

García, D. (2008). Manual de WEKA. *Disponível Através Do E-Mail Diego. Garcia. …*.

García, G., García, R., & Ledeneva, Y. (2014). Reglas que describen la deserción y permanencia en los estudiantes de la uap Tianguistenco de la uaem. *Ciencia Ergo Sum*, *21*, 121–132. http://www.redalyc.org/articulo.oa?id=10431177003

Garzón, J., & Landin, M. (2013). UNIVERSIDAD DE CUENCA - TESIS.pdf. *Articulo Ecuador*, *1*(5), 1–127.

Gomaa, W. H. (2019). A multi-layer system for semantic relatedness evaluation. *Journal of Theoretical and Applied Information Technology*, *97*(23), 3536–3544.

González-Ruiz, S. L., Gómez-Gallego, I., Pastrana-Brincones, J. L., & Hernández-Mendo, A. (2015). Classification algorithms and neural networks in automated observation records | Algoritmos de clasificación y redes neuronales en la observación automatizada de registros. *Cuadernos de Psicologia Del Deporte*, *15*(1), 31–40.

González Martínez, J. A. (2013). *Estado del arte del uso de la nube computacional para el apoyo al aprendizaje*. http://uvadoc.uva.es/handle/10324/4457

González, V. (2015). *El Uso Del Perceptrón Multicapa Para La Clasificación De Patrones En Conductas Adictivas*. 55.

Graus. 2017. (2017). *ESTADÍSTICA APLICADA A LA INVESTIGACIÓN CIENTÍFICA AUTOR: Michel Enrique Gamboa Graus 1*. 15.

Grijalva Arriaga, P. K. (2018). *Rev. Hallazgos21, Vol. 3, Suplemento Especial, 2018 TÉCNICAS DE MINERÍA DE DATOS EN LA EFICIENCIA ACADÉMICA*. *3*, 1–16.

Gutiérrez García, J. A. (2016). *Comenzando con Weka : Filtrado y selección de subconjuntos de atributos basada en su relevancia descriptiva para la clase* . 18.

Haro, S., Pazmiño Maji, R., Conde, M., & Peñalvo, F. (2018). Minería de datos para descubrir tendencias en la clasificación de los trabajos de titulación. *Congreso de Ciencia y Tecnología ESPE*, *13*(1), 125–128. https://doi.org/10.24133/cctespe.v13i1.739

Hernández G, C. L., & Dueñas R, M. X. (2009). *Hacia una metodología de gestión del conocimiento basada en minería de datos*.

Holgado-Apaza, L. A. (2018). Detección de patrones de bajo rendimiento académico mediante técnicas de minería de datos de los estudiantes de la Universidad Nacional Amazónica De Madre De Dios 2018. *Universidad Nacional Del Altiplano*, *051*, 1–7.

Hurtado, E. (2016). El Método de mínimos cuadrados. *Facultad de Ciencias UNAM*, *1*, 4.

Ibarra-piza, S., Segredo-santamaría, S., & Juárez-hernández, L. G. (2018). Estudio de validez de contenido y confiabilidad de un instrumento para evaluar la metodología socioformativa en el diseño de cursos. *Espacios*, *39*(53), 24.

Illinois, S. (2009). *exitosos de retención estudiantil universitaria : las vivencias de los estudiantes 1 Students ' Experiences Programs : Programmes qui ont du succès avec la rétention estudiantin universitaire : les expériences des étudiants*. *28*, 1–30.

Jorge, B., Lozano, T., & Decisión, A. D. E. (2015). *UN MODELO PARA EXPLICAR LA RETENCIÓN EN LA UNIVERSIDAD DE BOGOTÁ JORGE TADEO LOZANO: ÁRBOLES DE DECISIÓN*.

Kaiser, L., Meyers, J., Morrison, D., & Skelton, A. (2016). *Machine Learning to Predict Student Retention*. *147*, 11–40.

Kawano, H. (1997). Knowledge Discovery and Data Mining. *Journal of Japan Society for Fuzzy Theory and Systems*, *9*(6), 851–860. https://doi.org/10.3156/jfuzzy.9.6_851

Kuhn, M., & Johnson, K. (2013). *Applied Predictive Modeling with Applications in R*.

Lattuada, M. (2017). Deserción y retención en las unidades académicas de educación superior. Una aproximación a las causas, instrumentos y estrategias que contribuyen a conocer y morigerar su impacto. *Debate Universitario*, *5*(10), 100–113.

Lehr, S., Liu, H., Klinglesmith, S., Konyha, A., Robaszewska, N., & Medinilla, J. (2016). Use educational data mining to predict undergraduate retention. *Proceedings - IEEE 16th International Conference on Advanced Learning Technologies, ICALT 2016*, *1*, 428–430. https://doi.org/10.1109/ICALT.2016.138

Lin, J. J. J., Imbrie, P. K., & Reid, K. J. (2009). Student retention modelling: An evaluation of different methods and their impact on prediction results. *2009 Research in Engineering Education Symposium, REES 2009*, 1–6.

Lodhi, F. K., Abbasi, I., Khalid, F., Hasan, O., Awwad, F., & Hasan, S. R. (2016). A self-learning framework to detect the intruded integrated circuits. *Proceedings - IEEE International Symposium on Circuits and Systems*, *2016-July*, 1702–1705.

https://doi.org/10.1109/ISCAS.2016.7538895

Macías, M., Gómez, M., Tous, R., & Torres, J. (2016, May). *Introducción a Apache Spark - Mario Macías, Mauro Gómez, Ruben Tous, Jordi Torres - Google Libros*. https://books.google.com.ec/books?id=JqunDAAAQBAJ&lpg=PT115&dq=aprendizaje automático&hl=es&pg=PT25#v=onepage&q=aprendizaje automático&f=false

Mahajan, R. S., & Zaveri, M. A. (2017). Machine learning based paraphrase identification system using lexical syntactic features. *2016 IEEE International Conference on Computational Intelligence and Computing Research, ICCIC 2016*. https://doi.org/10.1109/ICCIC.2016.7919721

Malhotra, R. (2016). An empirical framework for defect prediction using machine learning techniques with Android software. *Applied Soft Computing Journal*, *49*, 1034–1050. https://doi.org/10.1016/j.asoc.2016.04.032

Martínez, C. (2019). *Desarrollo Y Evaluación De Un Modelo Predictivo Basado En Machine Learning Para Estudiar Y Predecir El Comportamiento Del Absentismo En Prestaciones Sanitarias*.

Martišius, I., Šidlauskas, K., & Damaševičius, R. (2013). Real-time training of Voted Perceptron for classification of EEG data. *International Journal of Artificial Intelligence*, *10*(13 S), 41–50.

Matas, A. (2018). Diseño del formato de escalas tipo Likert: Un estado de la cuestión. *Revista Electronica de Investigacion Educativa*, *20*(1), 38–47. https://doi.org/10.24320/redie.2018.20.1.1347

Medrana, R. F., & Romero, S. (2018). Los problemas familiares y el rendimiento académico de los y las estudiantes de 3er. nivel de la carrera de Trabajo Social de la Facultad de ciencias humanísticas y sociales. *Revista Caribeña de Ciencias Sociales*, 11.

Mellado, R, Cifuentes, M, Beltrán, A. (2017). *VARIABLES Y FACTORES ASOCIADOS AL FENÓMENO DE LA RETENCIÓN Y ABANDONO ESTUDIANTIL UNIVERSITARIO EN INVESTIGACIONES DE LATINOAMÉRICA Y EL CARIBE. Línea Temática: Factores asociados. Tipos y perfiles de abandono*. *2013*.

Méndez Suárez, M. (2018). *Análisis de datos con R: Una aplicación a la investigación de mercados - Méndez Suárez, Mariano - Google Libros*. ESIC EDITORIAL. https://books.google.com.ec/books?id=SnhJDwAAQBAJ&pg=PA12&dq=que+es+Rstudio&hl=es&sa=X&ved=2ahUKEwi649PBlazrAhUKw1kKHXGcCSMQ6AEwA3oECAQQAg#v=onepage&q&f=false

Mendoza Gutiérrez, L., Rubio, U. M., & Romero Meléndez, D. (2014). *Permanencia académica: Una preocupación de las instituciones de educación superior Academic tenure: A concern of the institutions of higher education*. 130–137.

Meneses, Paulina , Moraga, Ana , Puchi, R. (2015). Modelo de Apoyo Académico al Estudiante UFRO como aporte a la retención de los estudiantes de primer año. *V Clabes*, 1–16.

Merrill, B. (2011). *Access and Retention : Experiences of Non- traditional Learners in Higher Education Final Report*.

Moine, J. M. (2013). *Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo* . Universidad Nacional de La Plata .

Montaño, J. J. (2002). Redes Neuronales Artificiales aplicadas al Análisis de Datos. *Network*, 275.

Nason, R., Lloyd, P., & Ginns, I. (1996). Format-free databases and the construction of knowledge in primary school science projects. *Research in Science Education*, *26*(3), 353–373. https://doi.org/10.1007/BF02356945

Nell, L. (2014). *Estadística con SPSS 22 - Quezada, Nel - Google Libros*. Editora Marco E.I.R.L. https://books.google.com.ec/books?id=hg0wDgAAQBAJ&pg=PA15&dq=El+programa+SPSS?&hl=es&sa=X&ved=2ahUKEwiTvIjdoazrAhUK2FkKHUi8DFgQ6AEwA3oECAMQAg#v=onepage&q=El programa SPSS%3F&f=false

Oñate, A. A. (2016). *Análisis de la Deserción y Permanencia Académica en la Educación Superior Aplicando Minería De Datos*. 66. http://bdigital.unal.edu.co/53635/1/alvaroagustinoñatebowen.2016.pdf

Owens, J. W. (2013). *A Dissertation by*. *May*, 1–182.

Parada Rico, D. A. (2017). Factores relacionados con la permanencia estudiantil en programas de pregrado de una universidad pública. *Investigación En Enfermería: Imagen y Desarrollo*, *19*(1), 155. https://doi.org/10.11144/javeriana.ie19-1.frpe

Pardo, Y. M., & Rivera, S. I. (2017). Analisis de permanencia y retención de estudiantes de acuerdo con la metodologia que ofrece EL PROGRAMA DE PREGRADO DE ADMINISTRACIÓN FINANCIERA DE LA CORPORACION UNIVERSITARIA MINUTO DE DIOS VICERRECTORIA REGIONAL LLANOS SEDE VILLAVICENCIO. *Вестник Росздравнадзора*, *6*, 5–9.

Pedraza Ortiz, A., Díaz Pérez, V. R., & Cabrales Salazar, O. (2015). Una aproximación

conceptual a la retención estudiantil en Latinoamérica. In *Revista Interamericana de Investigación, Educación y Pedagogía, RIIEP* (Vol. 7, Issue 2). https://doi.org/10.15332/s1657-107x.2014.0002.05

Pellerano, B. D., & Matus, M. (2013). *Bianca Dapelo Pellerano y Manuel Matus Jara Revista de Orientación Educacional V27 Nº52, pp 15-33, 2013*. *27*(52), 15–33.

Pozón, J. R. (2015). Los Estudiantes Universitarios Ante Las Actividades Extracurriculares University Students and Extracurricular Activities. *Universidad Anáhuac México Sur*, *13*, 137–150. https://doi.org/10.12795/anduli.2014.i13.08

Pulido Polo, M. (2015). Ceremonial and protocol: Methods and techniques for scientific research. *Opcion*, *31*, 1137–1156.

Robalino, J, Rosillo, L, Leon, I, Martines, L, Acurio, G. (2020). *Relación entre la capacitación de estudiantes y docentes, y la tasa de retención estudiantil.* 1–15.

Roberti, P. De, Miranda, R., & Roberti, R. P. De. (2010). Visión del desempeño académico estudiantil en la Universidad Centroccidental Lisandro Alvarado Compendium. *Universidad Centroccidental Lisandro Alvarado Barquisimeto, Venezuela*, *13*, 5–21.

Rodríguez Montequín, T. M., Álvarez Cabal, J. V., Mesa Fernández, J. M., & González Valdés, A. (2005). *METODOLOGÍAS PARA LA REALIZACIÓN DE PROYECTOS DE DATA MINING*.

Sampedro Carmen. (2019). *La estadística y la probabilidad en la educación secundaria obligatoria - Google Libros*.

Sampieri, H., Fernández Collado, R., & Baptista Lucio, C. (2004). *METODOLOGÍA DE LA INVESTIGACIÓN*.

Sánchez-Hernández, G., Barboza-Palomino, M., & Castilla-Cabello, H. (2017). Análisis de la deserción y los factores asociados a la permanencia estudiantil en una universidad peruana. *Actualidades Pedagógicas*, *69*, 169–191. https://doi.org/10.19052/ap.4075

SAS Institue Inc. (2017). *SAS ® Enterprise Miner ™ 14.2: Reference Help*. 321–327.

Sassano, M. (2008). An Experimental Comparison of the Voted Perceptron and SUpport Vector Machines in Japanese Analysis Tasks. *International Joint Conference on Natural Language Processing*, 829–834.

Sharma, R. C., Hara, K., & Hirayama, H. (2017). A Machine Learning and Cross-Validation Approach for the Discrimination of Vegetation Physiognomic Types Using Satellite Based Multispectral and Multitemporal Data. *Scientifica*, *2017*. https://doi.org/10.1155/2017/9806479

SIES. (2014). Panorama de la educación superior en Chile 2014. *Servicio de Informacion de Educacion Superior*, 57.

Smola, A. J. (1998). *Learning with Kernels Diplom*. *November*.

Spady, W. G. (1970). Lament for the letterman: The effects of peer status and activities on goals and attainments. *American Journal of Sociology*, *75*, 680–702.

Strauss, A., & Corbin, J. (2002). *Bases de la investigación cualitativa: técnicas y procedimientos para*. https://books.google.com.ec/books?id=TmgvTb4tiR8C&dq=metodo+de+investigacion+cualitativa&hl=es&source=gbs_navlinks_s+

Támara, L. G. (2019). Correlación y regresión. *Análisis Exploratoria de Datos*, 143–178. https://doi.org/10.2307/j.ctvc5pc9g.6

Terraza-Beleño., W. (2019). *ESTRATEGIAS DE RETENCIÓN ESTUDIANTIL EN EDUCACIÓN SUPERIOR Y SU INFLUENCIA EN LA DESERCION*. *3*(4), 39–56. https://doi.org/http://dx.doi.org/10.15658/rev.electron.educ.pedagog19.03030403

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zanbrano. SJ., Hidalgo-Troya, A. y Alvarado-Pérez, J. c. (2016). Introduction. *Ingenierías*, *8*(26), 63–86.

Timarán, S., Hernández, I., Caicedo, S., Hidalgo, A., & Alvarado, J. (2016). *El proceso de descubrimiento de conocimiento en bases de datos*. https://ediciones.ucc.edu.co/index.php/ucc/catalog/download/36/40/230-1?inline=1

Tinto, V. (1975). Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, *45*(1), 89–125. https://doi.org/10.3102/00346543045001089

Tinto, V. A. (1993). *A Longitudinal Study of Freshman Interest Groups at the University of Washington.*

Tinto, V. C. J. (1973). *Abandono en la educación superior: una revisión y una teoria de sintesis de investigacione recientes*.

Tobon, M., Durán, M., & Áñez, A. (2016). Satisfacción Académica Y Profesional De Estudiantes Universitario. *Revista Electronica de Humanidades, Educacion y Comunicacion Social.*, *22*(1957), 111.

Tolegen, G., Toleu, A., & Mussabayev, R. (2020). *Voted-Perceptron Approach for Kazakh Morphological Disambiguation*. *May*, 258–264.

Torres, L. E. (2010). *Estado Del Arte De La Retención De Estudiantes De La Educación Superior*. 131.

https://contextoseducativosinteractivos.files.wordpress.com/2012/11/estado_del_arte_de_la_retencion_de_estudiantes.pdf

Trstenjak, B., & Donko, D. (2014). Determining the impact of demographic features in predicting student success in Croatia. *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2014 - Proceedings*, *May*, 1222–1227. https://doi.org/10.1109/MIPRO.2014.6859754

Univaso, P., Ale, J. M., & Gurlekian, J. A. (2015). Data mining applied to forensic speaker identification. *IEEE Latin America Transactions*, *13*(4), 1098–1111. https://doi.org/10.1109/TLA.2015.7106363

Urbina Cárdenas, J. E., & Ovalles Rodriguez, G. A. (2016). Abandono y Permanencia en la Educación Superior: Una aplicación de la Teoría Fundamentada. *Sophia*, *12*(1), 27. https://doi.org/10.18634/sophiaj.12v.1i.290

Valbuena, R. (2017). *CIENCIA PURA: LA LÓGICA DE PROCEDIMIENTOS Y RAZONAMIENTOS CIENTÍFICOS - ROIMAN VALBUENA - Google Libros*. https://books.google.es/books?hl=es&lr=&id=vJwrDwAAQBAJ&oi=fnd&pg=PR1&dq=Valbuena,+R.+(2017).+Ciencia+Pura:+La+lógica+de+procedimientos+y+razonamientos+científicos.+Maracaibo:+Roiman+Valbuena.&ots=saHtteCPHg&sig=AcAzym5jZLng04z2ssvHnAYpcgU#v=onepage&q=inv

Velázquez Narváez, Y., & González Medina, M. A. (2017). Factors associated with student persistence: The case of the UAMM-UAT. *Revista de La Educacion Superior*, *46*(184), 117–138. https://doi.org/10.1016/j.resu.2017.11.003

Velázquez, Y., & González, A. (2017). Factores asociados a la permanencia de estudiantes universitarios : caso uamm-uat Factors associated with student persistence : The case of the uamm-uat. *Revista de La Educación Superior*, *46*(184), 117–138. https://doi.org/10.1016/j.resu.2017.11.003

Villada, F., Muñoz, N., & García-Quintero, E. (2016). Redes neuronales artificiales aplicadas a la predicción del precio del oro. *Informacion Tecnologica*, *27*(5), 143–150. https://doi.org/10.4067/S0718-07642016000500016

Villavicencio Caparó, E. (2018). El Tamaño Muestral Para La Tesis.¿Cuántas Personas Debo Encuestar? *Odontología Activa Revista Científica*, *2*(1), 59. https://doi.org/10.26871/oactiva.v2i1.175

Virseda, J., Arias, J., Parra, F., & Beltran, M. (2019). *Métodos de Data Science aplicados a la Economía y a la Dirección y ... - VICENTE VÍRSEDA Juan Antonio , GONZÁLEZ*

ARIAS Julio , PARRA RODRÍGUEZ Francisco , BELTRÁN PASCUAL Mauricio - Google Libros. https://books.google.com.ec/books?id=rCi6DwAAQBAJ&pg=PT40&dq=que+es+R+studio&hl=es&sa=X&ved=2ahUKEwicxaHJ0KnrAhWxr1kKHXhfD6YQ6AEwAnoECAIQAg#v=onepage&q=que es R studio&f=false

Widyahastuti, F., & Tjhin, V. U. (2017). Predicting students performance in final examination using linear regression and multilayer perceptron. *Proceedings - 2017 10th International Conference on Human System Interactions, HSI 2017*, 188–192. https://doi.org/10.1109/HSI.2017.8005026

Wolff, A., Zdrahal, Z., Herrmannova, D., & Knoth, P. (2014). Predicting student performance from combined data sources. *Studies in Computational Intelligence*, *524*, 175–202. https://doi.org/10.1007/978-3-319-02738-8_7

16 ANEXOS

ANEXO 1. Encuesta

ENCUESTA

Encuesta para determinar factores que inciden en la retención estudiantil universitaria. La presente encuesta permitirá descubrir los factores que afecta a la tasa de Retención Estudiantil Universitaria. Esta encuesta está dirigida a personas que tengan experiencia con temas relevantes en la retención universitaria y estas respuestas se utilizaran con fines de investigación académica. La encuesta está dividida en dos grupos. El grupo 1 se encuentra relacionada la información personal e información relacionada a la institución, el grupo 2 incluye preguntas que contribuyen a determinar la influencia de factores que inciden la retención estudiantil.

Grupo 1 Información Personal e institucional

1. Cédula Pasaporte (N°)

---

2. Género

Masculino ( )
Femenino ( )

3. Edad

17 - 21 ( )
22 – 26 ( )
27 – 31 ( )
32 – 36 ( )
42 o más ( )

4. Estado civil

Casado(a) ( )
Unión libre ( )

Divorciado(a) ( )
Separado(a) ( )
Soltero(a) ( )
Viudo(a) ( )

5. Etnia

Afro ecuatoriano(a) ( )
Blanco(a) ( )
Indígena ( )
Mestizo(a) ( )
Mulato(a) ( )
No registra ( )
Otro ( )

6. Discapacidad

Si ( )
No ( )

7. Ubicación residencia vive cerca de la universidad

Si ( )
No ( )

8. Lugar de origen

Azuay ( )
Bolívar ( )
Cotopaxi ( )
Imbabura ( )
Loja ( )
Pichincha ( )
Tungurahua ( )

Chimborazo ( )
Carchi ( )
Los Ríos ( )
Santo Domingo ( )
Manabí ( )
Guayas ( )
Napo ( )
Orellana ( )
Pastaza ( )
Sucumbíos ( )
Zamora Chinchipe ( )
Galápagos ( )

9. Tipo de hogar

Inmediato funcional (Otros) ( )
Materno funcional (Madre y hermanos) ( )
Nuclear Funcional (Padre, madre y hermanos) ( )
Nuclear Funcional (Cónyuge e hijos) ( )
Nuclear funcional (Padre y hermanos) ( )

10. Cuál es el número de miembros de su familia

____

11. Tipo de vivienda

Casa arrendada ( )
Casa propia ( )
Departamento arrendado ( )
Departamento propio ( )

12. Ingresos familiares

386 - 772 ( )

773 – 1158 ( )
1159 – 1544 ( )
1545 – 1931 ( )
1932 – 2318 ( )
2706 – 3092 ( )
3867 más ( )

13. Origen de los Recursos para sus estudios

Beca estudio ( )
Crédito educativo ( )
Hermanos ( )
Otros familiares ( )
Otros miembros hogar ( )
Padres tutores ( )
Pareja sentimental ( )
Recursos propios ( )
No registra ( )

14 ¿Cuál es el nivel de formación padre?

Centro de Alfabetización ( )
Educación Básica ( )
Educación media ( )
Jardín de Infantes ( )
Ninguna ( )
Posgrado Maestría ( )
Superior no universitaria completa ( )
Superior no universitaria incompleta ( )
Superior universitaria completa ( )
Superior universitaria incompleta ( )

15 ¿Cuál es el nivel de formación madre?

Centro de Alfabetización ( )
Educación Básica ( )
Educación media ( )
Jardín de Infantes ( )
Ninguna ( )
Posgrado Maestría ( )
Superior no universitaria completa ( )
Superior no universitaria incompleta ( )
Superior universitaria completa ( )
Superior universitaria incompleta ( )

16. ¿Tipo de colegio que procede?

Fiscal ( )
Fisco misional ( )
Municipal ( )
Particular ( )
No registra ( )

Grupo 2: Factores de retención

En este grupo se presentan las preguntas relacionadas con los factores que pueden influencias en la retención de los estudiantes en la universidad.

Conteste las preguntas según lo que usted considere en una escala del 1 al 5.

17. ¿Considera usted que la educación recibida es satisfactoria?

Nada Satisfactorio ( )
Poco Satisfactorio ( )
Medianamente Satisfactorio ( )
Muy satisfactorio ( )
Totalmente Satisfactorio ( )

18. ¿Considera satisfactorio el proceso de formación académica para una inserción laboral?

Nada Satisfactorio ( )
Poco Satisfactorio ( )
Medianamente Satisfactorio ( )
Muy satisfactorio ( )
Totalmente Satisfactorio ( )

19. ¿Considera que la formación en nivelación influye en la adaptación a la institución?

No influye ( )
Baja influencia ( )
Mediana influencia ( )
Alta influencia ( )
Influye totalmente ( )

20. ¿Considera satisfactoria la experiencia del bachillerato para el ingreso a la universidad?

Nada Satisfactorio ( )
Satisfactorio ( )
Poco Satisfactorio ( )
Medianamente Satisfactorio ( )
Totalmente Satisfactorio ( )

21. ¿Considera usted que su rendimiento académico es satisfactorio?

Nada Satisfactorio ( )
Poco Satisfactorio ( )
Medianamente Satisfactorio ( )
Muy satisfactorio ( )
Totalmente Satisfactorio ( )

22. ¿Considera usted que las asignaturas y contenidos en la formación académica es adecuada para su carrera?

Nada Adecuada ( )
Poco adecuada ( )
Medianamente adecuada ( )
Muy Adecuada ( )
Totalmente adecuada ( )

23. ¿Usted está de acuerdo con el tiempo dedicado fuera del horario de clase?

Totalmente en desacuerdo ( )
En desacuerdo ( )
Indeciso ( )
De acuerdo ( )
Totalmente de acuerdo ( )

24. ¿Considera usted importante tener aspiraciones de obtener un título profesional?

Sin importancia ( )
De poca importancia ( )
Moderadamente importante ( )
Muy Importante ( )
Totalmente importante ( )

25. ¿Considera usted, que sus profesores valoran su esfuerzo y trabajo académico en clases?

Nada valorado ( )
Poco Valorado ( )
Mediamente valorado ( )
Altamente valorado ( )
Totalmente Valorado ( )

26. ¿Considera importante la iteración entre el profesor y alumno?

Sin importancia ( )
De poca importancia ( )

Moderadamente importante ( )
Importante ( )
Totalmente importante ( )

27. ¿Considera usted que la figura del profesor influye en su actitud hacia la materia?

No influye ( )
Baja influencia ( )
Mediana influencia ( )
Alta influencia ( )
Influye totalmente ( )

28. ¿Cuál es el grado de motivación, que los profesores incentivan en su aprendizaje?

Nada motivado ( )
Poco motivado ( )
Medianamente motivado ( )
Muy motivado ( )
Totalmente motivado ( )

29. ¿Considera usted, satisfactoria la participación en actividades extracurriculares para su experiencia en la universidad?

Nada Satisfactorio ( )
Satisfactorio ( )
Poco Satisfactorio ( )
Medianamente Satisfactorio ( )
Totalmente Satisfactorio ( )

30. ¿Usted, considera satisfactorio interactuar con sus compañeros de clase?

Nada Satisfactorio ( )
Poco Satisfactorio ( )
Medianamente Satisfactorio ( )

Muy Satisfactorio ( )
Totalmente Satisfactorio ( )

31. ¿Considera que disfrutar de la universidad influye en su preparación académica?

No influye ( )
Baja influencia ( )
Mediana influencia ( )
Alta influencia ( )
Influye totalmente ( )

32. ¿Está, usted de acuerdo con las actividades sociales que proponen sus compañeros?

Totalmente en desacuerdo ( )
En desacuerdo ( )
Indeciso ( )
De acuerdo ( )
Totalmente de acuerdo ( )

33. ¿Usted, está satisfecho con las asignaturas recibida en el aula de clases?

Nada Satisfactorio ( )
Satisfactorio ( )
Poco Satisfactorio ( )
Medianamente Satisfactorio ( )
Totalmente Satisfactorio ( )

34. ¿Considera, usted que los mecanismos para titularse son los más adecuado para la retención estudiantil?

Nada Adecuada ()
Adecuada ( )
Poco adecuada ()
Medianamente adecuada ()

Totalmente adecuada ()

35. ¿Considera importante la interacción con los miembros de la comunidad universitaria?

Sin importancia ()
De poca importancia ()
Moderadamente importante ( )
Importante ( )
Muy importante ( )

36. ¿Considera usted que el dialogo influye entre estudiantes y profesores para la retención estudiantil?

No influye ( )
Baja influencia ( )
Mediana influencia ( )
Alta influencia ( )
Influye totalmente ( )

37. ¿Considera, usted que es importante realizar preguntas a sus profesores en el aula?

Sin importancia ( )
De poca importancia ( )
Moderadamente importante ( )
Importante ( )
Muy importante ( )

38. ¿Cree usted que los trabajo en equipo influyen en la formación académica?

No influye ( )
Baja influencia ( )
Mediana influencia ( )
Alta influencia ( )
Influye totalmente ( )

39. ¿Usted considera que la formación de amistades influye en el grupo de clase?

No influye	( )
Baja influencia	( )
Mediana influencia	( )
Alta influencia	( )
Influye totalmente	( )

40. ¿Considera, importante la facilidad de comunicación entre compañeros?

Sin importancia	( )
De poca importancia	( )
Moderadamente importante	( )
Importante	( )
Muy importante	( )

41. ¿Considera usted que es satisfactorio mantener lazos de amistades fuera del aula?

Nada Satisfactorio	( )
Satisfactorio	( )
Poco Satisfactorio	( )
Medianamente Satisfactorio	( )
Totalmente Satisfactorio	( )

42. ¿Usted considera que el compromiso en la formación académica influye en permanencia estudiantil?

No influye	( )
Baja influencia	( )
Mediana influencia	( )
Alta influencia	( )
Influye totalmente	( )

43. ¿Considera usted importante realizar trabajos cerca de la universidad?

Sin importancia ( )
De poca importancia ( )
Moderadamente importante ( )
Muy Importante ( )
Totalmente importante ( )

44. ¿Considera usted que los problemas familiares influencian en sus estudios?

No influye ( )
Baja influencia ( )
Mediana influencia ( )
Alta influencia ( )
Influye totalmente ( )

45. ¿Considera usted que la visión profesional a futuro influye en la retención estudiantil?

No influye ( )
Baja influencia ( )
Mediana influencia ( )
Alta influencia ( )
Influye totalmente ( )

46. ¿Considera, usted que el esfuerzo dedicado en la formación académica será valorado por sus familiares?

Nada valorado ( )
Poco Valorado ( )
Mediamente valorado ( )
Muy valorado ( )
Valorado totalmente ( )

47. ¿Considera usted que el apoyo familiar en las expectativas a futuro influye en la retención universitaria?

No influye	( )
Baja influencia	( )
Mediana influencia	( )
Alta influencia	( )
Influye totalmente	( )