



ESCUELA DE POSTGRADO
UNIVERSIDAD CÉSAR VALLEJO

Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz – 2016

TESIS PARA OBTENER EL GRADO DE:
DOCTOR EN ADMINISTRACIÓN DE LA EDUCACIÓN

AUTOR:

Mag. Torres Alegre, Luis Augusto

ASESOR:

Dr. Alba Callacná, Rafael Arturo

SECCIÓN:

Educación e Idiomas

LÍNEA DE INVESTIGACIÓN

Gestión y Calidad Educativa

PERÚ - 2017

PÁGINA DEL JURADO

Dr. William Marcial Rodríguez Ramos
Presidente

Dra. Elizabeth Ysmenia Montoya Soto
Secretario

Dr. Rafael Arturo Alba Callacná
Vocal

DEDICATORIA

A la memoria de mi padre Luis, quien me sirvió como ejemplo de superación. A mi querida madre Alina, a quien tengo la dicha de tenerla aún a mi lado.

A mi amada esposa Norma, quien con su constante apoyo me dio la fortaleza para culminar mis estudios. A mis amados hijos Luis Alejandro y Rafael Augusto, a quienes espero servir de ejemplo en sus vidas.

AGRADECIMIENTO

Al personal del Instituto de Educación Superior Tecnológico Público “Daniel Villar” de Caraz, quienes me brindaron su amplia colaboración proporcionándome la información necesaria para desarrollar la presente investigación.

A la Universidad César Vallejo - Huaraz, por acogerme en sus aulas y darme la oportunidad de lograr una de mis grandes metas.

Al doctor Rafael Alba Callacná, por sus acertadas sugerencias y consejos que hicieron posible la conclusión de la presente investigación.

DECLARACIÓN DE AUTENTICIDAD

Yo, Luis Augusto Torres Alegre, estudiante del Programa de Doctorado en Administración de la Educación de la Escuela de Postgrado de la Universidad César Vallejo, identificado con DNI N° 32384844, con la tesis titulada “Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz – 2016”

Declaro bajo juramento que:

- 1) La tesis es de mi autoría.
- 2) He respetado las normas internacionales de citas y referencias para las fuentes consultadas. Por tanto, la tesis no ha sido plagiada ni total ni parcialmente.
- 3) La tesis no ha sido autoplagiada; es decir, no ha sido publicada ni presentada anteriormente para obtener algún grado académico previo o título profesional.
- 4) Los datos presentados en los resultados son reales, no han sido falseados, ni duplicados, ni copiados y por tanto los resultados que se presenten en la tesis se constituirán en aportes a la realidad investigada.

De identificarse fraude (datos falsos), plagio (información sin citar a autores), autoplagio (presentar como nuevo algún trabajo de investigación propio que ya ha sido publicado), piratería (uso ilegal de información ajena) o falsificación (representar falsamente las ideas de otros), asumo las consecuencias y sanciones que de mi acción se deriven, sometiéndome a la normatividad vigente de la Universidad César Vallejo.

Huaraz, 14 de Enero de 2017.



Luis Augusto Torres Alegre

D.N.I. N° 32384844

PRESENTACIÓN

Señores miembros de Jurado, presento ante ustedes la Tesis titulada “Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz – 2016”, con la finalidad de descubrir patrones de deserción estudiantil e identificar los factores que influyen en la deserción relacionados con los datos operativos de los estudiantes, mediante el procesamiento de los datos operativos (personales y académicos) de los estudiantes que abandonaron voluntariamente sus estudios, siguiendo los lineamientos del proceso KDD de descubrimiento del conocimiento en bases de datos y aplicando técnicas de Minería de Datos Educativa, en cumplimiento del reglamento de Grados y Títulos de la Universidad César Vallejo para obtener el Grado Académico de Doctor en Administración de la Educación.

Esperando cumplir con los requisitos de aprobación.

El Autor.

ÍNDICE

	Pág.
Página del jurado	ii
Dedicatoria	iii
Agradecimiento	iv
Declaración de autenticidad	v
Presentación	vi
Índice	vii
Resumen	ix
Abstract	x
I. INTRODUCCIÓN	12
1.1 Problema	23
1.2 Hipótesis	24
1.3 Objetivos	24
II. MARCO METODOLÓGICO	26
2.1 Variables	26
2.2 Operacionalización de variables	26
2.3 Metodología	31
2.4 Tipo de estudio	31
2.5 Diseño	31
2.6 Población, muestra y muestreo	32
2.7 Técnicas e instrumentos de recolección de datos	32
2.8 Métodos de análisis de datos	37
2.9 Aspectos éticos	40
III. RESULTADOS	42
IV. DISCUSIÓN	62

V. CONCLUSIONES	67
VI. RECOMENDACIONES	69
VII. REFERENCIAS BIBLIOGRÁFICAS	71
VIII. ANEXOS	77
8.1 Propuesta pedagógica	77
8.2 Validación de instrumento de recolección de datos por criterio de jueces o expertos	85
8.3 Matriz de consistencia	95

RESUMEN

La presente investigación abordó el estudio de la deserción estudiantil en el instituto Superior tecnológico “Daniel Villar” de Caraz. Tuvo como objetivo obtener conocimientos sobre el comportamiento de los estudiantes que abandonaron voluntariamente sus estudios que sirvan para formular alternativas de solución, mediante el descubrimiento de patrones de deserción estudiantil y los factores asociados que lo producen, siguiendo la metodología del proceso de descubrimiento del conocimiento en bases de datos (proceso KDD), contribuyendo al conocimiento y solución del problema en los institutos superiores tecnológicos del país. La deserción estudiantil fue del 41.75 % en la cohorte de estudiantes que ingresaron al instituto en los años 2010 al 2013. Se analizaron los datos personales y académicos de todos los estudiantes de la cohorte mencionada, que sumaron 594 estudiantes en 4 carreras profesionales. Se utilizaron los datos que fueron recopilados por la institución en las carpetas del postulante y las actas consolidadas de evaluación académica semestral, cuando el estudiante postuló y durante el desarrollo de sus estudios, respectivamente. Se aplicó la técnica de análisis de información para la recolección de datos, consistente en un proceso de selección, pre procesamiento y transformación orientado a la sistematización de los datos, que finalmente se almacenaron en un repositorio de datos. El análisis de los datos multivariantes obtenidos, se realizó mediante técnicas de minería de datos educacional, utilizándose algoritmos de árboles de decisión y reglas de asociación. Finalmente, los patrones descubiertos fueron analizados para determinar los factores asociados a la deserción estudiantil. Se determinó que los principales factores de deserción asociados a los datos analizados, son el promedio obtenido en secundaria y el promedio obtenido en el ciclo anterior al ciclo de deserción. En conclusión, los datos académicos de los estudiantes contenían información oculta que permitió identificar algunos patrones de deserción estudiantil y sus factores asociados.

PALABRAS CLAVE: Deserción estudiantil, Patrones de deserción estudiantil, proceso KDD, Minería de datos educacional, Árbol de decisión, Reglas de asociación.

ABSTRACT

The present investigation addressed the study of student desertion in the Higher Technological Institute "Daniel Villar" of Caraz. The objective of this study was to obtain knowledge about the behavior of students who voluntarily left their studies to formulate alternative solutions, through the discovery of patterns of student dropout and the associated factors that produce it, following the methodology of the process of discovery of knowledge in Databases (KDD process), contributing to the knowledge and solution of the problem in the technological institutes of the country. Student dropout was 41.75% in the cohort of students who entered the institute in the years 2010 to 2013. Personal and academic data of all the students of the mentioned cohort, which totaled 594 students in 4 professional races, were analyzed. We used the data that were collected by the institution in the applicant's files and the consolidated semester academic evaluation records, when the student applied and during the course of their studies, respectively. The data analysis technique was applied for data collection, consisting of a process of selection, pre-processing and transformation oriented to the systematization of the data, which were finally stored in a data repository. The analysis of the multivariate data obtained was performed using educational data mining techniques, using decision tree algorithms and association rules. Finally, the patterns discovered were analyzed to determine the factors associated with student dropout. It was determined that the main drop factors associated with the data analyzed are the average obtained in secondary and the average obtained in the cycle before the cycle of desertion. In conclusion, the academic data of the students contained hidden information that allowed to identify some patterns of student desertion and its associated factors.

KEY WORDS: Student dropout, Student dropout patterns, KDD process, Educational data mining, Decision tree, Association rules.

CAPÍTULO I

INTRODUCCIÓN

I. INTRODUCCIÓN

La deserción estudiantil en la educación superior, es un fenómeno que se produce a nivel mundial pues afecta a todas las naciones en mayor o menor proporción, y tiene asociadas una serie de consecuencias. El estudio de la deserción estudiantil no solo debe centrarse en su análisis estadístico e identificación de los factores que lo producen, sino también en establecer un modelo de predicción para poder predecirlo y prevenirlo (García Lázaro, 2015).

La deserción estudiantil produce un impacto económico en la sociedad, afecta al estudiante a nivel personal, al mercado laboral y al desarrollo del país. La inversión realizada por el estado y el estudiante durante el tiempo que estudió se desperdicia, asimismo el estudiante se siente frustrado al no haber cumplido su meta y además tiene problemas en encontrar empleo, y si lo encuentra, el sueldo que gane será menor en relación al de un profesional, y finalmente, el mercado laboral pierde un personal calificado, lo que repercute en el desarrollo del país (Rumberger, 2011). Las implicancias de la deserción estudiantil, convierte a este fenómeno en un problema que requiere urgente atención.

Los estudios realizados a nivel mundial sobre la deserción estudiantil en educación superior, en su gran mayoría están referidos a la deserción en el nivel universitario, y tan solo unos cuantos se han realizado a nivel de educación superior tecnológica.

En ese sentido, respecto a los estudios realizados para determinar la tasa de deserción en educación superior tecnológica, tenemos que la tasa de deserción durante el primer año en la educación superior tecnológica en Chile, fue del 38 % en el año 2006, mucho mayor a la registrada en la educación superior universitaria que fue del 18 % en el mismo año (Universidad de Chile, 2008). En un estudio realizado entre setiembre de 2009 a marzo de 2010 en el Instituto Superior "Benito Juárez" de Ecuador, se determinó que el 78.57 % de los estudiantes abandonó su carrera (Allauca Armas, 2012). La tasa de deserción estudiantil en el bachillerato tecnológico en México, fue del 15.6 % en el ciclo 2009-2010 (Zaragoya Loya, 2013).

En un estudio realizado por el Ministerio de Educación Nacional de Colombia, determinaron que la tasa de deserción en los institutos tecnológicos de Colombia fue del 53.8 % en el año 2013. Entre las conclusiones a las que arribaron, señalan que para garantizar el éxito en el control de dicho fenómeno, es necesario realizar trabajos de investigación y caracterización rigurosa del estudiante. Para la identificación de soluciones se deben consolidar los sistemas de información y realizar el análisis desde una óptica académica. Para ello, se debe impulsar el uso de las TIC haciendo uso de sistemas de información que produzcan indicadores, alertas tempranas y observaciones que posibiliten el diseño de acciones de mejoramiento (Ministerio de Educación Nacional de Colombia, 2015).

En nuestro país, existen escasos estudios sobre la deserción estudiantil en los institutos superiores tecnológicos. Según el diario Gestión publicado el 9 de mayo de 2013, el nivel de deserción en los institutos superiores en el año 2013, fué de 31 % en el nivel socioeconómico C, de 20 % en el nivel B y 13 % en el nivel A (Como se cita en <http://institutojuanbosco.blogspot.pe/2016/12/desercion-studiantil.html?view=magazine>).

La deserción estudiantil en el instituto superior tecnológico “Daniel Villar” de Caraz, fue de 41.75 % en la cohorte de 594 estudiantes que ingresaron durante los años 2010 al 2013, porcentaje que está muy por encima del nivel señalado para nuestro país en el año 2013. Los estudiantes que postulan al instituto son de bajo nivel económico, provenientes en su gran mayoría de barrios populares, zonas de expansión urbana y zonas rurales.

Son diversos los factores que influyen en la deserción, que se encuentran comprendidos dentro de las categorías institucional, académicas, personales y externas al entorno educativo (Castaño, Gallón, Gómez y Vásquez, 2006).

En muchos casos, la deserción está asociada a factores de rendimiento académico en el nivel secundario, que incide en la poca capacidad del estudiante para responder a las demandas académicas del nivel superior (Tafur, 2012).

Dapozo, Porcel, López, Bogado y Bargiele (2006) en un estudio sobre el nivel cognitivo y actitudinal con el que ingresan los estudiantes a la carrera de sistemas de la universidad FACENA-UNNE de Argentina, encontraron que éste es muy bajo, lo que indica deficiencias en el sistema educativo preuniversitario, situación que es determinante en el rendimiento académico de los estudiantes durante sus estudios en la universidad.

Por otro lado, el inmenso crecimiento en el almacenamiento y disponibilidad de datos en formato digital, junto con el gran desarrollo de la computación, han generado oportunidades de utilizar métodos modernos de análisis de datos como la minería de datos. La educación no es una excepción, pues analizando los datos de los estudiantes se puede obtener resultados que permitan anticipar y evitar efectos negativos para éstos. La comunidad anglosajona ha puesto especial interés en las herramientas de learning analytics y minería de datos, para desarrollar áreas de investigación e intervención en educación superior, especialmente para la generación de modelos predictivos de la deserción y el rendimiento académico (Celis, Moreno, Poblete, Villanueva y Weber, 2015).

Nos encontramos en la era de la información, donde día a día se incrementa la cantidad de datos almacenados en los sistemas de información de las instituciones, los cuales se almacenan en diversas fuentes, estructuras y formatos. Sin embargo, el disponer cada vez de mayor cantidad de datos, no siempre supone el aumento de conocimiento debido a que los métodos de procesamiento convencionales no siempre producen resultados satisfactorios. Las

instituciones educativas enfrentan problemas complejos como mejorar la calidad educativa y disminuir la deserción y la reprobación de los estudiantes, cuya solución requiere disponer de conocimientos que permitan gestionar estrategias apropiadas, por lo que se puede recurrir a la denominada minería de datos educacional para descubrir el conocimiento oculto en los datos (Eckert y Suénaga, 2013).

Tomando en cuenta lo mencionado líneas arriba, con la finalidad de determinar los patrones de comportamiento y los factores asociados a la deserción estudiantil en el instituto, se planteó analizar los datos operativos de los estudiantes, compuestos por sus datos personales y académicos que fueron recolectados en documentos oficiales de la institución. Uno de estos documentos es la carpeta del postulante, que contiene formularios que deben ser llenados con información personal del estudiante, y que además debe ser completada con copias de sus documentos personales y el certificado de estudios secundarios, al momento de inscribirse para el examen de admisión. Los otros documentos son las actas consolidadas de evaluación académica semestral, que contienen la información del rendimiento académico del estudiante cuando cursa sus estudios, que se encuentran en formato electrónico.

En los últimos años, el procesamiento de los datos obtenidos desde entornos de estas características, se viene realizando utilizando técnicas de minería de datos educacional, siguiendo el proceso de descubrimiento de conocimiento en bases de datos (proceso KDD – Knowledge Discovery in Databases). El objetivo es descubrir patrones de deserción ocultos en los datos de los estudiantes y sus factores asociados.

El presente estudio se fundamenta técnicamente en los lineamientos del proceso KDD de descubrimiento de conocimiento en bases de datos. El proceso KDD consiste en la extracción automatizada de patrones que representan el conocimiento oculto en grandes bases de datos, almacenes de datos, la Web, repositorios de datos o flujos de datos, utilizando como herramienta de análisis la minería de datos. La minería de datos es un campo multidisciplinario, se basa en disciplinas como la estadística, aprendizaje de máquinas, reconocimiento de patrones, tecnología de bases de datos, recuperación de información, ciencia de redes, sistemas basados en el conocimiento, inteligencia artificial, computación de alto rendimiento y visualización de datos (Han, Kamber y Pei, 2012).

El proceso KDD es un proceso interactivo e iterativo, que se compone de los siguientes pasos: comprensión del dominio de la aplicación, del conocimiento relevante y de los objetivos del usuario final, creación del conjunto de datos, limpieza y pre procesamiento de los datos, reducción de los datos y proyección, elegir la tarea de minería de datos, elección del algoritmo o algoritmos de minería de datos, minería de datos, interpretación de los patrones encontrados y consolidación del conocimiento descubierto, aplicándolo al sistema (García y Molina, 2012).

En la figura 1, se muestra en diagrama de bloques del proceso desarrollado en la presente investigación:

FIGURA N° 1
PROCESO KDD DESARROLLADO EN LA INVESTIGACIÓN



Fuente: elaboración propia

Un patrón de comportamiento es una manera de conducirse que se mantiene dentro de ciertos márgenes y que es común a un grupo de personas. Los datos operativos de las instituciones educativas de nivel superior tecnológico pueden tener ocultos patrones de comportamiento de los estudiantes, especialmente los relacionados a la deserción estudiantil que no se pueden descubrir con técnica estadísticas convencionales por la cantidad de datos a

analizarse, por lo que se necesita aplicar ciertas técnicas de análisis de datos para su descubrimiento, como la minería de datos educacional. La información obtenida con esta técnica se encuentra en forma gráfica como árboles de decisión, diagramas de conglomerados y dendogramas, o como reglas de asociación.

A nivel internacional, es escasa la información existente de tesis realizadas sobre la aplicación de técnicas de minería de datos educacional a los datos operativos de los estudiantes. Existe mayor información sobre investigaciones realizadas sobre este tema. A nivel nacional y local, no existen tesis, investigaciones ni estudios sobre descubrimiento de patrones de deserción estudiantil ocultos en los datos de los estudiantes, utilizando técnicas de minería de datos educacional.

Como antecedentes, existen investigaciones realizadas en diversos países del mundo, cuyo objetivo ha sido descubrir patrones de deserción estudiantil mediante la aplicación de técnicas de minería de datos educacional, a los datos de los estudiantes obtenidos de los sistemas de información de las instituciones educativas.

Márquez (2015) en su tesis doctoral titulada “Predicción del fracaso y el abandono escolar mediante técnicas de minería de datos”, propone una metodología para realizar el análisis de los datos de 419 estudiantes del primer año de bachillerato de la Universidad Autónoma de Zacatecas en México. Se utilizaron los datos de los estudiantes almacenados en los sistemas de información de la universidad (193 mujeres y 226 hombres) inscritos en el primer semestre (Agosto – Diciembre 2012), que es la etapa donde se presenta el mayor porcentaje de desaprobación y deserción. Se hicieron varios experimentos con la finalidad de comparar los resultados del algoritmo propuesto ICRM frente a otros cinco algoritmos de clasificación disponibles en el software de Minería de Datos Weka (clasificadores bayesianos, máquinas de vectores de soporte, aprendizaje perezoso basado en instancias, reglas de clasificación y árboles de decisión). El objetivo principal de esta investigación, fue obtener un modelo lo más preciso posible que sea capaz de predecir qué estudiantes presentan mayor riesgo de fracasar en sus estudios, ya sea por desertar o por desaprobación las asignaturas. El autor concluye que la tarea de predecir el fracaso escolar de los estudiantes es una tarea muy difícil de conseguir, principalmente por dos causas: la primera es que son muchos los factores que pueden influir en los estudiantes para que desapruében o abandonen sus estudios; y la segunda es que generalmente la información con la que se trabaja para predecir a estos estudiantes está desbalanceada, es decir, el número de estudiantes que aprueban y pasan de curso y el número de estudiantes que no pasan y/o desertan, son totalmente diferentes. Para solucionar estos dos problemas, el autor ha propuesto la utilización de técnicas de minería de datos que recientemente se han empezado a utilizar cada vez con más frecuencia y de manera exitosa, en

lugar de las tradicionales técnicas estadísticas. También en esta tesis el autor explica cómo se puede predecir a los estudiantes en riesgo de desaprobar o desertar, indicando que ya se mostró que esto se consigue fundamentalmente a través de recoger información de los estudiantes, y que sin embargo, es imposible desarrollar un método infalible, que pueda predecir a la totalidad de los estudiantes que fracasan, que siempre hay situaciones que no se pueden predecir y que pueden llevar a los estudiantes a su fracaso. Señala que también se pueden dar los casos en los que los estudiantes en riesgo son detectados y que la institución educativa les ofrece el apoyo que requieren para evitar su fracaso, pero que no están interesados en permanecer en la institución y simplemente esperan que termine el periodo escolar para desertar. Indica que lo anterior suele presentarse en el nivel medio superior, probablemente por la etapa que viven estos estudiantes que en su mayoría tienen entre quince y dieciocho años de edad. Asimismo, señala que en esas condiciones es imposible terminar con el fracaso escolar, pero que trabajos como el desarrollado, contribuyen a tratar de reducir el problema.

En ese mismo sentido, Timarán (2009) en su artículo titulado “Detección de patrones de bajo rendimiento académico y deserción estudiantil con técnicas de minería de datos”, presenta los resultados de la investigación cuyo objetivo fue descubrir patrones que determinen perfiles de bajo rendimiento académico y de deserción estudiantil aplicando técnicas de minería de datos, utilizando los datos de los estudiantes almacenados en las bases de datos de la Universidad de Nariño (Colombia) durante los últimos 15 años. El nivel de deserción está sobre el 32% en los programas de pregrado. Como fuente interna de datos utilizó la base de datos histórica de 46,173 estudiantes, y como fuente externa la información de los colegios de educación secundaria obtenida del Ministerio de Educación Nacional de Colombia. Concluye que las fases de preprocesamiento y transformación de datos fueron las que demandaron más tiempo, debido a la mala calidad de los datos de las bases de datos existentes. El objetivo de la investigación era detectar patrones de bajo rendimiento académico y deserción estudiantil utilizando las tareas de minería de datos de Clasificación y Asociación. Asimismo, concluye que en cuanto a los patrones obtenidos, la Universidad es quien debe tomar decisiones y proponer estrategias de seguimiento a estudiantes con estos perfiles con el fin de prevenir que los estudiantes tengan un bajo rendimiento y de esta manera disminuir los casos de deserción que se presentan.

Como resultado de un trabajo de investigación realizado posteriormente, Timarán, Calderón y Jiménez (2013) publican su artículo titulado “Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos”, en el cual presentan los resultados de la investigación cuyo objetivo fue detectar patrones de deserción estudiantil utilizando técnicas de minería de datos. En la investigación se recopilaron los datos socioeconómicos, académicos,

disciplinarios e institucionales de 6,870 estudiantes de los programas de pregrado de la Universidad de Nariño y de la Institución Universitaria IUCESMAG de la ciudad de Pasto (Colombia), construyéndose un repositorio de datos con la información de los estudiantes que ingresaron a la Universidad de Nariño entre el 2004 y el 2006, con un intervalo de observación hasta el año 2011. Los autores mencionan que aplicando las técnicas de clasificación y clustering sobre los datos de los estudiantes, se ha obtenido un patrón común de deserción estudiantil, consistente en haber obtenido un promedio bajo y el tener cursos desaprobados en los primeros semestres de la carrera. También señalan que entre los factores socioeconómicos asociados a la deserción estudiantil están el pagar una matrícula alta para el tipo de estudiantes que ingresan a la única universidad pública de la región, teniendo en cuenta que los estudiantes provienen de estratos bajos. Señalan también que la zona de procedencia influye en la deserción estudiantil. Otro factor académico asociado a la deserción estudiantil es un promedio bajo en el Instituto Colombiano para el Fomento de la Educación Superior (Icfes). Finalmente, señalan que la evaluación y análisis de estos patrones permitirá apoyar la toma de decisiones por parte de los directivos de las universidades, que deben estar enfocadas en desarrollar estrategias y habilidades orientadas a un mejor aprovechamiento de los programas de retención estudiantil establecidos.

Por su parte, Marcano y Rodríguez (2014) en su artículo sobre la investigación denominada “Minería de datos aplicada a la deserción estudiantil. Caso: Licenciatura en Computación de la Universidad del Zulia-NPF” señalan que el objetivo de su investigación fue descubrir los patrones de comportamiento de los estudiantes que desertaron de la licenciatura en computación de la Universidad del Zulia-NPF (Venezuela), en el periodo 2008 al 2011. Los autores concluyen que los estudiantes desertan principalmente por tres causas. Primero: los conocimientos adquiridos por los estudiantes en el nivel de educación secundaria no son suficientes para realizar estudios en la Licenciatura en Computación, principalmente por tener escasos conocimientos de matemática y lógica formal, materias importantes para los primeros semestres de la carrera. Segundo: la falta de recursos económicos es otro factor que influye en la deserción, debido a que puede impedir que el estudiante cuente con equipos de computación para practicar fuera de la universidad. Tercero: la poca concentración en los estudios y el escaso tiempo dedicado al mismo, son factores que reducen el desempeño académico del estudiante y por lo tanto se convierten en causas de deserción.

En relación al tema de predicción de la deserción estudiantil, Argote, Jiménez y Gómez (2014) en su artículo de investigación “Detección de patrones de deserción en los programas de pregrado de la Universidad Mariana de San Juan de Pasto, aplicando el proceso de descubrimiento de conocimiento sobre base de datos (KDD) y su implementación en modelos

matemáticos de predicción” presentado en la IV Conferencia Latinoamericana sobre el Abandono en la Educación Superior (IV CLABES), abordan el proceso de predicción de la deserción mediante modelos matemáticos. En esta investigación, uno de los objetivos fue determinar el perfil del estudiante que deserta mediante una descripción comparativa entre los patrones evidentes y los patrones no evidentes que influyen en la deserción. El estudio se realizó sobre los datos de 6,930 estudiantes recopilados del sistema de información SPADIES y del sistema de información de la Universidad Mariana de Pasto (Colombia), para las cohortes de estudiantes que ingresaron en los años 2005 al 2011. Los autores estiman que su investigación se convierte en el punto de partida para la aplicación de sistemas para la toma de decisiones en relación con el entorno educativo, apoyado en el desarrollo y ejecución de modelos de minería de datos, para analizar y evaluar los factores que influyen en la deserción universitaria. Asimismo, concluyen que la formulación de los modelos matemáticos de predicción permiten a los responsables de los programas académicos tomar una decisión adecuada, basándose en el estudio temprano de la información de los indicadores que más incidencia tienen en la deserción estudiantil en sus correspondientes programas, con el propósito de plantear estrategias o proyectos orientados a la retención de los estudiantes. En esta investigación, también se descubrió que la Universidad no hace una adecuada recolección de la información del estudiante al momento que éste ingresa a un programa académico, que permita una adecuada caracterización del estudiante y así poder establecer modelos de predicción para la retención de los educandos.

En esta misma línea de predecir la deserción estudiantil, Amaya, Barrientos y Heredia (2014), en su artículo denominado “Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos”, señalan que en su investigación se desarrolló un modelo predictivo de deserción estudiantil, en base al análisis de los datos socio económicos y académicos que se recopilaron en los últimos 6 años, de los estudiantes del programa de Ingeniería de Sistemas de la Universidad Simón Bolívar de Colombia. Para el procesamiento de los datos, se utilizaron técnicas de clasificación, específicamente árboles de decisión. Los autores señalan que para el establecimiento del modelo predictivo, se tuvieron en cuenta variables de tipo personal como edad actual, sexo, ocupación, estado civil, nivel de estudios del padre y de la madre; variables económicas como costo de la matrícula e ingresos familiares, y de carácter académico como semestre, materias cursadas, promedio, jornada y materias perdidas. Estas variables se discretizaron y almacenaron en 2 grupos para obtener un mejor resultado en el momento de clasificar los estudiantes que pueden desertar.

Méndez Alandete (2015), en su trabajo de investigación utiliza minería de datos para determinar los factores que afectan a los estudiantes que desertan, buscando la detección

temprana de los estudiantes en riesgo de deserción, de manera que se puedan definir estrategias para disminuir este fenómeno.

Valero Orea (2009) en su artículo de investigación titulado “Aplicación de técnicas de minería de datos para predecir deserción”, señala que es de vital importancia en los primeros cuatrimestres de estudios universitarios identificar a los estudiantes más propensos a desertar, cual es la probabilidad de que deserten y cuales son los factores que llevan a que lo hagan. Usando minería de datos se puede identificar y calcular la probabilidad que el estudiante deserte, de manera que se pueda proponer con mucha anticipación las estrategias necesarias para disminuir la deserción.

La minería de datos educacional no solo se utiliza en la detección de los factores que influyen en la deserción estudiantil, sino también en el estudio de las causas del bajo rendimiento académico que sería una de las causas de la deserción, y en la creación de modelos predictivos para determinar la probabilidad de que un estudiante deserte (Camana, 2016).

En relación a los algoritmos utilizados en la minería de datos, tenemos que Veitch (2004) en su artículo titulado “Identifying characteristics of high school dropouts: data mining with a decision tree model” señala que el propósito de su estudio fue investigar las causas de la deserción en la educación secundaria en los años 2001-2002, usando minería de datos en las fuentes de datos existentes, utilizando árboles de decisión. El autor concluye que el árbol que se descubrió en este estudio, presenta una cierta capacidad para predecir qué estudiantes están en riesgo de abandonar la escuela.

Aguirre, Valdovinos, Antonio, Alejo y Marcial (2015) en el artículo de investigación denominado “Análisis de deserción escolar con minería de datos”, indican que se realizó el análisis de los datos de 497 estudiantes de la carrera de Ingeniería en Computación de la Universidad de Ixtlahuaca (México). Realizaron el análisis de frecuencias de algunas de las variables y debido a que la mayoría de las variables son categóricas, decidieron utilizar las técnicas de clasificación mediante árboles de decisión para el descubrimiento de los factores que causan la deserción.

En la Universidad de Bogotá Jorge Tadeo Lozano se realizó un estudio para jerarquizar las variables explicativas de la matrícula, analizándose los datos de 1,477 ingresos registrados en el sistema integrado de información académica, utilizando árboles de decisión (Barragán y Gonzales, 2015).

El descubrimiento de patrones de deserción estudiantil, a partir de los datos almacenados en forma física o digital en los sistemas de información de las instituciones educativas, se fundamenta en el proceso KDD de descubrimiento del conocimiento en bases de datos (KDD: Knowledge Discovery in Databases). Este proceso consiste en analizar los datos

obtenidos del sistema de información, para descubrir patrones de comportamiento ocultos en los datos y obtener información relevante, e interpretarla para convertirla en conocimiento que sirva como soporte para la toma de decisiones.

Los datos son descripciones de eventos que suceden en el mundo real y que son registrados en el sistema de información en forma de transacciones o eventos. El almacenamiento de datos es una función básica de un sistema de información, definiéndose el sistema de información como un conjunto de procedimientos realizados en forma manual o automatizada, y de procedimientos orientados a almacenar y procesar los datos, convirtiéndolos en información (Rodríguez y Daureo, 2003).

El proceso KDD es un proceso de extracción de información, orientado a analizar grandes cantidades de datos y descubrir relaciones o patrones de comportamiento, utilizando técnicas de aprendizaje automático, bases de datos, estadística, razonamiento basado en casos, técnicas de representación del conocimiento, adquisición de conocimiento, razonamiento aproximado, redes neuronales y visualización de datos. El proceso KDD ejecuta tareas como la inducción de reglas, problemas de clasificación, agrupamiento o clustering, reconocimiento de patrones de comportamiento, el desarrollo de modelos predictivos, la detección de dependencias, etc. (García y Molina, 2012).

La minería de datos, conocida como descubrimiento de conocimiento en bases de datos (KDD), se puede definir como el campo del descubrimiento de información novedosa y potencialmente útil a partir de grandes cantidades de datos (Baker, 2014).

En la actualidad, los sistemas de información son fundamentales en toda organización para el almacenamiento de su información operativa. Los datos constituyen el soporte de las operaciones diarias de la organización. En ese sentido, en las instituciones educativas superiores se está produciendo un aumento explosivo de los datos almacenados, pero que no están produciendo ningún beneficio al no ser sometidos a un proceso de análisis. En las transacciones diarias los datos tienen un valor operativo, pero no se está obteniendo todo el valor potencial que puede ser extraído de los datos. Puesto que los datos son un recurso ya adquirido por el sistema de información, que se encuentra disponible y en constante crecimiento, habría que plantearse si se está explotando eficazmente todo su potencial, considerando que los datos son un recurso estratégico (Alcolea y Pavón, 2013).

La minería de datos tiene como labor fundamental, el descubrimiento de conocimiento en forma de reglas o patrones a partir de grandes volúmenes de datos, haciendo uso de técnicas y herramientas automáticas o asistidas, de modo que su uso apoye la toma de decisiones (Montalvo, André y Baldoquín 2010).

Algunas herramientas de minería de datos predicen futuras tendencias y comportamientos, utilizando para ello técnicas de varias áreas como la Inteligencia Artificial, las Bases de Datos, la Estadística, la Computación Gráfica y el Procesamiento Masivo, usando como materia prima los datos almacenados en bases de datos (Palomo, s.f.).

La mayoría de los estudios que han utilizado minería de datos, han encontrado que los métodos de aprendizaje automático (por ejemplo, las redes neuronales artificiales, las máquinas de vectores soporte y los árboles de decisión) son superiores a sus equivalentes estadísticos (por ejemplo, la regresión logística y el análisis discriminante) al estar menos restringidos por las hipótesis y producir mejores resultados de predicción (Delen, 2010).

Una de las características significativas de las instituciones de educación superior, es la acumulación explosiva de datos producto de las transacciones operativas que realizan, sin embargo, estos datos no son utilizados para apoyar la administración institucional. Las técnicas de minería de datos pueden ser utilizadas para producir información que explique y prediga el rendimiento académico de los estudiantes y los procesos de deserción estudiantil. La investigación en este campo aplicando la minería de datos está en aumento, utilizándose como materia prima para las investigaciones, los datos almacenados en los sistemas de información, que contienen información que no es evidente pero que es potencialmente útil, la cual se puede descubrir para generar conocimiento, aplicando las técnicas de minería de datos (Sembiring, Zarlis, Hartama, Ramliana, y Wani, 2011).

La minería de datos educacional, nombre utilizado para referirse a la minería de datos aplicado a los repositorios de datos en entornos estudiantiles, hace uso de diversas técnicas como los árboles de decisión, las redes neuronales, k-vecinos más cercanos, Naive Bayes, máquinas de vectores soporte, etc., para descubrir información en forma de reglas de asociación, árboles de decisión, reglas de clasificación y agrupaciones, que se pueden utilizar para muchos fines educativos, como la organización del plan de estudios, la predicción de la inscripción de estudiantes en una especialidad, la mejora de los modelos de enseñanza tradicionales, etc. (Kumar & Chadha, 2011).

La minería de datos educacional aparece como un modelo orientado a la generalización de tareas, métodos y algoritmos para la exploración de datos que provienen de un contexto educativo. Asimismo, permite descubrir y analizar patrones que caractericen los comportamientos que tienen los estudiantes, analizando los datos de los estudiantes almacenados en las bases de datos (Ballesteros, Sánchez-Guzmán y García, 2013).

Las tareas de minería de datos educacional comúnmente utilizadas para el descubrimiento de patrones de deserción estudiantil y bajo rendimiento académico, son las reglas de clasificación, el agrupamiento o clustering, los patrones secuenciales y las asociaciones

entre otras. El conocimiento contenido en los patrones descubiertos, puede ser utilizado para predecir la ocurrencia de esos fenómenos y tomar acciones para prevenirlos (Timarán, 2009).

Las instituciones educativas no escapan a esta revolución tecnológica, pues en los países desarrollados se vienen utilizando estas herramientas para identificar problemas de deserción estudiantil y apoyar la toma de decisiones, con la finalidad de mejorar la calidad educativa. Sin embargo, a nivel nacional no existe literatura referida al tema de la aplicación de minería de datos para descubrir patrones de deserción estudiantil en los institutos superiores tecnológicos.

La investigación se justifica porque sirvió para determinar los factores que inciden en la deserción estudiantil, a partir de los patrones de deserción estudiantil descubiertos al aplicar técnicas de minería de datos a los datos personales y académicos de los estudiantes. Es importante porque se produjo conocimiento sobre los estudiantes que abandonaron voluntariamente sus estudios, identificando algunos de los factores que influyeron en la deserción estudiantil en el Instituto de Educación Superior Tecnológico Público “Daniel Villar”. Se aplicó minería de datos para el análisis de los datos de los estudiantes que ingresaron entre los años 2010 al 2013, desde la aplicación del nuevo diseño curricular básico de la educación superior tecnológica.

Los principales beneficiados con la presente investigación son los estudiantes, pues en base a la información y el conocimiento generados mediante el descubrimiento y análisis de los patrones de deserción estudiantil, se identificaron algunos factores que intervienen en dicho fenómeno, de manera que los responsables educativos podrán tomar decisiones tendientes a disminuir la deserción estudiantil en la institución. Asimismo, se beneficia al personal responsable de la toma de decisiones, pues contarán con una herramienta que les apoyará en dicha función. El presente estudio contribuye a identificar algunas causas de la deserción asociadas a los datos personales y académicos de los estudiantes, conocimiento que posibilita la planificación de acciones oportunas para prevenir ocurrencias futuras. La metodología que se utilizó, puede ser aplicada en otras instituciones de educación superior tecnológica del país, con la finalidad de entender las causas de la deserción estudiantil y tomar acciones orientadas a disminuirla. Con el presente estudio, se sientan las bases para la investigación de la deserción estudiantil mediante la aplicación de la minería de datos dentro del área de la administración educativa en los institutos superiores tecnológicos del país.

1.1. Problema

¿Cuáles son los patrones de deserción estudiantil ocultos en los datos operativos de los estudiantes que abandonaron voluntariamente sus estudios en el Instituto de Educación Superior Tecnológico “Daniel Villar” de Caraz, que serán relevantes y permitirán descubrir los factores asociados que influyen en la deserción?

1.2. Hipótesis

Hi: Los patrones de deserción estudiantil descubiertos mediante el procesamiento de los datos operativos de los estudiantes con técnicas de minería de datos educacional, permitirán descubrir algunos de los factores que influyen en la deserción estudiantil en el Instituto de Educación Superior Tecnológico Público “Daniel Villar”.

Ho: Los patrones de deserción estudiantil descubiertos mediante el procesamiento de los datos operativos de los estudiantes con técnicas de minería de datos educacional, no permitirán descubrir algunos de los factores que influyen en la deserción estudiantil en el Instituto de Educación Superior Tecnológico Público “Daniel Villar”.

1.3. Objetivos

1.3.1 General:

Descubrir patrones de deserción estudiantil e identificar los factores que influyen en la deserción relacionados con los datos operativos de los estudiantes, mediante el procesamiento de los datos operativos (personales y académicos) de los estudiantes que abandonaron voluntariamente sus estudios, siguiendo los lineamientos del proceso KDD de descubrimiento del conocimiento en bases de datos, aplicando técnicas de Minería de Datos Educacional, estableciendo un procedimiento para el estudio de la deserción estudiantil en los institutos superiores tecnológicos del país.

1.3.2 Específicos:

- Recolectar los datos secundarios para la investigación, mediante la aplicación de los procesos de selección, pre procesamiento y transformación de datos definidos en el proceso de descubrimiento del conocimiento en bases de datos (KDD), sobre los datos personales y académicos de los estudiantes, teniendo como fuente de datos secundarios las carpetas del postulante y las actas consolidadas de evaluación semestral de los estudiantes que ingresaron durante los años 2010 al 2013 y establecer un repositorio de datos para su almacenamiento electrónico.
- Utilizar técnicas de Minería de Datos Educacional para el descubrimiento de patrones de deserción estudiantil, e interpretarlos y evaluarlos para identificar los factores relacionados con los datos operativos de los estudiantes, que influyen en la deserción estudiantil.
- Establecer un procedimiento para el estudio de la deserción estudiantil en los institutos superiores tecnológicos del país, utilizando los datos personales y académicos de los estudiantes, siguiendo los lineamientos del proceso KDD de descubrimiento del conocimiento en bases de datos, aplicando técnicas de Minería de Datos Educacional.

CAPÍTULO II

MARCO METODOLÓGICO

II. MARCO METODOLÓGICO

2.1 Variables

Variable dependiente: Patrones de deserción estudiantil.

Los patrones de deserción estudiantil, son los patrones o modelos de comportamiento de aquellos estudiantes que abandonaron voluntariamente y de manera permanente sus estudios.

Variable independiente: Datos operativos de los estudiantes.

Los datos operativos de los estudiantes, son los datos personales y académicos que fueron recolectados en los documentos oficiales de la institución. Uno de estos documentos es la carpeta del postulante, que contiene el formato único de trámite y la ficha de inscripción que deben ser llenados con información personal del estudiante, la copia del documento nacional de identidad, la partida de nacimiento y el certificado de estudios secundarios, recabado al momento de la inscripción para el examen de admisión. Los otros documentos son las actas consolidadas de evaluación académica semestral, que contienen la información del rendimiento académico del estudiante cuando cursa sus estudios, que se encuentran en formato electrónico.

2.2 Operacionalización de variables

Variable	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Escala de medición
Datos operativos de los estudiantes.	Los datos operativos constituyen la gestión de la información del día a día, reflejan los procesos físicos de las organizaciones, como transacciones monetarias, movimientos de mercancías, carga de máquinas, entradas y salidas del personal, etc. (Hernández, 2005)	Datos personales y académicos de los estudiantes, recopilados por la institución en el proceso de admisión y el desarrollo de sus estudios, que luego de un proceso de selección, pre procesamiento y transformación de datos, serán almacenados electrónicamente en un repositorio de datos.	Datos personales	Distrito de nacimiento Provincia de nacimiento Departamento de nacimiento Edad al ingresar Distrito de residencia Provincia de residencia Departamento de residencia Colegio de procedencia Tipo de colegio	Nominal Nominal Nominal De razón, discreta Nominal Nominal Nominal Nominal Nominal (Nacional, Particular)

				Modalidad del colegio	Nominal (EBR, EBA)
				Distrito donde está ubicado el colegio	Nominal
				Provincia donde está ubicado el colegio	Nominal
				Departamento donde está ubicado el colegio	Nominal
				Diferencia entre el año de ingreso al instituto y el año de egreso del colegio	De razón, discreta
				Sexo	Nominal (Varón, Mujer)
				Estado civil	Nominal (Soltero, Casado)
				Trabaja	Nominal (Sí, No)
				Nivel académico	Nominal (Secundario, Superior)

			Datos académicos	Promedio final del estudiante en el colegio	De razón, discreta (0 a 20)
				Año de ingreso	De razón, discreta (2010 a 2013)
				Carrera profesional	Nominal (Enfermería técnica, Computación e informática, Secretariado ejecutivo, Mecánica automotriz)
				Ciclo de abandono	Nominal (I, II, III, IV, V, VI)
				Promedio I ciclo	De razón, discreta (0 a 20)
				Promedio II ciclo	De razón, discreta (0 a 20)
				Promedio III ciclo	De razón, discreta (0 a 20)
				Promedio IV ciclo	De razón, discreta (0 a 20)

				Promedio V ciclo	De razón, discreta (0 a 20)
				Promedio VI ciclo	De razón, discreta (0 a 20)
Patrones de deserción estudiantil.	Maneras de actuar que exhiben los estudiantes y que están relacionados con el abandono voluntario de sus estudios.	Modelos que describen el comportamiento de los estudiantes que abandonan voluntariamente sus estudios, obtenidos mediante técnicas de minería de datos.	Naturaleza	Trivial Útil	Nominal (Sí, No) Nominal (Sí, No)
			Confiabilidad	Porcentaje de confianza del modelo	De razón, continua (0 – 100 %)

2.3 Metodología

La presente investigación tuvo un enfoque empírico, toda vez que el objetivo fue descubrir los perfiles de los estudiantes desertores y sus factores causantes asociados, analizando los datos cuantitativos obtenidos de los datos operativos de los estudiantes, que fueron recolectados por el instituto mediante la carpeta del postulante y las actas consolidadas de evaluación académica semestral. En estos documentos se recolectaron cuantitativamente características intrínsecas de los estudiantes que no son susceptibles de manipulación, y constituyeron las fuentes de datos secundarios que se utilizaron para obtener los datos utilizados en la investigación. Los datos obtenidos fueron de naturaleza multivariante, y con la finalidad de obtener los patrones de deserción ocultos en ellos, para su análisis se utilizaron las modernas técnicas de minería de datos educacional. Se siguió la metodología del proceso KDD (Knowledge Discovery in Databases) de descubrimiento de conocimiento en bases de datos, compuesta de los siguientes procesos: selección y pre procesamiento de los datos, transformación de los datos, análisis de los datos con técnicas de minería de datos, e interpretación de los patrones encontrados y consolidación del conocimiento descubierto (García y Molina, 2012). Los datos operativos seleccionados para realizar el estudio fueron transcritos y sistematizados almacenándose en una base de datos, luego se prepararon y acondicionaron almacenándose en un repositorio de datos, a continuación se analizaron estos datos con técnicas de minería de datos, encontrándose los patrones de deserción que fueron analizados obteniéndose para cada uno los factores que influyen en la deserción, lo que constituye el conocimiento descubierto que puede ser utilizado para el tratamiento de la deserción en el instituto bajo estudio.

2.4 Tipo de estudio

Estudio no experimental.

Los datos personales y académicos de los estudiantes han sido recolectados por el sistema de información de la institución y por lo tanto, la variable independiente ya ha ocurrido y no puede ser manipulada.

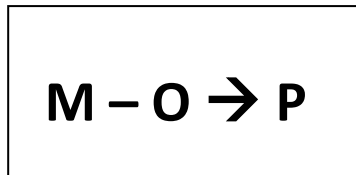
2.5 Diseño

Diseño descriptivo con propuesta.

El estudio se realizó con el objetivo de obtener conocimientos sobre la deserción estudiantil en el instituto bajo estudio, mediante el descubrimiento de patrones de deserción estudiantil que describan el comportamiento de los estudiantes que abandonan voluntariamente sus estudios y los factores asociados a éstos, estableciéndose un método de análisis que podrá ser utilizado en otras instituciones educativas para estudiar dicho fenómeno.

Los datos personales y académicos de los estudiantes fueron recolectados por el sistema de información de la institución, sin administrar ningún tratamiento.

La propuesta es la creación de un sistema que utilice el conocimiento descubierto para predecir la deserción de los estudiantes, permitiendo diseñar estrategias para su prevención y la adopción de acciones correctivas, como alternativa de solución.



Donde:

M: Población censal

O: Patrones de deserción estudiantil

P: Propuesta

2.6 Población, muestra y muestreo

La población censal estudiada, estuvo compuesta de 594 estudiantes de ambos sexos que ingresaron durante los años 2010 al 2013, y cursaron sus estudios en cuatro especialidades del Instituto de Educación Superior Tecnológico “Daniel Villar” del distrito de Caraz, provincia de Huaylas, departamento de Ancash, y que algunos de ellos han desertado. Esta población-muestra presentó un promedio de deserción estudiantil de 41.92 % durante el desarrollo de sus estudios entre los años 2010 al 2015.

2.7 Técnicas e instrumentos de recolección de datos

2.7.1 Técnicas

La técnica utilizada para la recolección de datos, fue el análisis de datos secundarios cuantitativos, que es una técnica que se refiere a la investigación con datos contruidos por otros, consistente en buscar y seleccionar entre los datos existentes aquellos que le permitan abordar el tema de interés, para producir conocimiento socialmente relevante en el actual contexto de enorme proliferación de información (Otamendi y Otero, 2009).

No se utilizaron otras técnicas como la entrevista o encuesta, por la imposibilidad de ubicar a los estudiantes que desertaron, debido a que la institución no cuenta con

información sobre su domicilio o medio de comunicación, por lo que no se recolectaron sus datos socioeconómicos.

Los datos utilizados en la presente investigación, han sido datos secundarios transcritos y sistematizados desde el sistema de información del instituto. Fueron recolectados de las carpetas del postulante y de las actas consolidadas de evaluación académica semestral de los estudiantes que ingresaron durante los años 2010 al 2013, teniendo como instrumento de recolección de datos una base de datos.

Los datos de la carpeta del postulante se encuentran consignados en papel (formulario único de trámite, solicitando su inscripción en el concurso de admisión, ficha de inscripción, certificado oficial de estudios secundarios, copia de D.N.I. del postulante y partida de nacimiento), y los datos académicos se encuentran consignadas en las actas consolidadas de evaluación académica semestral, en formato electrónico (hoja de cálculo de Microsoft Excel).

Los datos personales y académicos primarios de los estudiantes han sido recolectados por el sistema de información de la institución. Las técnicas utilizadas por la institución han sido el llenado de la ficha de recolección de datos personales y la recolección de documentos personales (carpeta del postulante), para el caso de los datos personales; y la evaluación académica y el llenado de los registros de evaluación y las actas consolidadas de evaluación académica semestral, para los datos académicos.

Los instrumentos que utilizó la institución para la recolección de los datos personales y académicos de los estudiantes, fueron la carpeta del postulante (formularios y documentos personales), y las actas de evaluación de las unidades didácticas según el ciclo de estudios, que son transcritas en las actas consolidadas de evaluación académica semestral, en formato electrónico. Estos instrumentos son documentos oficiales y han sido formulados por el Ministerio de Educación y la institución. Los datos primarios recolectados mediante estos instrumentos tienen la mayor confiabilidad pues son información oficial de la institución.

La carpeta del postulante registra los siguientes datos:

- Año del examen de admisión.
- Carrera profesional a la que postula.
- Apellidos y nombres del postulante.
- Documento nacional de identidad.
- Lugar y fecha de nacimiento.
- Lugar y dirección de residencia.

- Números telefónicos.
- Estado civil.
- Sexo.
- Edad.
- Colegio de procedencia.
- Ubicación del colegio.
- Tipo de colegio.
- Modalidad de enseñanza.
- Año de egreso de la institución educativa de nivel secundario
- Nivel académico.
- Situación laboral.
- Segunda opción de postulación.
- Notas obtenidas en los cursos del nivel secundario.

Por su parte, las actas consolidadas de evaluación académica semestral registran los siguientes datos:

- Carrera profesional.
- Tipo de gestión.
- Ubicación del instituto.
- Período y semestre académico.
- Código de matrícula, DNI, sexo y apellidos y nombres de los estudiantes.
- Unidades didácticas y créditos, correspondientes al semestre académico.
- Notas obtenidas por cada estudiante en las unidades didácticas.
- Puntaje promedio ponderado de cada estudiante en el semestre académico.
- Nombres de los docentes por unidad didáctica.

2.7.2 Instrumentos

El instrumento para la recolección de los datos secundarios fue un repositorio de datos compuesto por una base de datos creada en el sistema de gestión de bases de datos MySQL, cuya estructura diseñada para almacenar los indicadores de la variable independiente, así como las escalas de medición correspondientes, fue validada mediante criterio de jueces o expertos. El instrumento de recolección de datos validado se muestra en el anexo 2. Los datos recolectados, posteriormente fueron trasladados a una hoja de cálculo de Excel para su procesamiento. Para el proceso de selección, pre procesamiento y transformación de datos del proceso KDD, se creó un software con el IDE Visual Basic .NET, mediante el cual se ingresaron manualmente los datos de las carpetas del postulante y se

importó automáticamente los datos consignados en las actas consolidadas de evaluación en Excel. También, se calculó el promedio de las notas obtenidas en secundaria.

Los datos considerados en el repositorio de datos fueron los siguientes:

- DNI del postulante.
- Apellido paterno.
- Apellido materno.
- Nombres.
- Dirección domiciliaria.
- Distrito del domicilio.
- Provincia del domicilio.
- Departamento del domicilio.
- Teléfono.
- Especialidad.
- Fecha de nacimiento.
- Distrito de nacimiento.
- Provincia de nacimiento.
- Departamento de nacimiento.
- Sexo.
- Año de ingreso a la institución.
- Estado civil.
- Nombre del colegio.
- Tipo de colegio.
- Modalidad del colegio.
- Distrito de ubicación del colegio.
- Provincia de ubicación del colegio.
- Departamento de ubicación del colegio.
- Nivel académico del postulante.
- Año de egreso del colegio.
- Segunda opción de postulación.
- Situación laboral.
- Promedio de las notas del colegio.
- Puntaje promedio ponderado del I ciclo de estudios en el instituto.
- Puntaje promedio ponderado del II ciclo de estudios en el instituto.
- Puntaje promedio ponderado del III ciclo de estudios en el instituto.

- Puntaje promedio ponderado del IV ciclo de estudios en el instituto.
- Puntaje promedio ponderado del V ciclo de estudios en el instituto.
- Puntaje promedio ponderado del VI ciclo de estudios en el instituto.

Para el procesamiento de los datos, se transcribieron solamente aquellos datos relevantes para el objetivo del estudio, es decir datos que pudieran estar relacionados con algunos de los factores de la deserción estudiantil en la institución, y que mantuvieran el anonimato de los estudiantes por cuestiones éticas.

También, se identificó a los estudiantes que habían abandonado sus estudios en la institución, tomando como criterio para determinar la deserción, la condición de que no se hayan matriculado en dos semestres seguidos, y se determinó el ciclo en que abandonaron los estudios.

Los datos considerados en la hoja de cálculo de Excel para su procesamiento fueron los siguientes:

- Desertó.
- Ciclo de abandono.
- Distrito del domicilio.
- Provincia del domicilio.
- Departamento del domicilio.
- Especialidad.
- Fecha de nacimiento.
- Distrito de nacimiento.
- Provincia de nacimiento.
- Departamento de nacimiento.
- Sexo.
- Año de ingreso a la institución.
- Estado civil.
- Tipo de colegio.
- Modalidad del colegio.
- Distrito de ubicación del colegio.
- Provincia de ubicación del colegio.
- Departamento de ubicación del colegio.
- Nivel académico del postulante.
- Año de egreso del colegio.
- Situación laboral.

- Promedio de las notas del colegio.
- Puntaje promedio ponderado del I ciclo de estudios en el instituto.
- Puntaje promedio ponderado del II ciclo de estudios en el instituto.
- Puntaje promedio ponderado del III ciclo de estudios en el instituto.
- Puntaje promedio ponderado del IV ciclo de estudios en el instituto.
- Puntaje promedio ponderado del V ciclo de estudios en el instituto.
- Puntaje promedio ponderado del VI ciclo de estudios en el instituto.

En esta etapa se calculó la nota promedio obtenida por el estudiante en educación secundaria.

Se determinó que 9 estudiantes hicieron abandono de carrera, es decir, abandonaron su carrera inicial para postular a otra carrera dentro de la misma institución, por lo que los registros de estos estudiantes en la primera carrera a la que postularon fueron excluidos.

2.8 Métodos de análisis de datos

En primer lugar, se utilizaron técnicas de análisis descriptivo para el análisis de los datos correspondientes a las variables independientes, con la finalidad de determinar sus frecuencias y proporcionar una idea general de sus características, y también para ayudar a determinar si eran relevantes para el estudio, utilizando el software Excel. Luego se utilizaron las técnicas de minería de datos educacional para el análisis de los datos, con la finalidad de descubrir los patrones de deserción estudiantil de los estudiantes de la institución, específicamente técnicas de clasificación mediante el algoritmo de árbol de decisión C5.0 y sus correspondientes reglas de asociación, utilizando el software SPSS Modeler. Finalmente, se procedió al análisis de los modelos obtenidos, para identificar los patrones de deserción relevantes e identificar los factores de deserción asociados a ellos.

La minería de datos es parte de las técnicas estadísticas de análisis multivariante (Aluja, 2001), técnicas que se utilizaron porque los patrones de deserción estudiantil dependen del valor de varias variables. Estas técnicas de análisis para grandes volúmenes de datos, se aplicaron sobre los datos almacenados en el repositorio de datos, para encontrar patrones de comportamiento que se relacionen con la deserción estudiantil. La minería de datos es una tecnología que extrae información útil de grandes volúmenes de datos, y que viene transformando los métodos de análisis un campo de investigación tras otro (Baker y Siemens, 2013).

Los métodos de predicción estadística univariante como la regresión y el análisis discriminante, implican ajustar un modelo a los datos, evaluar el ajuste y estimar los parámetros, que más tarde se utilizan en una ecuación de predicción. Los modelos de

árboles de decisión realizan la partición de un conjunto de datos, basándose en las relaciones entre las variables predictoras y la variable objetivo (resultado). El árbol creado indica las variables predictoras relacionadas más estrechamente con la variable objetivo (Veitch, 2004).

Los datos que fueron materia de análisis, se almacenaron en una hoja de cálculo de Excel para su acondicionamiento y posterior procesamiento. En la terminología utilizada en la minería de datos, se denomina variable objetivo a la variable dependiente, y predictor a la variable independiente. Asimismo, se denominan variables continuas a las variables numéricas y de marca a las variables dicotómicas.

Al acondicionar los datos para su procesamiento, algunas variables continuas fueron transformadas a nominales o de intervalo, por ser más conveniente para el análisis con los algoritmos utilizados. También se calcularon los valores de los datos derivados como la edad y el tiempo de egresado del colegio al momento de ingresar.

Se utilizó el software SPSS Modeler como software de minería de datos para realizar el análisis de datos. El repositorio de datos ha estado compuesto de las siguientes variables:

- deserto. Variable objetivo dicotómica, que indica si el estudiante desertó o no.
- ciclo_ab. Ciclo de abandono. Predictor, nominal.
- dist_dom. Distrito donde domicilia. Predictor, nominal.
- prov_dom. Provincia donde domicilia. Predictor, nominal.
- especialidad. Especialidad que estudió. Predictor, nominal.
- edad. Edad del estudiante al ingresar. Predictor, continuo.
- dist_nac. Distrito de nacimiento. Predictor, nominal.
- prov_nac. Provincia de nacimiento. Predictor, nominal.
- dep_nac. Departamento de nacimiento. Predictor, nominal.
- sexo. Sexo del estudiante. Predictor, dicotómico.
- a_ing. Año de ingreso a la institución. Predictor, nominal.
- e_civil. Estado civil del estudiante. Predictor, dicotómico.
- tipo_col. Tipo de colegio. Predictor, nominal.
- mod_col. Modalidad del colegio. Predictor, nominal.
- dist_col. Distrito de ubicación del colegio. Predictor, nominal.
- prov_col. Provincia de ubicación del colegio. Predictor, nominal.
- dep_col. Departamento de ubicación del colegio. Predictor, nominal.
- t_egreso. Tiempo de egresado del colegio. Predictor, continuo.
- trabaja. Situación laboral del estudiante. Predictor, dicotómico.

- prom_sec. Promedio de las notas del colegio. Predictor, nominal.
- prom1. Puntaje promedio ponderado del I ciclo de estudios en el instituto. Predictor, nominal.
- prom2. Puntaje promedio ponderado del II ciclo de estudios en el instituto. Predictor, nominal.
- prom3. Puntaje promedio ponderado del III ciclo de estudios en el instituto. Predictor, nominal.
- prom4. Puntaje promedio ponderado del IV ciclo de estudios en el instituto. Predictor, nominal.
- prom5. Puntaje promedio ponderado del V ciclo de estudios en el instituto. Predictor, nominal.
- prom6. Puntaje promedio ponderado del VI ciclo de estudios en el instituto. Predictor, nominal.

De este conjunto de datos, en muchos casos se derivaron subconjuntos de datos para realizar los análisis, por ser necesario clasificar los datos relevantes o relacionados para determinado análisis, y excluir aquellos que eran irrelevantes o no necesarios.

También, se hicieron análisis de la importancia de los predictores relacionados con los datos sometidos a estudio, para determinar la importancia de cada predictor en relación a la variable objetivo. Se realizaron estimaciones con los indicadores de Pearson, Razón de verosimilitud, V de Cramer y Lambda, disponibles en el software SPSS Modeler. Se seleccionaron aquellas variables que tenían una mayor frecuencia de aparición en los resultados obtenidos en cada prueba.

Por otro lado, teniendo en cuenta que los algoritmos de aprendizaje automático tienden a concentrarse más en el aprendizaje de los datos mayoritarios, es necesario equilibrar los datos de manera que exista más o menos la misma cantidad de registros que generen los dos valores de la variable objetivo, caso contrario se puede producir un sesgo. En el repositorio de datos utilizado los datos están desbalanceados, por lo que se utilizó un nodo equilibrar con un factor de multiplicación adecuado para cada caso, para que el sesgo que pudiera producirse fuera mínimo o nulo. Debido a que el nodo equilibrar lo que hace es crear artificialmente más registros de acuerdo al factor de balanceo, en muchos resultados aparecieron una cantidad mayor de registros que los que contenía inicialmente el repositorio. Estos registros creados son descartados automáticamente luego del análisis, quedando sin cambios el repositorio de datos.

2.9 Aspectos éticos

El tema de investigación y el contenido del presente documento son originales, por lo tanto no contiene plagios de ninguna índole.

El autor fue autorizado por la dirección del instituto, para la utilización en el presente estudio de los datos registrados en las carpetas de los postulantes así como de las actas consolidadas de evaluación académica semestral, de los estudiantes que ingresaron en los años 2010 al 2013.

Los datos personales y académicos de los estudiantes utilizados han sido extraídos de documentos oficiales como las carpetas de los postulantes y las actas consolidadas de evaluación académica semestral del Instituto de Educación Superior tecnológico Público “Daniel Villar” de Caraz, correspondientes a los estudiantes que ingresaron en el período 2010-2013, y fueron utilizados expresamente para la realización del presente estudio.

CAPÍTULO III

RESULTADOS

III. RESULTADOS

Los siguientes gráficos muestran la información analítica obtenida del procesamiento de los datos contenidos en el repositorio de datos construido en una hoja de cálculo Excel, que registra los datos de 594 estudiantes. Los datos están organizados en registros, correspondiendo un registro a cada estudiante. Algunos de los registros de datos están incompletos debido a que las carpetas de postulante no tenían la documentación completa o los postulantes no llenaron adecuadamente algunos datos. Sin embargo, los datos académicos obtenidos de las actas consolidadas de evaluación académica semestral están completos, permitiendo determinar la información referente a quienes desertaron, su especialidad, sexo y ciclo de deserción.

La tabla N° 1 muestra la frecuencia y el porcentaje de deserción estudiantil en el instituto bajo estudio.

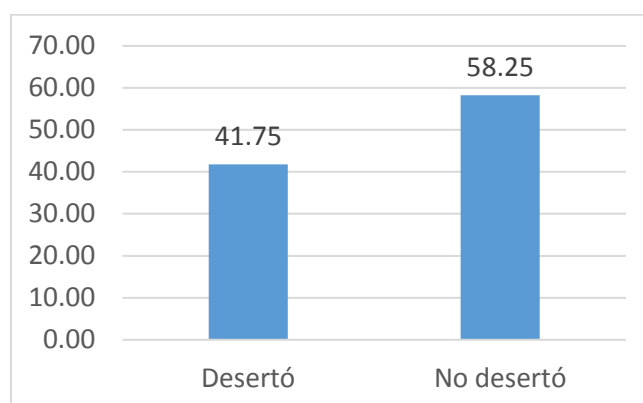
TABLA N° 1
DESERCIÓN ESTUDIANTIL EN EL INSTITUTO BAJO ESTUDIO

Condición	Frecuencia	Porcentaje
Desertó	248	41.75
No desertó	346	58.25
Total	594	100.00

Fuente: Actas consolidadas de evaluación académica semestral.

Como se puede apreciar, la deserción estudiantil en el instituto es alta, con 248 estudiantes desertores que representan un 41.75 % de deserción, de un total de 594 estudiantes que ingresaron a la institución durante los años 2010 al 2013.

GRÁFICO N° 1
PORCENTAJE DE DESERCIÓN ESTUDIANTIL EN EL INSTITUTO



Fuente: tabla N° 1.

La tabla N° 2 muestra la frecuencia y el porcentaje de deserción estudiantil según la especialidad que cursan.

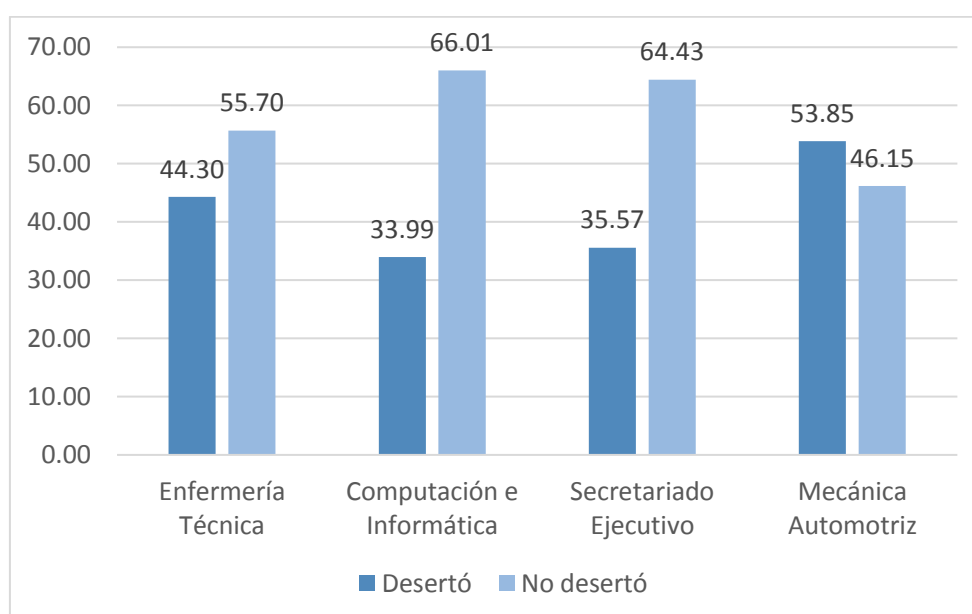
TABLA N° 2
DESERCIÓN POR ESPECIALIDAD

ESPECIALIDAD	DESERTARON				TOTAL POR ESPECIALIDAD
	SI		NO		
	Frecuencia	%	Frecuencia	%	
Enfermería Técnica	66	44.30	83	55.70	149
Computación e Informática	52	33.99	101	66.01	153
Secretariado Ejecutivo	53	35.57	96	64.43	149
Mecánica Automotriz	77	53.85	66	46.15	143
TOTAL ESTUDIANTES					594

Fuente: Actas consolidadas de evaluación académica semestral.

La especialidad de Mecánica Automotriz es la que presenta la mayor tasa de deserción con un 53.85 % (77 estudiantes), seguido por Enfermería Técnica con un 44.30 % (66 estudiantes). Las especialidades de Secretariado Ejecutivo y Computación e Informática presentan una menor tasa de deserción con un 35.57 % (53 estudiantes) y 33.99 % (52 estudiantes) respectivamente.

GRÁFICO N° 2
PORCENTAJE DE DESERCIÓN POR ESPECIALIDAD



Fuente: tabla N° 2.

En la tabla N° 3 se muestra la frecuencia y porcentaje de deserción estudiantil por ciclo de estudios.

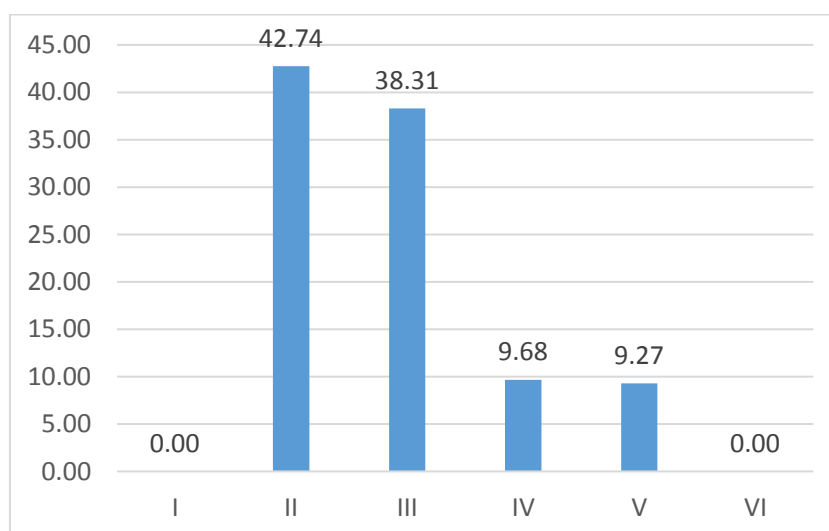
TABLA N° 3
DESERCIÓN ESTUDIANTIL POR CICLO DE ESTUDIOS

Ciclo	Desertaron	
	Frecuencia	%
I	0	0.00
II	106	42.74
III	95	38.31
IV	24	9.68
V	23	9.27
VI	0	0.00
Total:	248	100.00

Fuente: Actas consolidadas de evaluación académica semestral.

En la tabla 3, se puede observar que la mayor tasa de deserción se produce en el segundo ciclo, con un 42.74 % (106 estudiantes) del total de desertores, seguido del tercer ciclo con un 38.31 % (95 estudiantes), disminuyendo notoriamente en el cuarto ciclo con 9.68 % (24 estudiantes) y el quinto ciclo con 9.27 % (23 estudiantes). Se consideró como ciclo de deserción, aquel en el que el estudiante dejó de matricularse y no lo volvió a hacer. La deserción es 0 % para el primer ciclo, lo cual era de esperarse porque todos los ingresantes se matriculan en el primer ciclo, y para el sexto ciclo también es 0 %, lo que indica que todos los estudiantes matriculados en el quinto ciclo lo hacen también en el sexto ciclo y culminan sus estudios.

GRÁFICO N° 3
PORCENTAJE DE DESERCIÓN POR CICLO DE ESTUDIOS



Fuente: tabla N° 3.

En la tabla N° 4 se muestra la frecuencia y el porcentaje de estudiantes por género:

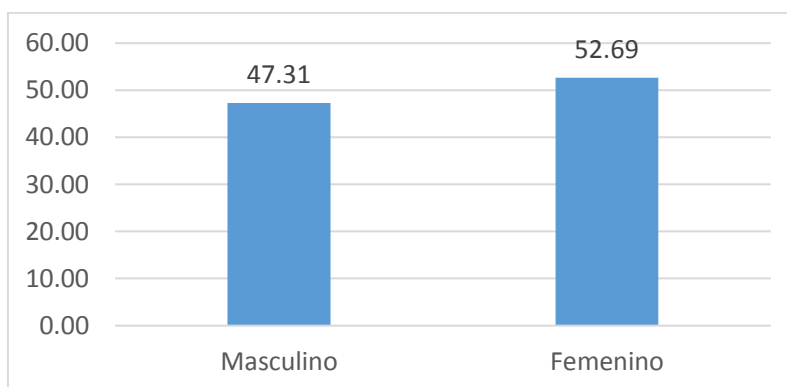
TABLA N° 4
ESTUDIANTES POR GÉNERO

Género	Frecuencia	%
Masculino	281	47.31
Femenino	313	52.69
Total:	594	100.00

Fuente: Actas consolidadas de evaluación académica semestral.

Se observa que en el instituto hay un porcentaje ligeramente mayor de estudiantes de sexo femenino.

GRÁFICO N° 4
PORCENTAJE DE ESTUDIANTES POR GÉNERO



Fuente: tabla N° 4.

En la tabla N° 5 se muestra la frecuencia y el porcentaje de estudiantes por género y especialidad.

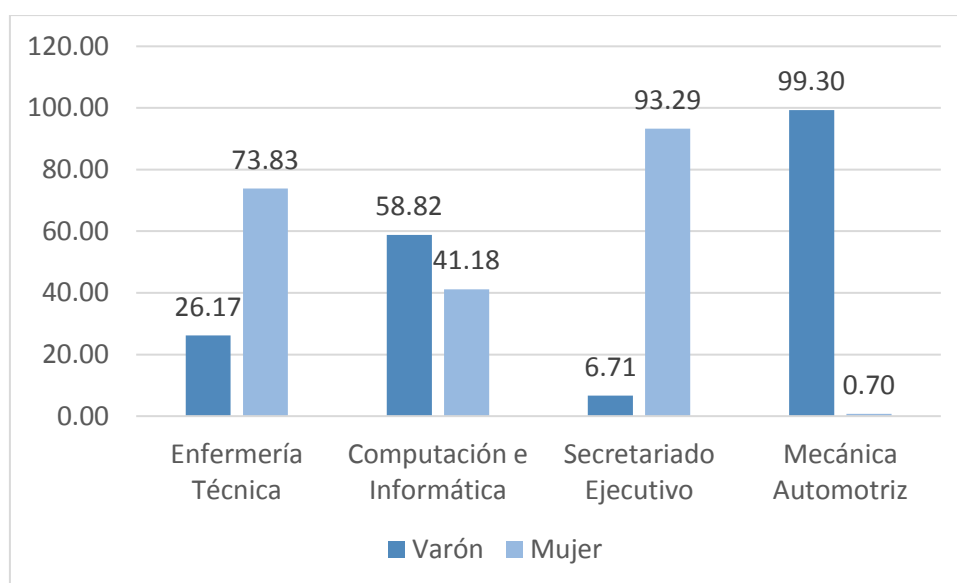
TABLA N° 5
ESTUDIANTES POR GÉNERO Y ESPECIALIDAD

ESPECIALIDAD	Varón		Mujer		TOTAL POR ESPECIALIDAD
	Frecuencia	%	Frecuencia	%	
Enfermería Técnica	39	26.17	110	73.83	149
Computación e Informática	90	58.82	63	41.18	153
Secretariado Ejecutivo	10	6.71	139	93.29	149
Mecánica Automotriz	142	99.30	1	0.70	143
TOTAL ESTUDIANTES					594

Fuente: Actas consolidadas de evaluación académica semestral.

Se puede apreciar que en algunas especialidades hay un gran desequilibrio en el género de los estudiantes. Así tenemos que en la especialidad de Mecánica Automotriz, el 99.30 % (142) de los estudiantes han sido varones, y solo ha habido una estudiante mujer. Asimismo, en la especialidad de Secretariado Ejecutivo el 93.29 % (139) de los estudiantes han sido mujeres, y solo un 6.71 % (10) de los estudiantes han sido varones. En la especialidad de Enfermería Técnica, también hay un marcado desequilibrio, pues 73.83 % (110 estudiantes) han sido mujeres y solo el 26.17 % (39 estudiantes) varones. En la especialidad de Computación e Informática hay un mayor equilibrio, con el 58.82 % (90) de varones y el 41.18 % (63) de mujeres.

GRÁFICO N° 5
PORCENTAJE DE ESTUDIANTES POR GÉNERO Y ESPECIALIDAD



Fuente: tabla N° 5.

En la tabla N° 6 se muestra la frecuencia y el porcentaje de deserción estudiantil por género.

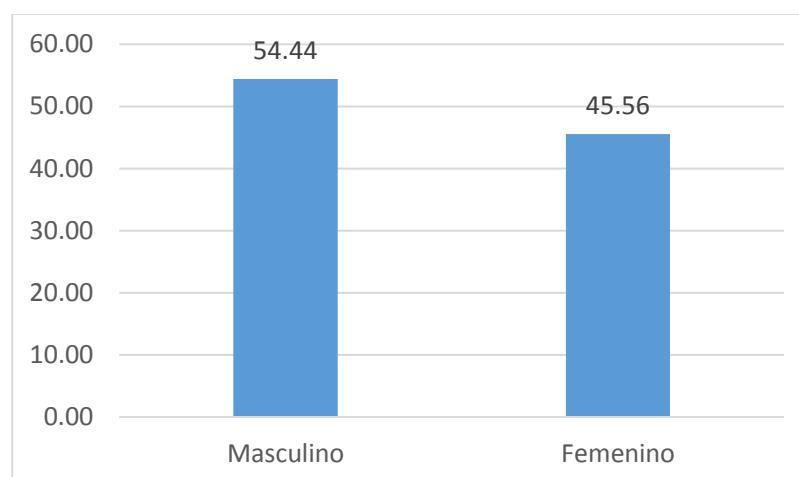
TABLA N° 6
DESERCIÓN ESTUDIANTIL POR GÉNERO

Género	Desertaron	
	Frecuencia	%
Masculino	135	54.44
Femenino	113	45.56
Total:	248	100.00

Fuente: Actas consolidadas de evaluación académica semestral.

En la tabla 6 se puede observar que el mayor porcentaje de deserción estudiantil corresponde a los estudiantes de sexo masculino, a pesar de que son el género minoritario a nivel de estudiantes. Esto se puede explicar porque según la tabla 2, la mayor tasa de deserción corresponde a la especialidad de Mecánica Automotriz, en la cual según la tabla 5, el 99.30 % (142 estudiantes) son varones.

GRÁFICO Nº 6
PORCENTAJE DE DESERCIÓN ESTUDIANTIL POR GÉNERO



Fuente: tabla Nº 6.

Hasta este punto, se han analizado parte de los datos contenidos en las actas consolidadas de evaluación semestral. Para realizar el análisis de datos sobre otros aspectos relacionados con la deserción estudiantil, es necesario utilizar los datos obtenidos de las carpetas de los postulantes, por lo que se verificó la integridad de los datos obtenidos de estas carpetas, ya que en el proceso de selección, pre procesamiento y transformación de datos se observó que algunas carpetas estaban incompletas, ya sea porque se extraviaron las carpetas o porque se habían sustraído algunos documentos. Para ello, los registros se sometieron a una verificación de contenido para determinar si están completos y si podrían utilizarse en el estudio. Se identificaron 69 registros con datos incompletos, que fueron excluidos para realizar el resto de los análisis.

Se excluyeron un total de 69 registros, por lo que en adelante, los análisis se realizarán sobre un universo de 525 estudiantes.

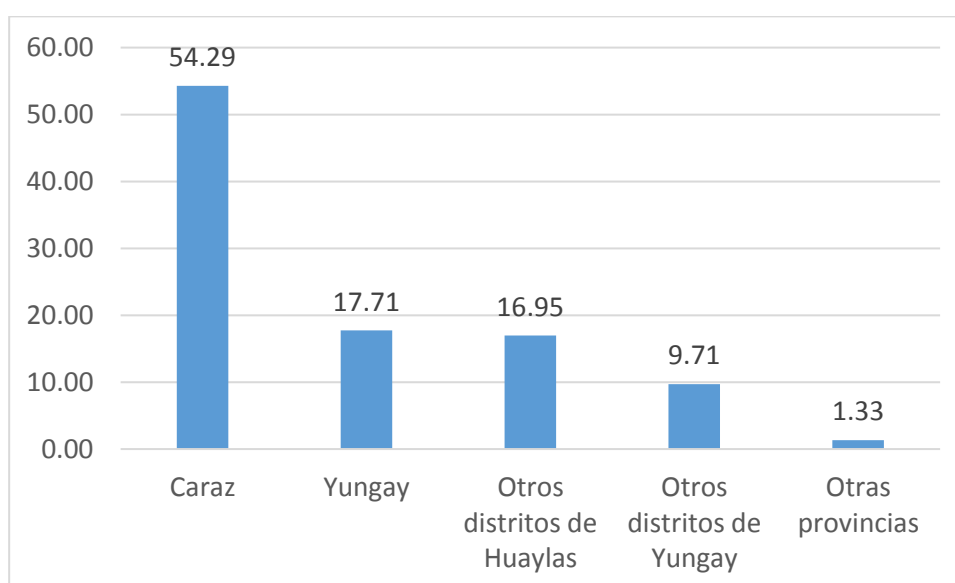
En la tabla Nº 7, se muestra la frecuencia y el porcentaje de estudiantes por distrito donde domicilian, en el cual se han considerado los distritos de mayor frecuencia en forma individual, y el resto de distritos en forma agrupada por la provincia a la que pertenecen.

TABLA Nº 7**ESTUDIANTES POR DISTRITO DONDE DOMICILIAN**

Distrito	Frecuencia	%
Caraz	285	54.29
Yungay	93	17.71
Otros distritos de Huaylas	89	16.95
Otros distritos de Yungay	51	9.71
Otras provincias	7	1.33
Total	525	100.00

Fuente: Actas consolidadas de evaluación académica semestral y carpetas del postulante.

Se aprecia que el distrito con mayor procedencia de estudiantes es Caraz con 54.29 % (285 estudiantes), y en segundo lugar se encuentra el distrito de Yungay con 17.71 % (93 estudiantes). El distrito de Caraz pertenece a la provincia de Huaylas, y conjuntamente con el resto de distritos de la provincia totalizan 374 estudiantes, que representan el 71.24 % del total. El distrito de Yungay pertenece a la provincia del mismo nombre, y conjuntamente con el resto de distritos de la provincia totalizan 144 estudiantes, que representan el 27.43 %. Solo el 1.33 % (7) de los estudiantes proceden de otras provincias, lo que indica que el ámbito de influencia del instituto comprende las provincias de Huaylas y Yungay.

GRÁFICO Nº 7**PORCENTAJE DE ESTUDIANTES POR DISTRITO DONDE DOMICILIAN**

Fuente: tabla Nº 7.

En la tabla N° 8 se muestra la frecuencia y el porcentaje de deserción estudiantil por el distrito donde domicilian, en relación con el total de estudiantes del correspondiente distrito.

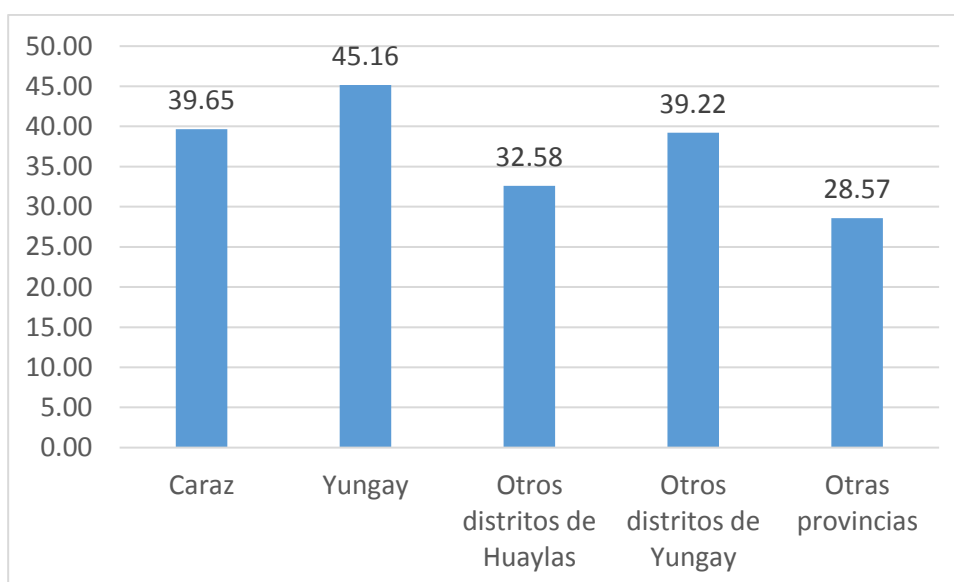
TABLA N° 8
DESERCIÓN ESTUDIANTIL POR DISTRITO DONDE DOMICILIAN

Distrito	Desertó		Porcentaje de deserción
	Si	No	
Caraz	113	172	39.65
Yungay	42	51	45.16
Otros distritos de Huaylas	29	60	32.58
Otros distritos de Yungay	20	31	39.22
Otras provincias	2	5	28.57

Fuente: Actas consolidadas de evaluación académica semestral y carpetas del postulante.

Como se puede ver en la tabla N° 8, el mayor porcentaje de deserción de estudiantes del distrito corresponde al distrito de Yungay con el 45.16 % (42 estudiantes), seguido de la provincia de Huaylas con el 39.65 % (113 estudiantes).

GRÁFICO N° 8
PORCENTAJE DE DESERCIÓN ESTUDIANTIL POR DISTRITO DONDE DOMICILIAN



Fuente: tabla N° 8.

En la tabla N° 9 se muestra la frecuencia y el porcentaje de estudiantes según el promedio que obtuvieron en el colegio:

TABLA N° 9
ESTUDIANTES SEGÚN EL PROMEDIO QUE OBTUVIERON EN EL COLEGIO

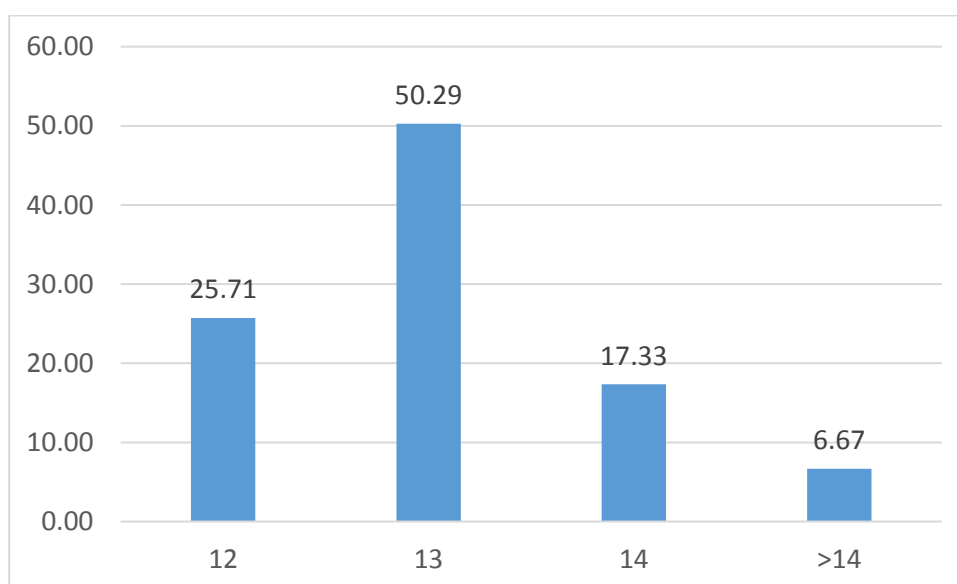
Promedio del colegio	Frecuencia	Porcentaje
12	135	25.71
13	264	50.29
14	91	17.33
>14	35	6.67
Total	525	100.00

Fuente: Actas consolidadas de evaluación académica semestral y carpetas del postulante.

En la tabla 9 se puede apreciar que los estudiantes que obtuvieron un promedio de 13 en el colegio, constituyen la clase mayoritaria con un 50.29 % (264 estudiantes), seguidos de aquellos que obtuvieron un promedio de 12 con un 25.71 % (135 estudiantes). Solo 91 estudiantes obtuvieron un promedio de 14, que constituyen el 17.33 %, y escasamente 35 estudiantes obtuvieron un promedio mayor que 14, que constituyen el 6.67 %.

De estos datos se puede deducir que el nivel académico de los estudiantes que ingresaron al instituto, es bastante modesto.

GRÁFICO N° 9
PORCENTAJE DE ESTUDIANTES SEGÚN EL PROMEDIO QUE OBTUVIERON EN EL COLEGIO

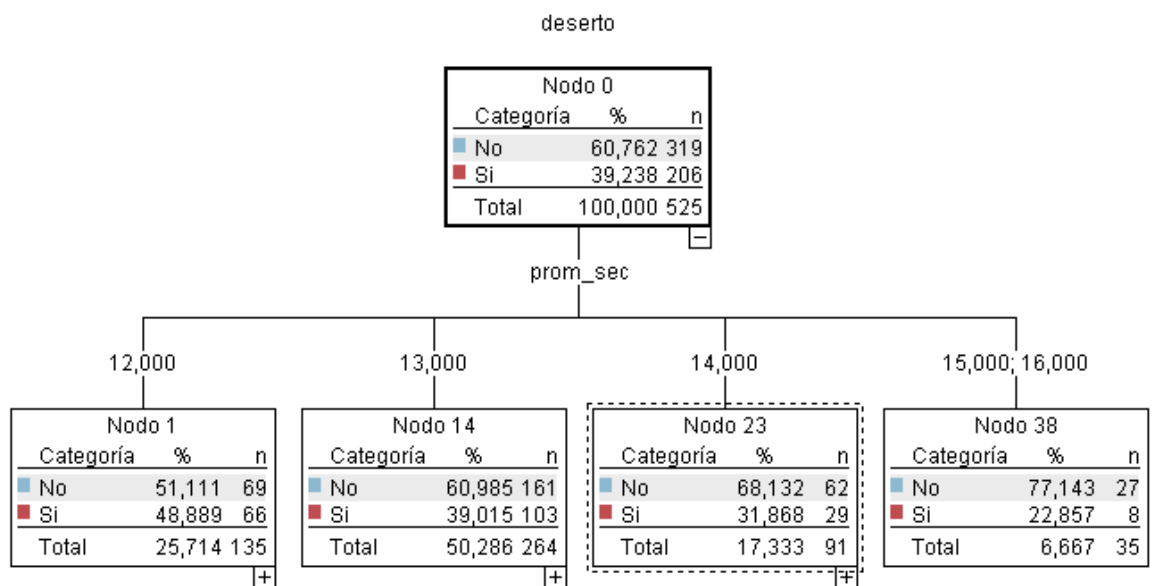


Fuente: tabla N° 9.

Para realizar el análisis multivariado de los datos y descubrir patrones de deserción de los estudiantes, se utilizaron técnicas de minería de datos, considerándose para el presente estudio, solamente los patrones más relevantes.

Se aplicó el algoritmo C5.0 de minería de datos sobre los datos personales de los estudiantes, obteniéndose el árbol de decisión de la figura 2, mostrado hasta el segundo nivel.

FIGURA Nº 2
ÁRBOL DE DECISIÓN PARA EL PROMEDIO DE SECUNDARIA



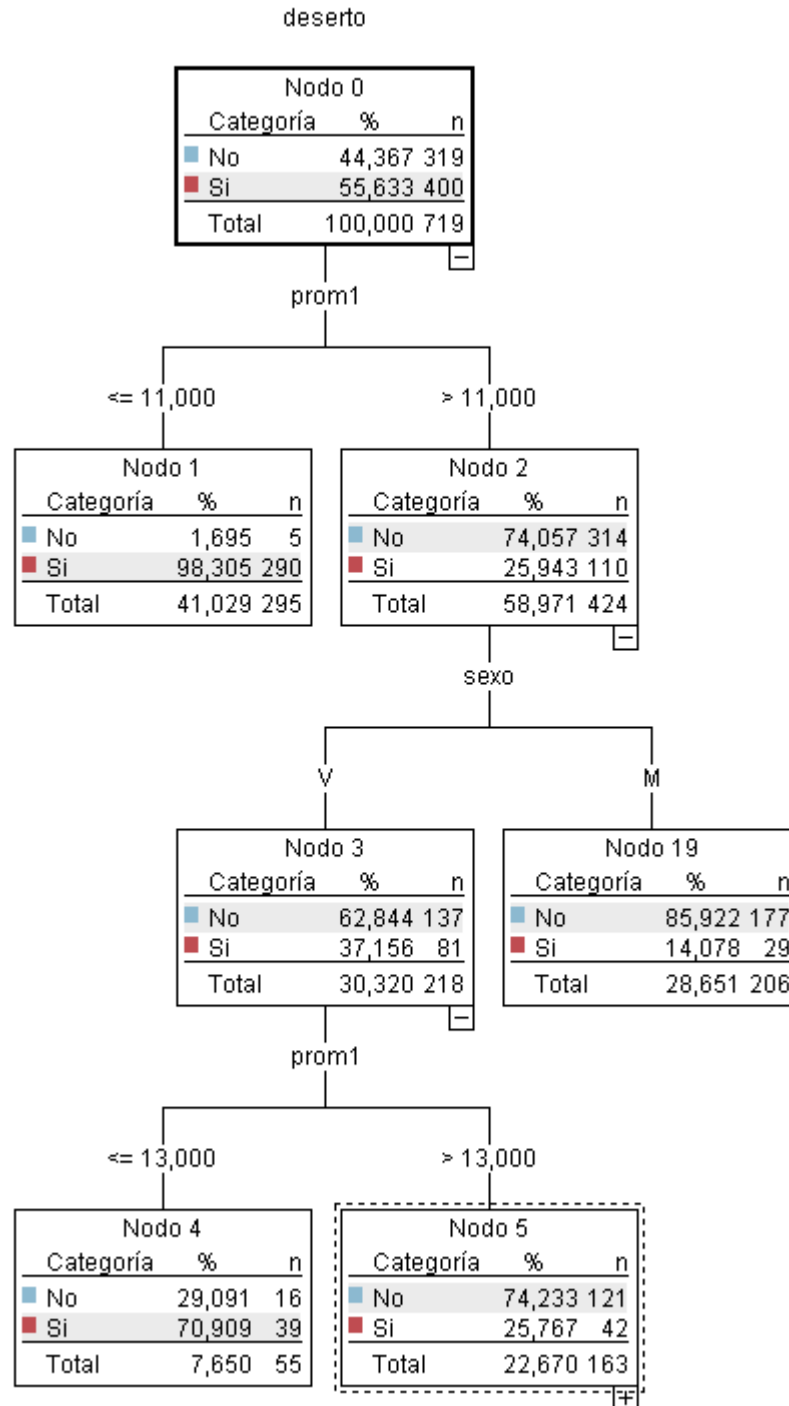
Fuente: Software SPSS Modeler

En la figura 2, se puede deducir que el predictor promedio de secundaria sin duda influye en la tasa de deserción de los estudiantes en cada categoría, disminuyendo el porcentaje de deserción conforme aumenta el promedio de secundaria obtenido. Así tenemos que para un promedio de secundaria de 12, la tasa de deserción es de 48.889 % (66 estudiantes) de un total de 135 estudiantes. Para un promedio de secundaria de 13, la tasa de deserción es de 39.015 % (103 estudiantes) de un total de 264 estudiantes. Para un promedio de secundaria de 14, la tasa de deserción es de 31.868 % (29 estudiantes) de un total de 91 estudiantes, y para un promedio de secundaria de 15 y 16, la tasa de deserción es de 22.857 % (8 estudiantes) de un total de 35 estudiantes.

Se aplicó el algoritmo C5.0 de minería de datos sobre los datos personales y académicos de los estudiantes, subconjunto correspondiente a los registros relacionados con la deserción en el segundo ciclo, el cual fue sometido al proceso de balanceo y determinación de los predictores más importantes, obteniéndose el árbol de decisión de la figura 3, mostrado hasta el cuarto nivel.

FIGURA N° 3

ÁRBOL DE DECISIÓN PARA LA DESERCIÓN EN EL SEGUNDO CICLO



Fuente: Software SPSS Modeler

Como se puede apreciar en la figura 3, el predictor que es determinante para la deserción en el segundo ciclo, es el promedio del primer ciclo. Cuando el estudiante obtiene un promedio igual o menor a 11, deserta el 98.3 % (290 estudiantes) de un total de 295 estudiantes. Si el estudiante obtiene un promedio mayor a 11, deserta el 25.94 % (110 estudiantes) de un total de 424 estudiantes. De este grupo de 424 estudiantes, si son mujeres, deserta el 14.08 % (29 estudiantes) de un total de 206 estudiantes. Si son varones, deserta el 37.16 % (81 estudiantes) de un total de 218 estudiantes. A su vez, de estos 218 estudiantes, si la nota del primer ciclo es menor o igual a 13, deserta el 70.91 % (39 estudiantes) de un total de 55 estudiantes, y si la nota del primer ciclo es mayor a 13, deserta el 25.77 % (42 estudiantes) de un total de 163 estudiantes. Hay que recordar que la cantidad de estudiantes desertores está balanceada con el fin de evitar sesgos en los análisis y resultados, por lo que las cantidades reales de estudiantes desertores en los resultados es menor.

Las reglas de asociación obtenidas para la deserción, asociadas al árbol de decisión de la figura 3, son las siguientes:

Reglas para Si - contiene 2 regla(s)

Regla 1 para Si (295; 0,983)

si $\text{prom1} \leq 11$

entonces Si

Regla 2 para Si (55; 0,709)

si $\text{prom1} > 11$

y $\text{sexo} = V$

y $\text{prom1} \leq 13$

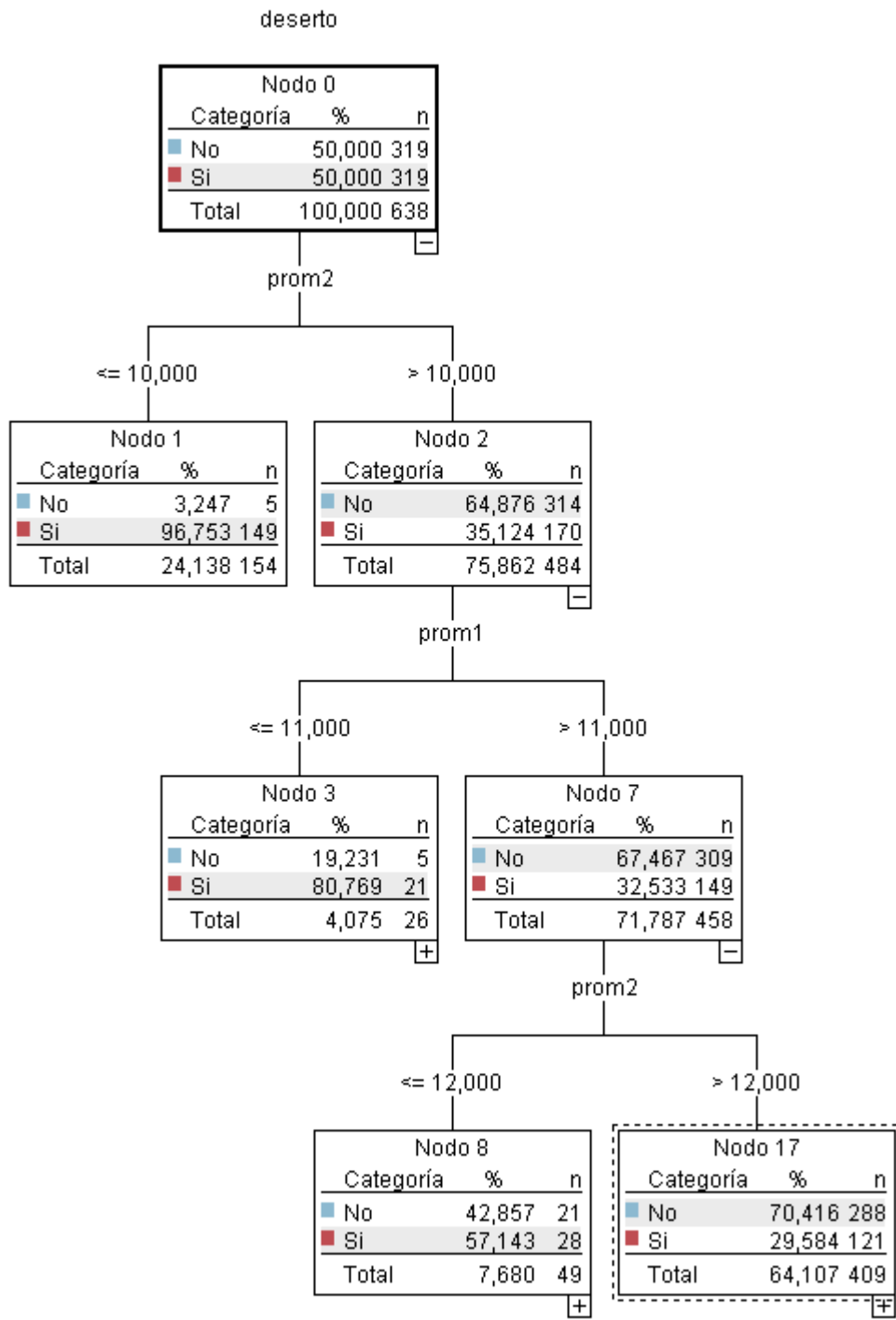
entonces Si

Como se puede advertir, la regla 1 se aplica a los estudiantes que obtuvieron un promedio en el primer ciclo menor o igual a 11, con un nivel de confianza de 0.983 y aplicable a 295 registros. La regla 2 se aplica a los estudiantes que obtuvieron un promedio en el primer ciclo entre 12 y 13 y son varones, con un nivel de confianza de 0.709 y aplicable a 55 registros.

Se aplicó el algoritmo C5.0 de minería de datos sobre los datos personales y académicos de los estudiantes, subconjunto correspondiente a los registros relacionados con la deserción en el tercer ciclo, el cual fue sometido al proceso de balanceo y determinación de los predictores más importantes, obteniéndose el árbol de decisión de la figura 4, mostrado hasta el cuarto nivel.

FIGURA N° 4

ÁRBOL DE DECISIÓN PARA LA DESERCIÓN EN EL TERCER CICLO



Fuente: Software SPSS Modeler

Como se puede apreciar en la figura 4, el predictor que es determinante para la deserción en el tercer ciclo, es el promedio del segundo ciclo, y en segunda instancia, el

promedio del primer ciclo. Cuando el estudiante obtiene un promedio menor o igual a 10, deserta el 96.5 % (149 estudiantes) de un total de 154 estudiantes. Si el estudiante obtiene un promedio mayor a 10, deserta el 35.12 % (170 estudiantes) de un total de 484 estudiantes. De este grupo de 484 estudiantes, si en el primer ciclo obtuvo un promedio menor o igual a 11, deserta el 80.77 % (21 estudiantes) de un total de 26 estudiantes. Si en el primer ciclo obtuvo un promedio mayor a 11, deserta el 32.53 % (149 estudiantes) de un total de 458 estudiantes. A su vez, de estos 458 estudiantes, si la nota del segundo ciclo es menor o igual a 12, deserta el 57.14 % (28 estudiantes) de un total de 49 estudiantes, y si la nota del segundo ciclo es mayor a 12, deserta el 29.58 % (121 estudiantes) de un total de 409 estudiantes. Hay que recordar que la cantidad de estudiantes desertores está balanceada con el fin de evitar sesgos en los análisis y resultados, por lo que las cantidades reales de estudiantes desertores en los resultados es menor.

Las reglas de asociación obtenidas para la deserción, asociadas al árbol de decisión de la figura 4, son las siguientes:

Reglas para Si - contiene 3 regla(s)

Regla 1 para Si (154; 0,968)

si $\text{prom2} \leq 10$

entonces Si

Regla 2 para Si (24; 0,875)

si $\text{prom2} > 10$

y $\text{prom1} \leq 11$

y $\text{prov_dom} = \text{Huaylas}$

entonces Si

Regla 3 para Si (23; 0,87)

si $\text{prom2} > 10$

y $\text{prom1} > 11$

y $\text{prom2} \leq 12$

y $\text{sexo} = \text{M}$

y $\text{dist_nac} \in [\text{"Otros distritos de Huaylas"} \text{"Yungay"}]$

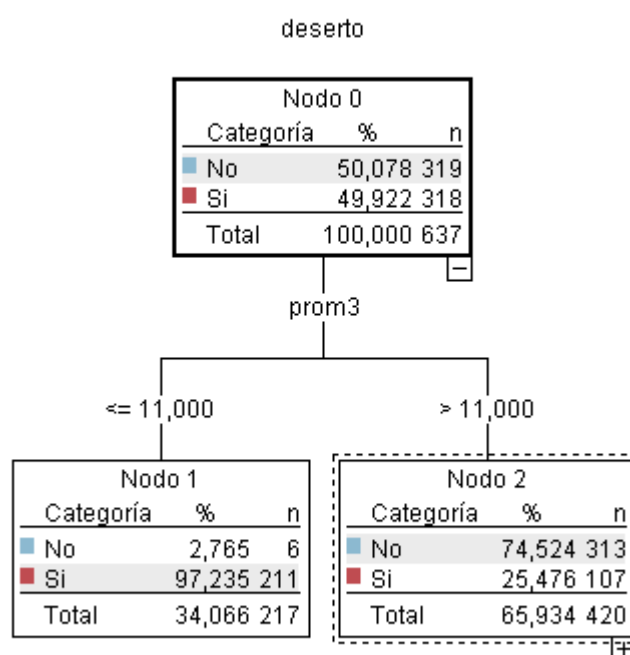
entonces Si

La regla 1 se aplica a los estudiantes que obtuvieron un promedio en el segundo ciclo menor o igual a 10, con un nivel de confianza de 0.968 y aplicable a 154 registros. La regla 2 se aplica a los estudiantes que obtuvieron un promedio en el segundo ciclo mayor a 10, un promedio en el primer ciclo menor o igual a 11 y domicilian en la provincia de Huaylas, con un nivel de confianza de 0.875 y aplicable a 24 registros. La regla 3 se aplica a los estudiantes que

obtuvieron un promedio en el segundo ciclo de 11 o 12, un promedio en el primer ciclo mayor a 11 y nacieron en un distrito de la provincia de Huaylas que no sean Caraz, Pueblo Libre o Santa Cruz, o nacieron en el distrito de Yungay, con un nivel de confianza de 0.87 y aplicable a 23 registros.

Se aplicó el algoritmo C5.0 de minería de datos sobre los datos personales y académicos de los estudiantes, subconjunto correspondiente a los registros relacionados con la deserción en el cuarto ciclo, el cual fue sometido al proceso de balanceo y determinación de los predictores más importantes, obteniéndose el árbol de decisión de la figura 5, mostrado hasta el segundo nivel.

FIGURA Nº 5
ÁRBOL DE DECISIÓN PARA LA DESERCIÓN EN EL CUARTO CICLO



Fuente: Software SPSS Modeler

Como se puede apreciar en la figura 5, el predictor que es determinante para la deserción en el cuarto ciclo, es el promedio del tercer ciclo. Cuando el estudiante obtiene un promedio menor o igual a 11, deserta el 97.24 % (211 estudiantes) de un total de 217 estudiantes. Si el estudiante obtiene un promedio mayor a 11, deserta el 25.48 % (107 estudiantes) de un total de 420 estudiantes. Hay que recordar que la cantidad de estudiantes desertores está balanceada con el fin de evitar sesgos en los análisis y resultados, por lo que las cantidades reales de estudiantes desertores en los resultados es menor.

Las reglas de asociación obtenidas para la deserción, asociadas al árbol de decisión de la figura 5, son las siguientes:

Reglas para Si - contiene 1 regla(s)

Regla 1 para Si (217; 0,972)

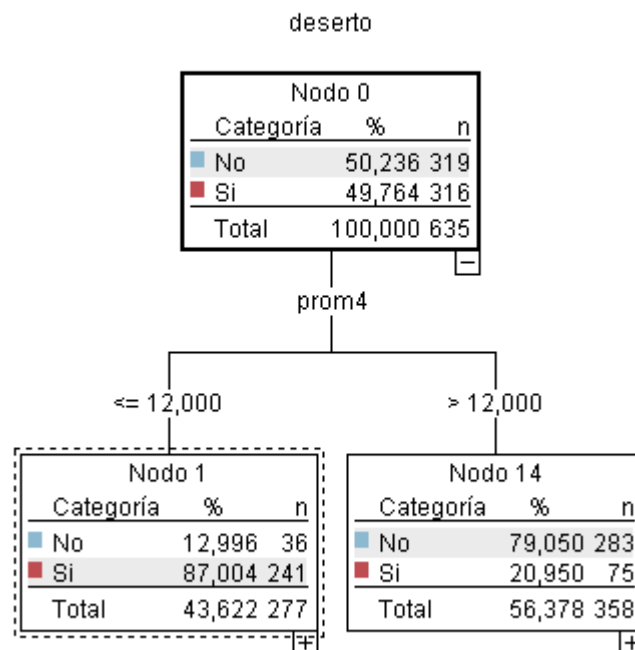
si prom3 <= 11

entonces Si

La regla 1 se aplica a los estudiantes que obtuvieron en el tercer ciclo un promedio menor o igual a 11, con un nivel de confianza de 0.972 y aplicable a 217 registros.

Se aplicó el algoritmo C5.0 de minería de datos sobre los datos personales y académicos de los estudiantes, subconjunto correspondiente a los registros relacionados con la deserción en el quinto ciclo, el cual fue sometido al proceso de balanceo y determinación de los predictores más importantes, obteniéndose el árbol de decisión de la figura 6, mostrado hasta el segundo nivel.

FIGURA Nº 6
ÁRBOL DE DECISIÓN PARA LA DESERCIÓN EN EL QUINTO CICLO



Fuente: Software SPSS Modeler

Como se puede apreciar en la figura 6, el predictor que es determinante para la deserción en el cuarto ciclo, es el promedio del cuarto ciclo. Cuando el estudiante obtiene un promedio menor o igual a 12, deserta el 87.00 % (241 estudiantes) de un total de 277 estudiantes. Si el estudiante obtiene un promedio mayor a 12, deserta el 20.95 % (75 estudiantes) de un total de 358 estudiantes. Hay que recordar que la cantidad de estudiantes desertores está balanceada con el fin de evitar sesgos en los análisis y resultados, por lo que las cantidades reales de estudiantes desertores en los resultados es menor.

Las reglas de asociación obtenidas para la deserción, asociadas al árbol de decisión de la figura 6, son las siguientes:

Reglas para Si - contiene 1 regla(s)

Regla 1 para Si (241; 0,870)

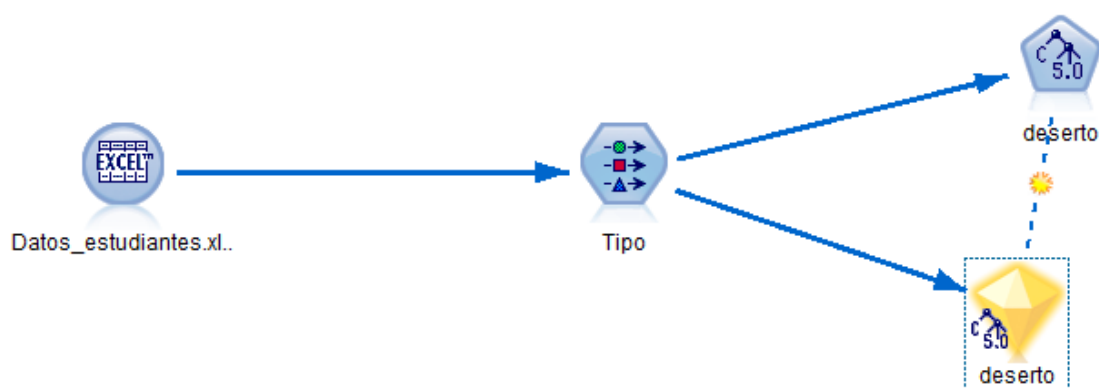
si $prom4 \leq 12$

entonces Si

La regla 1 se aplica a los estudiantes que obtuvieron en el cuarto ciclo un promedio menor o igual a 12, con un nivel de confianza de 0.870 y aplicable a 241 registros.

En la figura 7 se muestra la ruta utilizada es SPSS Modeler para realizar los análisis de los datos personales de los estudiantes:

FIGURA N° 7
RUTA EN SPSS MODELER PARA EL ANÁLISIS DE LOS DATOS PERSONALES DE LOS ESTUDIANTES



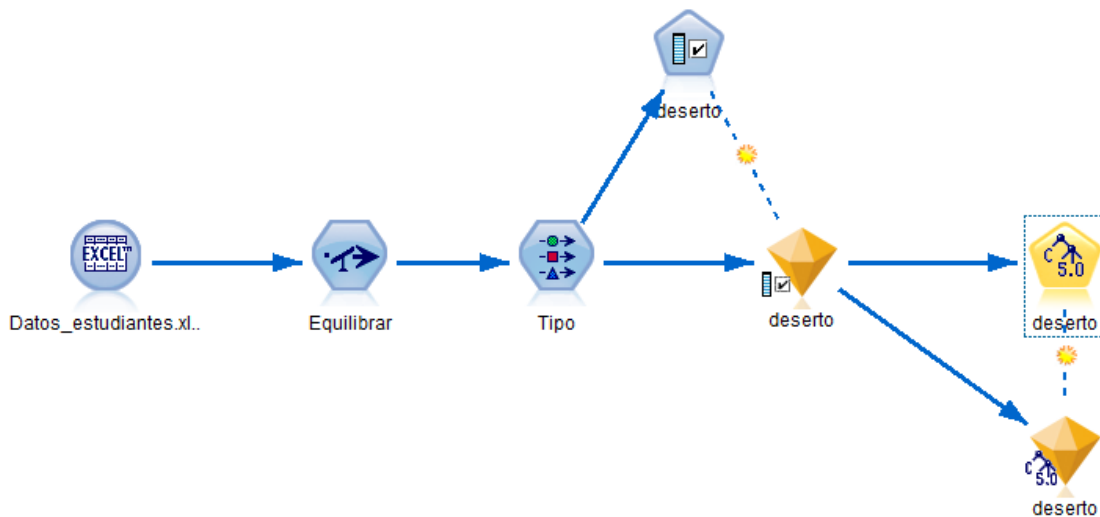
Fuente: Software SPSS Modeler

El libro de Excel en el que se almacenaron los datos de los estudiantes se denominó Datos_estudiantes.xlsx, el cual estaba compuesto por varias hojas de cálculo conteniendo los datos personales y académicos de los estudiantes, organizados de acuerdo al análisis en el que se iban a utilizar. El nodo de origen de datos Excel se cargó con los datos de la hoja de cálculo que contenía los datos personales de los estudiantes, incluido el promedio del colegio.

El nodo de tipo tiene como fuente de datos al nodo de origen, y su salida está conectada al nodo de modelado del algoritmo C5.0 y al nugget de resultados. En el nodo de tipo se estableció el campo de destino y se seleccionaron los predictores que se utilizaron para el análisis y la generación del árbol de decisión. El nodo de modelado implementa el algoritmo de análisis y generación del árbol de decisión, y puede configurar sus características. El nugget asociado al nodo de modelado contiene los resultados e información del análisis y el árbol de decisión.

En la figura 8 se muestra la ruta utilizada en SPSS Modeler para realizar los análisis de los datos académicos de los estudiantes:

FIGURA N° 8
RUTA EN SPSS MODELER PARA EL ANÁLISIS DE LOS DATOS ACADÉMICOS DE LOS ESTUDIANTES



Fuente: Software SPSS Modeler

La ruta utilizada para el análisis de los datos académicos de los estudiantes, utiliza adicionalmente un nodo equilibrar para corregir el desequilibrio presente en los datos utilizados para el estudio de la deserción en los ciclos de estudio. También utiliza un nodo selección de características cuya función es la de seleccionar los predictores más importantes para el análisis de la deserción. El nugget asociado que contiene el resultado de la selección de los descriptores más importantes, sirve como fuente de datos al nodo de modelado del algoritmo C5.0 y el nugget de resultados asociado.

Los resultados obtenidos en el descubrimiento de patrones de deserción estudiantil y los factores asociados a estos, validan la hipótesis planteada para el presente estudio. Además, el establecimiento del repositorio de datos, y la instauración de un procedimiento para la

investigación de la deserción estudiantil, demuestran que se ha cumplido con los objetivos planteados para la presente investigación.

CAPÍTULO IV

DISCUSIÓN

IV. DISCUSIÓN

La deserción estudiantil es un problema de dimensión mundial, que afecta el desarrollo y la economía de los países, por lo que muchos de ellos han implementado políticas de estado para su control, y han desarrollado o están desarrollando estudios y sistemas para prevenir este fenómeno. El Perú no es ajeno a este fenómeno, sin embargo, no existen estudios desarrollados por el Ministerio de Educación sobre la deserción estudiantil en los institutos superiores tecnológicos de nuestro país. Asimismo, existen escasos estudios realizados en algunos institutos del país sobre este tema, de ahí la importancia de establecer un procedimiento que facilite el estudio de la deserción estudiantil en los institutos superiores tecnológicos, en base a la información que ya tiene recopilada cada institución, como son los datos contenidos en las carpetas del postulante y las actas consolidadas de evaluación académica semestral, el cual fue el tercer objetivo del presente estudio.

El primer objetivo del presente estudio, fue establecer un repositorio de datos para almacenar los datos obtenidos mediante el proceso de selección, pre procesamiento y transformación de datos del proceso KDD de descubrimiento del conocimiento en bases de datos.

Como fuentes de datos se utilizaron las carpetas de postulante y las actas consolidadas de evaluación académica semestral de los estudiantes que ingresaron entre los años 2010 al 2013. Los datos de las carpetas de postulante, que son los datos personales de los estudiantes y de su colegio de procedencia, se encuentran en forma escrita en formatos oficiales del Ministerio de Educación y de la institución, mientras que los datos de las actas consolidadas, que contienen la información académica del estudiante, se encuentra en forma digital en hojas de cálculo de Excel.

Esta etapa es la que demanda mayor cantidad de tiempo, pues se transcribieron miles de datos de las carpetas del postulante al repositorio de datos, que consistió en una base de datos creada en el sistema de gestión de bases de datos MySQL. Se introdujeron los datos a la base de datos mediante una aplicación creada con el IDE de Visual Basic .NET, la que contaba con un formulario para el ingreso de los datos de cada campo de los registros de la base de datos. Asimismo, la aplicación importó automáticamente los datos académicos de los estudiantes, contenidos en hojas de cálculo especialmente acondicionadas para tal fin. Posteriormente, la información integrada en la base de datos, se exportó a hojas de cálculo para su procesamiento.

El segundo objetivo del presente estudio, fue descubrir patrones de deserción estudiantil asociados con los datos personales y académicos de los estudiantes que ingresaron

al instituto superior tecnológico “Daniel Villar” de Caraz, entre los años 2010 al 2013. Para ello, se siguieron los lineamientos del proceso KDD (Knowledge Discovery from Databases) de descubrimiento del conocimiento en bases de datos.

Los datos analizados, son datos secundarios obtenidos de las carpetas del postulante y de las actas consolidadas de evaluación académica semestral, documentos oficiales del Ministerio de Educación y de la institución.

Según la tabla N° 1, el porcentaje de deserción estudiantil fue de 41.75 % (248 estudiantes) en la cohorte de estudiantes ingresantes en los años 2010 al 2013, compuesta por un total de 594 estudiantes. Esta tasa de deserción es mayor que el nivel de deserción en los institutos superiores del país en el año 2013, que fué de 31 % en el nivel socioeconómico C, de 20 % en el nivel B y 13 % en el nivel A (Como se cita en <http://institutojuanbosco.blogspot.pe/2016/12/desercion-studiantil.html?view=magazine>). Asimismo, esta tasa de deserción es cercana a la del primer año en la educación superior tecnológica en Chile en el año 2006, que fue del 38 % (Universidad de Chile, 2008).

De los resultados mostrados en la tabla N° 2, se encuentra que la mayor tasa de deserción se produce en las especialidades de Mecánica Automotriz con un 53.85 % (77 estudiantes), seguido por Enfermería Técnica con un 44.30 % (66 estudiantes). Las especialidades que presentan una menor tasa de deserción son Secretariado Ejecutivo con un 35.57 % (53 estudiantes) y Computación e Informática con un 33.99 % (52 estudiantes), que siguen siendo significativas, lo que ameritó un estudio más exhaustivo para determinar los factores que intervinieron, con técnicas de minería de datos educacional.

Analizando el contenido de la tabla N° 3, se puede apreciar que los estudiantes desertan mayoritariamente en el segundo y tercer ciclo, en un 81.05 % (201 estudiantes) de un total de 248 estudiantes desertores, lo cual concuerda con los resultados consignados en el artículo de investigación de Zarría Torres, Arce Ramos y Lam Moraga (2016) realizada en la universidad Arturo Pratt de Chile. La deserción en los ciclos IV y V es del 18.95 % (47 estudiantes), que es apreciablemente menor a los ciclos anteriores. De estos resultados, se concluyó que este comportamiento establecía un patrón de deserción estudiantil, y que ameritaba un estudio más exhaustivo para determinar los factores que intervienen en dicho fenómeno, utilizando técnicas de minería de datos educacional.

Según la información de la tabla N° 5, el género de los estudiantes en dos especialidades tiene un marcado sesgo. En la especialidad de Mecánica Automotriz, de los 143 estudiantes matriculados durante los años 2010 al 2013, 142 han sido varones y solo ha habido una estudiante mujer. Igualmente, en la especialidad de Secretariado Ejecutivo de los 149 estudiantes matriculados durante los años 2010 al 2013, 139 han sido mujeres y solo 10 han

sido varones. Estos resultados demuestran que de acuerdo a la idiosincrasia de la zona, se considera que algunas profesiones son exclusivas para un determinado género. La proporción de género a nivel institucional, está mucho más equilibrado como se aprecia en la tabla N° 4.

Según la información contenida en la tabla 9, se aprecia que el promedio obtenido por los estudiantes en secundaria, en su gran mayoría es bastante bajo, puesto que el porcentaje de estudiantes que obtuvieron un promedio de 13 en el colegio es de 50.29 % (264 estudiantes), y de los que obtuvieron un promedio de 12 es de 25.71 % (135 estudiantes), que en conjunto suman un 76.00 % (399 estudiantes). Esta información establece un patrón de comportamiento de los estudiantes que postulan al instituto, y que según la información obtenida de la figura 2, la nota promedio obtenida en el colegio es uno de los factores que influye en la deserción estudiantil, resultado que no es concordante con el resultado de la investigación de Zarria Torres et al. (2016) realizada en la universidad Arturo Pratt de Chile.

Al realizar un análisis de la información contenida en la figura 3, se puede concluir que uno de los factores determinantes para la deserción en el segundo ciclo, es el promedio del primer ciclo. El nodo 1 del árbol de decisión de la figura 3, establece que cuando el promedio obtenido por el estudiante en el primer ciclo es menor o igual a 11, el estudiante deserta en un porcentaje de 98.31 %. Esta información es corroborada por la regla 1 de asociación generada por el software para el árbol de decisión, que tiene un nivel de confianza de 0.983.

Estos resultados establecen un patrón de deserción para los estudiantes que estudian en el primer ciclo y que desertan en el segundo, indicando que el promedio obtenido en el primer ciclo de estudios es uno de los factores que incide en la deserción estudiantil. La segunda regla de asociación generada por el software para el árbol de decisión, no se considera por tener un bajo nivel de confianza.

De la información contenida en el nodo 1 del árbol de decisión de la figura 4, se puede identificar como uno de los principales factores de deserción al promedio obtenido por el estudiante en el segundo ciclo. La información contenida en el nodo 1 establece que si un estudiante obtiene un promedio menor o igual a 10 en el segundo ciclo, deserta en un porcentaje de 96.75 %. Esta información es corroborada por la regla 1 de asociación generada por el software para el árbol de decisión de la figura 4, que tiene un nivel de confianza de 0.968.

Estos resultados establecen un patrón de deserción para los estudiantes que estudian en el segundo ciclo y que desertan en el tercero, indicando que el promedio obtenido en el segundo ciclo de estudios es uno de los factores que incide en la deserción estudiantil.

Asimismo, de lo establecido por la regla 2 de asociación para el árbol de decisión de la figura 4, se puede deducir que si un estudiante obtiene un promedio mayor a 10 en el segundo

ciclo, pero ha obtenido un promedio menor o igual a 11 en el primer ciclo y domicilia en la provincia de Huaylas, deserta con un nivel de confianza de 0.875.

Este resultado establece un patrón de deserción para los estudiantes que estudian en el segundo ciclo y que tuvieron un promedio menor o igual a 11 en el primer ciclo, que domicilian en la provincia de Huaylas y que desertan en el tercer ciclo, lo que indica que otros factores asociados a la deserción estudiantil en el tercer ciclo son el promedio obtenido en el primer ciclo y la provincia en la que domicilian.

De igual manera, de lo establecido por la regla 3 de asociación para el árbol de decisión de la figura 4, se puede deducir que si un estudiante obtiene un promedio de 11 o 12 en el segundo ciclo y un promedio mayor a 11 en el primer ciclo, es del género masculino y ha nacido en uno de los distritos de la provincia de Huaylas que no sean Caraz, Pueblo Libre o Santa Cruz, o ha nacido en el distrito de Yungay, deserta con un nivel de confianza de 0.87.

Este resultado establece un patrón de deserción para los estudiantes que estudian en el segundo ciclo y que obtienen un promedio de 11 o 12, y un promedio mayor a 11 en el primer ciclo, del género masculino y ha nacido en uno de los distritos de la provincia de Huaylas que no sean Caraz, Pueblo Libre o Santa Cruz, o ha nacido en el distrito de Yungay y que desertan en el tercer ciclo, lo que indica que otros factores asociados a la deserción estudiantil en el tercer ciclo son el género y el distrito donde nacieron.

De la información contenida en el nodo 1 del árbol de decisión de la figura 5, se establece que el promedio obtenido en el tercer ciclo es determinante para la deserción estudiantil en el cuarto ciclo. Así tenemos que si el estudiante obtiene un promedio menor o igual a 11, deserta en un porcentaje de 97.24 %. Esta información es corroborada por la regla 1 de decisión para el árbol de decisión de la figura 5, con un nivel de confianza de 0.972.

Estos resultados establecen un patrón de comportamiento para los estudiantes del tercer ciclo que obtienen un promedio menor o igual a 11 y que desertan en el cuarto ciclo. El promedio obtenido en el tercer ciclo es un factor que influye en la deserción estudiantil.

De la información contenida en el nodo 1 del árbol de decisión de la figura 6, se colige que el promedio obtenido en el cuarto ciclo es determinante para la deserción estudiantil en el quinto ciclo. Así tenemos que si el estudiante obtiene un promedio menor o igual a 12, deserta en un porcentaje de 87.00 %. Esta información es corroborada por la regla 1 de decisión para el árbol de decisión de la figura 6, con un nivel de confianza de 0.87.

Estos resultados establecen un patrón de comportamiento para los estudiantes del cuarto ciclo que obtienen un promedio menor o igual a 12 y que desertan en el quinto ciclo, indicando que el promedio obtenido en el cuarto ciclo es un factor que influye en la deserción estudiantil

CAPÍTULO V

CONCLUSIONES

V. CONCLUSIONES

Los datos académicos de los estudiantes, recopilados por el sistema de información de la institución bajo estudio, contienen información oculta que permitió identificar algunos patrones de deserción estudiantil y sus factores asociados.

Los datos personales de los estudiantes no revelan patrones de deserción relevantes, lo que sugiere que los datos que se están recopilando en la carpeta del postulante están escasamente relacionados con las causas de la deserción estudiantil en el instituto, por lo que se concluye que se debería ampliar el cuestionario de la mencionada carpeta para recopilar los datos relacionados con su entorno familiar, social y económico.

El único patrón descubierto con los datos personales de los estudiantes, está relacionado con el promedio obtenido por los estudiantes en el colegio, que mayoritariamente tienen niveles muy bajos. Se puede concluir que los estudiantes con bajos índices de aprovechamiento escolar son los que prefieren postular al instituto.

Los patrones de deserción estudiantil descubiertos tienen relación principalmente con el rendimiento académico de los estudiantes, por lo que es imperativo implementar algún tipo de sistema que alerte cuando el rendimiento académico de algún estudiante se encuentra dentro de los niveles identificados con la deserción.

El proceso de selección, pre procesamiento y transformación de datos, es la actividad que ha demandado mayor tiempo, debido a que se ha tenido que ingresar manualmente la información personal de los postulantes al repositorio de datos, ya que ésta se encuentra escrita en formularios. Se concluye que para facilitar futuras investigaciones, se debe automatizar el registro de la información personal de los postulantes.

Las técnicas de minería de datos educacional han demostrado ser útiles para el estudio de la deserción estudiantil, permitiendo descubrir patrones ocultos en los datos operativos de los estudiantes. Para realizar los análisis se utilizó el algoritmo C5.0, que es el algoritmo más utilizado en estudios realizados por otros investigadores sobre temas similares, quedando abierta la posibilidad de aplicación de otros algoritmos de minería de datos en estudios futuros.

Se cumplieron los objetivos planteados para la presente investigación, pues se estableció un repositorio de datos para su análisis mediante minería de datos, se descubrieron algunos patrones de deserción estudiantil y sus factores asociados, y se estableció un procedimiento para el estudio de la deserción estudiantil con minería de datos educacional.

La presente investigación sienta las bases para trabajos futuros de investigación sobre la deserción estudiantil y otros temas relacionados con el comportamiento estudiantil, en los institutos superiores tecnológicos de nuestro país.

CAPÍTULO VI

RECOMENDACIONES

VI. RECOMENDACIONES

Se formulan las siguientes recomendaciones:

Ampliar la cobertura del cuestionario de la carpeta del postulante, con la finalidad de recopilar los datos relacionados con el entorno familiar, social y económico de los estudiantes, para ser utilizados en estudios futuros.

Realizar acciones de seguimiento del rendimiento escolar de los estudiantes, con la finalidad de realizar labores de prevención de la deserción estudiantil, toda vez que el promedio obtenido en los ciclos de estudio es un factor de deserción.

Automatizar el registro de la información de los estudiantes, con la finalidad de disponer de un repositorio de datos digital que facilite la aplicación de técnicas de minería de datos educacional, orientadas a realizar estudios sobre el comportamiento estudiantil y tomar acciones de prevención de la deserción estudiantil, así como obtener información que permita la toma de decisiones y la mejora de la calidad educativa.

Crear un sistema de predicción de la deserción estudiantil, que permita alertar sobre la situación de los estudiantes en riesgo de desertar, utilizando los patrones de deserción estudiantil descubiertos con técnicas de minería de datos educacional.

CAPÍTULO VII

REFERENCIAS BIBLIOGRÁFICAS

VII. REFERENCIAS BIBLIOGRÁFICAS

- Aguirre Mendiola, J. L., Valdovinos Rosas, R. M., Antonio Velazquez, J. A., Alejo Eleuterio, R. y Marcial Romero, J. R. (2015). Análisis de deserción escolar con minería de datos. *Research in computing sciences*, 93, 71-82. Recuperado de http://www.rcs.cic.ipn.mx/2015_93/Analisis%20de%20desercion%20escolar%20con%20mineria%20de%20datos.pdf
- Alcolea Picazo, J. J. y Pavón De Paula, S. (2013). Los datos como recurso estratégico. En Alcolea Picazo, J. J. (Ed.), *Libro blanco de inteligencia institucional en universidades* (17-44). Madrid, España: Oficina de Cooperación Universitaria, S.A. Recuperado de http://www.ocu.es/wp-content/uploads/2015/09/OCU_LB_I2_013_Digital.pdf
- Allauca Armas, A. B. (2012). Influencia de la evaluación de los aprendizajes en la deserción estudiantil del instituto tecnológico superior "Benito Juárez" Propuesta de un manual de evaluación de los aprendizajes (Tesis de maestría). Recuperado de <http://www.dspace.uce.edu.ec/bitstream/25000/658/1/T-UCE-0010-165.pdf>
- Aluja, T. (2001). La minería de datos, entre la estadística y la inteligencia artificial. Recuperado de www.raco.cat/index.php/Questiio/article/download/27009/26843
- Amaya Torrado, Y. K., Barrientos Avendaño, E., y Heredia Vizcaíno, D. J. (2014). Modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos. *Actas TICAL 2014*. 553-562. Recuperado de <http://hdl.handle.net/10786/759>
- Argote, I., Jiménez, R. y Gómez, J. (2014). Detección de patrones de deserción en los programas de pregrado de la Universidad Mariana de San Juan de Pasto, aplicando el proceso de descubrimiento de conocimiento sobre base de datos (KDD) y su implementación en modelos matemáticos de predicción. *Libro de actas IV CLABES 2014*, 185-193. <http://dx.doi.org/10.21803%2Fpenamer.6.10.137>
- Baker, R. S. y Siemens, G. (2013). Educational Data Mining and Learning Analytics. Recuperado de <http://www.columbia.edu/~rsb2162/BakerSiemensHandbook2013.pdf>
- Baker, R. S. (2014). Data Mining for Education. *International encyclopedia of education*, 112-118. Recuperado de <http://www.columbia.edu/~rsb2162/Encyclopedia%20Chapter%20Draft%20v10%20-fw.pdf>
- Ballesteros Román, A., Sánchez-Guzmán, D. y García Salcedo, R. (2013). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un

- contexto educativo. *Latin-American Journal of Physics Education*, 662-668. Recuperado de http://www.lajpe.org/dec13/22-LAJPE_814_bis_Alejandro_Ballesteros.pdf
- Barragán, S. y Gonzales, L. (2015). Un modelo para explicar la retención en la Universidad de Bogotá Jorge Tadeo Lozano: Árboles de decisión. Recuperado de http://www.alfaguia.org/www-alfa/images/ponencias/clabesv/L1_Ponencias/5_CLABES_paper_6.pdf
- Camana Fiallos, R. (2016). Aplicación de técnicas de minería de datos educacionales, para predecir la deserción académica. Recuperado de <http://acceso.virtualeduca.red/documentos/ponencias/puerto-rico/750-a3a6.pdf>
- Castaño, E., Gallón, S., Gómez, K. y Vásquez, J. (2006). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. *Lecturas de Economía*, 65, 9-36. Recuperado de http://www.alfaguia.org/alfaguia/files/1319760796_15.pdf
- Celis, S., Moreno, L., Poblete, P., Villanueva, J. y Weber, R. (2015). Un modelo analítico para la predicción del rendimiento académico de estudiantes de ingeniería. Recuperado de <http://www.dii.uchile.cl/~ris/RIS2015/rendimientoac.pdf>
- Dapozo, G., Porcel, E., López, M. V., Bogado, V. y Bargiele, R. (2006). Aplicación de minería de datos con una herramienta de software libre en la evaluación del rendimiento académico de los alumnos de la carrera de Sistemas de la FACENA-UNNE. Recuperado de http://sedici.unlp.edu.ar/bitstream/handle/10915/20797/Documento_completo.pdf?sequence=1
- Delen, D. (2010). A comparative analysis of machine learning techniques for student retention management. *Decision Support Systems*, 498-506. <http://dx.doi.org/10.1016/j.dss.2010.06.003>
- Eckert, K. y Suénaga, R. (2013). Aplicación de técnicas de Minería de Datos al análisis de situación y comportamiento académico de alumnos de la UGD. Recuperado de http://sedici.unlp.edu.ar/bitstream/handle/10915/27103/Documento_completo.pdf?sequence=1
- García Herrero, J. y Molina López, J.M. (2012) Técnicas de análisis de datos. Recuperado de <http://ocw.uc3m.es/ingenieria-informatica/analisis-de-datos/libroDataMiningv5.pdf>

- García Lázaro, D. (2015) Propuesta de un modelo de estimación del abandono universitario (Tesis doctoral). Recuperado de <http://hdl.handle.net/10115/13222>
- Han, J., Kamber, M. y Pei, J. (2012). *Data Mining*. San Francisco: Morgan Kaufmann Publishers
- Hernández Lamillar, R. M. (2005) Herramientas del marketing: data warehousing, tecnología necesaria para el comercio internacional. *Revista CENIC ciencias biológicas*. Recuperado de <http://revista.cnic.edu.cu/revistaCB/sites/default/files/articulos/CB-2005-4-CB-061.pdf>
- Kumar, V. y Chadha, A. (2011). An empirical study of applications of data mining techniques in education. *International Journal of Advanced Computer Science and Applications*, 80-84. Recuperado de <http://thesai.org/Downloads/Volume2No3/Paper%2014-%20An%20Empirical%20Study%20of%20the%20Applications%20of%20Data%20Mining%20Techniques%20in%20Higher%20Education.pdf>
- Marcano Y. J. y Rodríguez R.H. (2014). Minería de datos aplicada a la deserción estudiantil. Caso: Licenciatura en Computación de la Universidad del Zulia-NPF. *Revista EDUCARE*, 31-51. Recuperado de <http://revistas.upel.edu.ve/index.php/educare/article/view/2600/1255>
- Márquez Vera, C. (2015). *Predicción del fracaso y abandono escolar mediante técnicas de minería de datos* (Tesis doctoral). Recuperado de <http://helvia.uco.es/xmlui/bitstream/handle/10396/12852/2015000001157.pdf?sequence=1>
- Méndez Alandete, J. J. (2015). Proyección de estudiantes en riesgo de desertar mediante técnicas de minería de datos. Recuperado de revistas.cecar.edu.co/ingenieria/article/download/179/168
- Ministerio de Educación Nacional de Colombia (2015). Estrategias para la Permanencia en educación superior: Experiencias significativas. Recuperado de http://www.colombiaaprende.edu.co/html/micrositios/1752/articles-350844_pdf.pdf
- Montalvo Fránquiz, M., André Ampuero, M. y Baldoquín de la Peña, M. G. (2010), Determinación de patrones que contribuyan a la formación de un equipo de proyecto de software, utilizando técnicas del modelo predictivo de la minería de datos. Recuperado de <http://www.bdigitalitc.mx/cgi-bin/koha/opac-detail.pl?biblionumber=52844#>

- Otamendi, A. y Otero, M. P. (2009). Análisis de datos secundarios cuantitativos. Recuperado de metodo3.sociales.uba.ar/files/2015/03/DC_57_Otamendi_Otero_2009.pdf
- Palomo Miñambres, O. (s.f.). Minería de datos. Recuperado de <http://www.it.uc3m.es/jvillena/irc/practicas/11-12/12mem.pdf>
- Rodríguez Rodríguez, J.M. y Daureo Campillo, M.J. (2003). Sistemas de información: Aspectos técnicos y legales. Recuperado de <http://www.ual.es/~jmrodri/sistemasdeinformacion.pdf>
- Rumberger, R. W. (2011). Dropping out. Why students drop out of high school and what can be done about it. Recuperado de mina.education.ucsb.edu/rumberger/book/ch1.pdf
- Sembiring, S., Zarlis, M., Hartama, D., Ramlina, S. y Wani, E. (2011). Prediction of student academic performance by an application of Data Mining techniques. *2011 International Conference on Management and Artificial Intelligence*, 6, 110-114. Recuperado de <http://www.ipedr.com/vol6/21-A10015.pdf>
- Tafur Puente, R.M. (2012). Los sistemas de acceso, normativas de permanencia, y estrategias de tutoría y retención de estudiantes de educación superior – INFORME NACIONAL PERÚ –. Recuperado de <http://departamento.pucp.edu.pe/educacion/files/2014/11/Tafur-R-2012-Accedes.pdf>
- Timarán Pereira, R., Calderón Romero, A. y Jiménez Toledo, J. (2013). Descubrimiento de perfiles de deserción estudiantil con técnicas de minería de datos. *Revista vínculos*, 10(1), 373-383. Recuperado de <http://revistas.udistrital.edu.co/ojs/index.php/vinculos/article/view/4687/6419>
- Timarán Pereira, R. (2009). Detección de patrones de bajo rendimiento académico y deserción estudiantil con técnicas de minería de datos. Octava Conferencia Iberoamericana en Sistemas, Cibernética e Informática CИСCI 2009. Florida, U.S.A.: International Institute of Informatics and Systemics. Recuperado de www.iiis.org/CDs2008/CD2009CSC/CISCИ2009/PapersPdf/C692YV.pdf
- Universidad de Chile. (2008). Estudio sobre las causas de la deserción universitaria. Recuperado de www.oei.es/historico/pdf2/causas-desercion-universitaria-chile.pdf
- Valero Orea, S. (2009). Aplicación de técnicas de minería de datos para predecir deserción. Recuperado de <http://www.utim.edu.mx/~svalero/docs/MineriaDesercion.pdf>

- Veitch, W. (2004). Identifying characteristics of high school dropouts: data mining with a decision tree model. *Annual meeting of the american educational research association, 1-11*. Recuperado de files.eric.ed.gov/fulltext/ED490086.pdf
- Zaragoza Loya, J. E. (2013). Factores determinantes de abandono y permanencia en los estudios de bachillerato en música y arte de la facultad de música de la Universidad Autónoma de Tamaulipas (México) (Tesis doctoral). Recuperado de <http://hera.ugr.es/tesisugr/2281050x.pdf>
- Zarria Torres, C., Arce Ramos, C. y Lam Moraga, J. (2016). Estudio de variables que influyen en la deserción de estudiantes universitarios de primer año, mediante minería de datos. *Universidad Científica del Perú Ciencia amazónica (Iquitos)*, 6 (1), 73-84. Recuperado de <http://dx.doi.org/10.22386/ca.v6i1.110>

CAPÍTULO VIII

ANEXOS

VIII. ANEXOS

ANEXO 01

PROPUESTA PEDAGÓGICA

I. DENOMINACIÓN:

<p style="text-align: center;">Sistema de predicción y tratamiento de la deserción estudiantil en el Instituto de Educación Superior Tecnológico Público “Daniel Villar” de Caraz</p>
--

II. AUTOR:

Apellidos y Nombres	DNI	Correo electrónico
Torres Alegre Luis Augusto	32384844	lutor1@gmail.com

III. LOCALIZACIÓN:

Región	Ancash
Provincia	Huaylas
Distrito	Caraz
Localidad	Caraz
Institución educativa	Instituto de Educación Superior Tecnológico Público “Daniel Villar”

IV. RESPONSABLES:

- 4.1 Coordinador del proyecto
- 4.2 Investigador
- 4.3 Profesional en programación

V. BENEFICIARIOS:

- 5.1 Directos:
 - Estudiantes del Instituto de Educación Superior Tecnológico “Daniel Villar” de Caraz.
- 5.2 Indirectos:

Dirección y área administrativa de la institución.

Población de las localidades de influencia del instituto.

Dirección Regional de Educación de Ancash y Ministerio de Educación.

VI. DESCRIPCIÓN DE LA PROPUESTA:

Abordar el problema de la deserción estudiantil en el Instituto de Educación Superior Tecnológico Público “Daniel Villar” de Caraz, con la finalidad de disminuir el porcentaje de deserción estudiantil mediante el desarrollo de una aplicación informática que disponga de un sistema de recolección de datos, y realice la medición y el diagnóstico del riesgo de deserción, produciendo información en tiempo real para realizar acciones que permitan prevenir la deserción estudiantil. El sistema de diagnóstico deberá implementar las expresiones lógicas y matemáticas que modelan el comportamiento de los estudiantes, obtenidas mediante técnicas de minería de datos educacional.

VII. INTRODUCCIÓN:

7.1 Fundamento teórico

La deserción estudiantil produce un impacto económico en la sociedad, afecta al estudiante a nivel personal, al mercado laboral y al desarrollo del país. La inversión realizada por el estado y el estudiante durante el tiempo que estudió se desperdicia, asimismo el estudiante se siente frustrado al no haber cumplido su meta y además tiene problemas en encontrar empleo, y si lo encuentra, el sueldo que gane será menor en relación al de un profesional, y finalmente, el mercado laboral pierde un personal calificado, lo que repercute en el desarrollo del país (Rumberger, 2011). Las implicancias de la deserción estudiantil, convierte a este fenómeno en un problema que requiere urgente atención.

El estudio de la deserción estudiantil no solo debe centrarse en su análisis estadístico e identificación de los factores que lo producen, sino también en establecer un modelo de predicción para poder predecirlo y prevenirlo (García Lázaro, 2015).

El procesamiento de los datos operativos de los estudiantes almacenados en el sistema de información de la institución, se puede realizar utilizando técnicas de minería de datos educacional. El objetivo es descubrir patrones de deserción ocultos en los datos de los estudiantes y sus factores asociados, que sirvan para desarrollar un sistema de predicción temprana de los estudiantes que se encuentran en riesgo de desertar.

La minería de datos educacional se proyecta como un proceso para el análisis de datos que provienen de un contexto educativo. Permite descubrir y analizar patrones de comportamiento de los estudiantes, analizando los datos de los estudiantes almacenados en las bases de datos (Ballesteros, Sánchez-Guzmán y García, 2013).

Las técnicas de minería de datos educacional utilizadas generalmente para el descubrimiento de patrones de deserción estudiantil y bajo rendimiento académico, son las reglas de clasificación, el clustering, los patrones secuenciales, las asociaciones, los árboles de decisión entre otros (Timarán, 2009).

Dentro de las técnicas de la minería de datos educacional, tenemos los algoritmos de reglas de asociación, los árboles de decisión y la regresión logística. Como resultado de la utilización de éstas técnicas, se obtienen expresiones lógicas y matemáticas que modelan el comportamiento de los estudiantes, y que pueden ser implementados mediante un lenguaje de programación, para realizar la predicción de los estudiantes en riesgo de deserción.

Con los modelos de deserción obtenidos, se puede desarrollar un software de predicción en un lenguaje de programación visual. Se utilizarán los datos operativos de los estudiantes como datos de entrada del software y se determinarán los estudiantes que se encuentran en riesgo de deserción. Con esta información, se pueden tomar las medidas correctivas necesarias.

El manejo del software lo deberán realizar aquellos actores que interactúan directamente con los estudiantes, como los docentes y los tutores. Por esta razón, se deberá realizar un programa de capacitación para dichos actores, que puede extenderse a otras áreas administrativas.

7.2 Justificación

- Social

La deserción estudiantil afecta de manera significativa al Instituto de Educación Superior Tecnológico Público “Daniel Villar” de Caraz. La deserción estudiantil en el instituto fue de 41.75 % en la cohorte de estudiantes que ingresaron durante los años 2010 al 2013. Según el diario Gestión, el nivel de deserción en los institutos superiores en el año 2013, fué de 31 % en el nivel socioeconómico C, de 20 % en el nivel B y 13 % en el nivel A (Como se cita en <http://institutojuanbosco.blogspot.pe/2016/12/desercion-studiantil.html?view=magazine>). De lo expuesto, se

puede observar que el porcentaje de deserción en el instituto está muy por encima del nivel señalado para nuestro país para el año 2013.

Poder predecir qué estudiantes se encuentran en riesgo de desertar, es a todas luces un aspecto muy importante pues permitirá tomar las acciones oportunas para disminuir la deserción. Los principales beneficiados serán los estudiantes, pues en base a la información y el conocimiento obtenido, los responsables educativos podrán tomar decisiones tendientes a disminuir la deserción estudiantil en la institución. Asimismo, se beneficiará al personal responsable de la toma de decisiones, pues contarán con una herramienta que les apoyará en dicha función.

Las acciones que se tomarán para tratar de corregir los factores que ponen al estudiante en riesgo de deserción, incidirán en una mejora de la calidad educativa en la institución.

- Tecnológica

En los sistemas de información de las instituciones educativas superiores se está produciendo un gran aumento de los datos almacenados de los estudiantes, pero que no están produciendo ningún beneficio al no ser sometidos a un proceso de análisis. Los datos constituyen un recurso ya adquirido, disponible y en constante crecimiento, por lo que parece razonable plantearse si se está haciendo una explotación eficaz de ellos, considerando a los datos como un recurso estratégico (Alcolea y Pavón, 2013).

La institución tiene almacenados los datos operativos de los estudiantes, compuestos por sus datos personales y académicos que fueron recopilados en documentos oficiales de la institución. Uno de estos documentos es la carpeta del postulante, que contiene formularios que deben ser llenados con información personal del estudiante, y que además debe ser completada con copias de sus documentos personales y el certificado de estudios secundarios, al momento de inscribirse para el examen de admisión. Los otros documentos son las actas consolidadas de evaluación académica semestral, que contienen la información del rendimiento académico del estudiante cuando cursa sus estudios. Estos datos tienen oculta información sobre los patrones de comportamiento de los estudiantes que desertaron y los factores de deserción asociados, por lo que es razonable utilizar técnicas de minería de datos educacional para descubrir dicha información que puede ser utilizada para crear un sistema de predicción temprana que permita

conocer con anticipación a los estudiantes en riesgo de deserción, para tomar las acciones correctivas correspondientes.

VIII. OBJETIVOS:

8.1 Objetivo general

Disminuir el porcentaje de deserción estudiantil en el Instituto de Educación Superior Tecnológico Público “Daniel Villar” de Caraz, mediante la predicción de los estudiantes en riesgo de deserción.

8.2 Objetivos específicos

- Desarrollar e implementar un software de predicción, utilizando expresiones lógicas y matemáticas para implementar las reglas que modelan las características y el comportamiento de los estudiantes en mayor riesgo de deserción.
- Capacitar a los docentes y personal responsable del bienestar estudiantil, en el uso del software de predicción, para diagnosticar, monitorear e identificar a los estudiantes en riesgo de deserción, para proceder a su pronta atención.

IX. DISEÑO METODOLÓGICO:

Etapas	Estrategias/actividades	Recursos /Materiales
1ra. Etapa: Diseño de la propuesta	Formulación del proyecto Estudio de factibilidad	Personal responsable Computadora Impresora Papel
2da. Etapa: Evaluación diagnóstica de la propuesta	Establecer el grado de conocimiento de los docentes sobre las causas de deserción Determinar el nivel de manejo de sistemas informáticos de los docentes	Personal responsable Computadora Impresora Papel
3ra. Etapa: Diseño del plan de la propuesta	Diseñar un plan operativo Establecer horarios de reunión del equipo responsable	Personal responsable Computadora

	<p>Desarrollar un plan de programación del software</p> <p>Diseñar un plan de capacitación de manejo del software para los docentes y área administrativa</p>	<p>Impresora</p> <p>Papel</p>
<p>4ta. Etapa:</p> <p>Ejecución de la propuesta</p>	<p>Establecer las expresiones lógicas y matemáticas que servirán de base para el sistema de predicción</p> <p>Desarrollar el software de predicción</p> <p>Desarrollar el plan de capacitación de los docentes, tutores y personal administrativo</p>	<p>Personal responsable</p> <p>Software de minería de datos</p> <p>Software de lenguaje de programación</p> <p>Computadora</p> <p>Impresora</p> <p>Papel</p>
<p>5ta. Etapa:</p> <p>Evaluación e informe de la propuesta</p>	<p>Evaluación del nivel de manejo del sistema por parte de los docentes, tutores y personal administrativo</p> <p>Evaluación de la capacidad de predicción del software</p> <p>Evaluación del establecimiento de medidas correctivas</p> <p>Informe de las metas alcanzadas</p>	<p>Personal responsable</p> <p>Computadora</p> <p>Impresora</p> <p>Papel</p>

X. METAS:

Descubrir los patrones de deserción estudiantil en el instituto mediante técnicas de minería de datos educacional, y determinar las expresiones lógicas y matemáticas que modelan dichos patrones.

Crear un software que implemente dichos modelos, teniendo como entrada los datos personales y académicos de los estudiantes, y como salida una indicación del nivel de riesgo de deserción del estudiante.

Capacitar a los docentes, tutores y personal administrativo en el uso del software de predicción, y sensibilizarlos respecto a la importancia de su utilización.

Establecer medidas correctivas para minimizar el riesgo de deserción del estudiante, incidiendo en la calidad educativa de la institución, disminuyendo el nivel de deserción.

XI. RESULTADOS OBSERVABLES

Disminución del porcentaje de deserción estudiantil en la institución, alcanzando niveles mínimos.

Nivel de manejo óptimo del software de predicción por parte de los docentes, tutores y personal administrativo, y compromiso con la meta de disminución del nivel de deserción.

Mejora de la calidad educativa en la institución, reflejándose en la satisfacción y bienestar del estudiante.

XII. BIBLIOGRAFÍA

Alcolea Picazo, J. J. y Pavón De Paula, S. (2013). Los datos como recurso estratégico. En Alcolea Picazo, J. J. (Ed.), Libro blanco de inteligencia institucional en universidades (pp. 17-44). Madrid, España: Oficina de Cooperación Universitaria, S.A. Recuperado de http://www.ocu.es/wp-content/uploads/2015/09/OCU_LB_I2_013_Digital.pdf

Ballesteros Román, A., Sánchez-Guzmán, D. y García Salcedo, R. (2013). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. *Latin-American Journal of Physics Education*, 662-668. Recuperado de http://www.lajpe.org/dec13/22-LAJPE_814_bis_Alejandro_Ballesteros.pdf

García Lázaro, D. (2015) Propuesta de un modelo de estimación del abandono universitario (Tesis doctoral). Recuperado de <http://hdl.handle.net/10115/13222>

Rumberger, R. W. (2011). Dropping out. Why students drop out of high school and what can be done about it. Recuperado de [mina.education.ucsb.edu/rumberger /book/ch1.pdf](http://mina.education.ucsb.edu/rumberger/book/ch1.pdf)

Timarán Pereira, R. (2009). Detección de patrones de bajo rendimiento académico y deserción estudiantil con técnicas de minería de datos. Octava Conferencia Iberoamericana en Sistemas, Cibernética e Informática CИСCI 2009. Florida, U.S.A.: International Institute of Informatics and Systemics. Recuperado de www.iiis.org/CDs2008/CD2009CSC/CISCИ2009/PapersPdf/C692YV.pdf

ANEXO 02

**VALIDACIÓN DE INSTRUMENTO DE RECOLECCIÓN DE DATOS POR CRITERIO DE JUECES O
EXPERTOS**

VALIDEZ POR CRITERIO DE JUECES O EXPERTOS

MATRIZ DE VALIDACIÓN

TÍTULO DE LA TESIS: Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz – 2017

VARIABLE	DIMENSIÓN	INDICADOR	ESCALA DE MEDICIÓN	CRITERIOS DE EVALUACIÓN						OBSERVACIÓN Y/O RECOMENDACIONES	
				Relación entre la variable y la dimensión		Relación entre la dimensión y el indicador		Relación entre el indicador y la escala de medición			
				SI	NO	SI	NO	SI	NO		
Datos operativos de los estudiantes	Datos personales y académicos de los estudiantes, recopilados por la institución en el proceso de admisión y el desarrollo de sus estudios.	Datos personales	Distrito de nacimiento	Nominal	X		X		X		
		Datos personales de los estudiantes obtenidos de la carpeta del postulante	Provincia de nacimiento	Nominal			X		X		
		Departamento de nacimiento	Nominal	X				X			
		Edad al ingresar	De razón, discreta	X				X			
		Distrito de residencia	Nominal	X				X			
		Provincia de residencia	Nominal	X				X			
		Departamento de residencia	Nominal	X				X			
		Colegio de procedencia	Nominal	X				X			
		Tipo de colegio	Nominal (Nacional, Particular)	X				X			
		Modalidad del colegio	Nominal (EBA, EBR)	X				X			
		Distrito donde está ubicado el colegio	Nominal	X				X			
		Provincia donde está ubicado el colegio	Nominal	X				X			
		Departamento donde está ubicado el colegio	Nominal	X				X			
		Diferencia entre el año de ingreso al instituto y el año de egreso del colegio	De razón, discreta	X				X			
Sexo	Nominal (Varón, Mujer)	X		X							

		Estado civil	Nominal (Soltero, Casado)			X		X		
		Trabaja	Nominal (Sí, No)			X		X		
		Nivel académico	Nominal (Secundario, Superior)			X		X		
	Datos académicos	Promedio final del estudiante en el colegio	De razón, discreta (0 a 20)			X		X		
	Datos del rendimiento académico de los estudiantes obtenidos de las actas consolidadas de evaluación académica semestral y del certificado de estudios secundarios	Año de ingreso	De razón, discreta (2010 a 2013)			X		X		
		Carrera profesional	Nominal			X		X		
		Ciclo de abandono	Nominal (I, II, III, IV, V, VI)			X		X		
		Promedio I ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio II ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio III ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio IV ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio V ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio VI ciclo	De razón, discreta (0 a 20)			X		X		

MATRIZ DE VALIDACIÓN DE INSTRUMENTO

NOMBRE DEL INSTRUMENTO: Base de datos para la transcripción de los datos personales y académicos de los estudiantes.

OBJETIVO: Almacenamiento de los datos producto de los procesos de selección, pre procesamiento y transformación de los datos operativos de los estudiantes.

UNIDAD DE ANÁLISIS: Carpetas del postulante y actas consolidadas de evaluación académica semestral de los estudiantes que ingresaron durante los años 2010 al 2013.

APELLIDOS Y NOMBRES DEL EVALUADOR: Dr. JESÚS VÍCTOR ZÚÑIGA HUERTA

GRADO ACADÉMICO DEL EVALUADOR: _____

DOCTOR GESTIÓN EDUCACIÓN SUPERIOR

VALORACIÓN:

APLICABLE X	APLICABLE DESPUÉS DE LEVANTAR LAS OBSERVACIONES	NO APLICABLE
-----------------------	---	--------------

FECHA: 24/01/2017



Firma del evaluador


Dr. Jesús Víctor Zúñiga Huerta
Ponente
Escuela de Gestión Educativa
PCB - UENCA

VALIDEZ POR CRITERIO DE JUECES O EXPERTOS

MATRIZ DE VALIDACIÓN

TÍTULO DE LA TESIS: Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz – 2017

VARIABLE	DIMENSIÓN	INDICADOR	ESCALA DE MEDICIÓN	CRITERIOS DE EVALUACIÓN						OBSERVACIÓN Y/O RECOMENDACIONES	
				Relación entre la variable y la dimensión		Relación entre la dimensión y el indicador		Relación entre el indicador y la escala de medición			
				SI	NO	SI	NO	SI	NO		
Datos operativos de los estudiantes	Datos personales y académicos de los estudiantes, recopilados por la institución en el proceso de admisión y el desarrollo de sus estudios.	Datos personales	Distrito de nacimiento	Nominal	X		X		X		
		Datos personales de los estudiantes obtenidos de la carpeta del postulante	Provincia de nacimiento	Nominal			X		X		
		Departamento de nacimiento	Nominal	X				X			
		Edad al ingresar	De razón, discreta	X				X			
		Distrito de residencia	Nominal	X				X			
		Provincia de residencia	Nominal	X				X			
		Departamento de residencia	Nominal	X				X			
		Colegio de procedencia	Nominal	X				X			
		Tipo de colegio	Nominal (Nacional, Particular)	X				X			
		Modalidad del colegio	Nominal (EBA, EBR)	X				X			
		Distrito donde está ubicado el colegio	Nominal	X				X			
		Provincia donde está ubicado el colegio	Nominal	X				X			
		Departamento donde está ubicado el colegio	Nominal	X				X			
		Diferencia entre el año de ingreso al instituto y el año de egreso del colegio	De razón, discreta	X				X			
Sexo	Nominal (Varón, Mujer)	X		X							

		Estado civil	Nominal (Soltero, Casado)			X		X		
		Trabaja	Nominal (Sí, No)			X		X		
		Nivel académico	Nominal (Secundario, Superior)			X		X		
	Datos académicos	Promedio final del estudiante en el colegio	De razón, discreta (0 a 20)			X		X		
	Datos del rendimiento académico de los estudiantes obtenidos de las actas consolidadas de evaluación académica semestral y del certificado de estudios secundarios	Año de ingreso	De razón, discreta (2010 a 2013)			X		X		
		Carrera profesional	Nominal			X		X		
		Ciclo de abandono	Nominal (I, II, III, IV, V, VI)			X		X		
		Promedio I ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio II ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio III ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio IV ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio V ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio VI ciclo	De razón, discreta (0 a 20)			X		X		

MATRIZ DE VALIDACIÓN DE INSTRUMENTO

NOMBRE DEL INSTRUMENTO: Base de datos para la transcripción de los datos personales y académicos de los estudiantes.

OBJETIVO: Almacenamiento de los datos producto de los procesos de selección, pre procesamiento y transformación de los datos operativos de los estudiantes.

UNIDAD DE ANÁLISIS: Carpetas del postulante y actas consolidadas de evaluación académica semestral de los estudiantes que ingresaron durante los años 2010 al 2013.

APELLIDOS Y NOMBRES DEL EVALUADOR: _____

DE LA CRUZ RAMÍREZ YULIANA MERCEDES

GRADO ACADÉMICO DEL EVALUADOR: DOCTORA EN GESTIÓN Y CIENCIAS DE LA EDUCACIÓN

VALORACIÓN:

APLICABLE X	APLICABLE DESPUÉS DE LEVANTAR LAS OBSERVACIONES	NO APLICABLE
---------------------------	---	--------------

FECHA: 25/01/17



Firma del evaluador

VALIDEZ POR CRITERIO DE JUECES O EXPERTOS

MATRIZ DE VALIDACIÓN

TÍTULO DE LA TESIS: Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz – 2017

VARIABLE	DIMENSIÓN	INDICADOR	ESCALA DE MEDICIÓN	CRITERIOS DE EVALUACIÓN						OBSERVACIÓN Y/O RECOMENDACIONES	
				Relación entre la variable y la dimensión		Relación entre la dimensión y el indicador		Relación entre el indicador y la escala de medición			
				SI	NO	SI	NO	SI	NO		
Datos operativos de los estudiantes	Datos personales y académicos de los estudiantes, recopilados por la institución en el proceso de admisión y el desarrollo de sus estudios.	Datos personales	Distrito de nacimiento	Nominal	X		X		X		
		Datos personales de los estudiantes obtenidos de la carpeta del postulante	Provincia de nacimiento	Nominal			X		X		
		Departamento de nacimiento	Nominal	X				X			
		Edad al ingresar	De razón, discreta	X				X			
		Distrito de residencia	Nominal	X				X			
		Provincia de residencia	Nominal	X				X			
		Departamento de residencia	Nominal	X				X			
		Colegio de procedencia	Nominal	X				X			
		Tipo de colegio	Nominal (Nacional, Particular)	X				X			
		Modalidad del colegio	Nominal (EBA, EBR)	X				X			
		Distrito donde está ubicado el colegio	Nominal	X				X			
		Provincia donde está ubicado el colegio	Nominal	X				X			
		Departamento donde está ubicado el colegio	Nominal	X				X			
		Diferencia entre el año de ingreso al instituto y el año de egreso del colegio	De razón, discreta	X				X			
Sexo	Nominal (Varón, Mujer)	X		X							

		Estado civil	Nominal (Soltero, Casado)			X		X		
		Trabaja	Nominal (Sí, No)			X		X		
		Nivel académico	Nominal (Secundario, Superior)			X		X		
	Datos académicos	Promedio final del estudiante en el colegio	De razón, discreta (0 a 20)			X		X		
	Datos del rendimiento académico de los estudiantes obtenidos de las actas consolidadas de evaluación académica semestral y del certificado de estudios secundarios	Año de ingreso	De razón, discreta (2010 a 2013)			X		X		
		Carrera profesional	Nominal			X		X		
		Ciclo de abandono	Nominal (I, II, III, IV, V, VI)			X		X		
		Promedio I ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio II ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio III ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio IV ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio V ciclo	De razón, discreta (0 a 20)			X		X		
		Promedio VI ciclo	De razón, discreta (0 a 20)			X		X		

MATRIZ DE VALIDACIÓN DE INSTRUMENTO

NOMBRE DEL INSTRUMENTO: Base de datos para la transcripción de los datos personales y académicos de los estudiantes.

OBJETIVO: Almacenamiento de los datos producto de los procesos de selección, pre procesamiento y transformación de los datos operativos de los estudiantes.

UNIDAD DE ANÁLISIS: Carpetas del postulante y actas consolidadas de evaluación académica semestral de los estudiantes que ingresaron durante los años 2010 al 2013.

APELLIDOS Y NOMBRES DEL EVALUADOR: _____

OLAZA MAGUIÑA AUGUSTO FÉLIX

GRADO ACADÉMICO DEL EVALUADOR: DOCTOR EN GESTIÓN Y CIENCIAS DE LA EDUCACIÓN

VALORACIÓN:

APLICABLE X	APLICABLE DESPUÉS DE LEVANTAR LAS OBSERVACIONES	NO APLICABLE
---------------------------	---	--------------

FECHA: 25/01/17



Firma del evaluador

ANEXO 03

MATRIZ DE CONSISTENCIA

Título: Patrones de deserción estudiantil asociados a los datos operativos de los estudiantes en el instituto superior tecnológico Daniel Villar de Caraz - 2016

Problema	Objetivos	Hipótesis	Variables	Dimensiones	Indicadores	Diseño	Población y muestra
¿Cuáles son los patrones de deserción estudiantil ocultos en los datos operativos de los estudiantes que abandonaron voluntariamente sus estudios en el	General: Descubrir patrones de deserción estudiantil e identificar los factores que influyen en la deserción relacionados con	Hi: Los patrones de deserción estudiantil descubiertos mediante el procesamiento de los datos operativos de los	Variable independiente: datos operativos de los estudiantes.	Datos personales	Distrito de nacimiento Provincia de nacimiento Departamento de nacimiento Edad al ingresar Distrito de residencia Provincia de residencia Departamento de residencia	Diseño: Descriptivo con propuesta Tipo: Estudio no experimental	Población censal: 594 estudiantes del IESTP "Daniel Villar" de Caraz que ingresaron

<p>Instituto de Educación Superior Tecnológico “Daniel Villar” de Caraz, que serán relevantes y permitirán descubrir los factores asociados que influyen en la deserción?</p>	<p>los datos operativos de los estudiantes, mediante el procesamiento de los datos operativos (personales y académicos) de los estudiantes que abandonaron voluntariamente sus estudios, siguiendo los lineamientos del proceso KDD de descubrimiento del conocimiento en bases de datos, aplicando técnicas</p>	<p>estudiantes con técnicas de minería de datos educacional, permitirán descubrir algunos de los factores que influyen en la deserción estudiantil en el Instituto de Educación Superior Tecnológico Público “Daniel Villar”. Ho: Los patrones de deserción estudiantil</p>			<p>Colegio de procedencia Tipo de colegio Modalidad del colegio Distrito donde está ubicado el colegio Provincia donde está ubicado el colegio Departamento donde está ubicado el colegio Diferencia entre el año de ingreso al instituto y el</p>	<p>Esquema: M – O -> P M: Población censal O: Patrones de deserción estudiantil P: Propuesta</p>	<p>entre los años 2010 al 2013. Técnicas e instrumentos: Técnicas: Técnicas de análisis de datos cuantitativos. Instrumentos: Repositorio de datos consistente en una base de datos para almacenar los datos secundarios</p>
---	--	---	--	--	---	---	---

	de Minería de Datos Educativa, estableciendo un procedimiento para el estudio de la deserción estudiantil en los institutos superiores tecnológicos del país.	descubiertos mediante el procesamiento de los datos operativos de los estudiantes con técnicas de minería de datos educativa, no permitirán descubrir algunos de los factores que influyen en la deserción estudiantil en el Instituto de Educación Superior Tecnológico			año de egreso del colegio Sexo Estado civil Trabaja Nivel académico	recolectados de los documentos oficiales de la institución. Métodos de análisis de datos:
	Específicos: • Recolectar los datos secundarios para la investigación, mediante la			Datos académicos	Promedio final del estudiante en el colegio Año de ingreso Carrera profesional Ciclo de abandono Promedio por ciclo	Técnicas estadísticas descriptivas y técnicas de minería de datos educativa: reglas de asociación y árboles de

	<p>aplicación de los procesos de selección, pre procesamiento y transformación de datos definidos en el proceso de descubrimiento del conocimiento en bases de datos (KDD), sobre los datos personales y académicos de los estudiantes, teniendo como fuente de datos</p>	<p>Público “Daniel Villar”.</p>	<p>Variable dependiente: patrones de deserción estudiantil.</p>	Naturaleza	Utilidad		<p>decisión, disponibles en el software SPSS Modeler.</p>
				Confiabilidad	Porcentaje de confianza del modelo		

	<p>secundarios las carpetas del postulante y las actas consolidadas de evaluación semestral de los estudiantes que ingresaron durante los años 2010 al 2013 y establecer un repositorio de datos para su almacenamiento electrónico.</p> <ul style="list-style-type: none"> • Utilizar técnicas de Minería de Datos Educativa para el descubrimiento de patrones de 					
--	--	--	--	--	--	--

	<p>deserción estudiantil, e interpretarlos y evaluarlos para identificar los factores relacionados con los datos operativos de los estudiantes, que influyen en la deserción estudiantil.</p> <ul style="list-style-type: none"> • Establecer un procedimiento para el estudio de la deserción estudiantil en los institutos superiores 					
--	--	--	--	--	--	--

	tecnológicos del país, utilizando los datos personales y académicos de los estudiantes, siguiendo los lineamientos del proceso KDD de descubrimiento del conocimiento en bases de datos, aplicando técnicas de Minería de Datos Educativa.						
--	--	--	--	--	--	--	--