

Actividad 04

Minería de datos

Nombre y código de los integrantes:

Alvarez Gutiérrez David Alejandro 217294016

Padilla Martin Juan Pablo 217294261

Ruiz Ortiz Cesar Adrian 208020817

Carrera: Informática

Profesor: Israel Román Godínez

Sección D01

Ciclo 2020A



A. Objetivos del negocio

1. Antecedentes

La recopilación de información referente a las causales de la Deserción permite identificar una gran diversidad de resultados, demostrándose la influencia de aspectos tales como: sociales y económicos, cómo la como la ocupación y nivel educacional de los padres, la valoración y expectativas educativas de los jóvenes y el compromiso con la meta y los objetivos futuros; situación financiera del estudiante y su familia, el trabajo remunerado del estudiante y las responsabilidades familiares, factores motivacionales, que provocan déficits cognitivos, valores del arancel, acceso a préstamos y becas, ubicación geográfica, clase y raza de los estudiante.

2. Objetivo comercial o de investigación del proyecto

El objetivo de esta investigación fue analizar la deserción de los estudiantes mediante técnicas de minería de datos y obtener un modelo que fuese capaz de clasificar estudiantes desertores a partir de los datos socioeconómicos y académicos de los estudiantes de carreras de pregrado en la Universidad Arturo Prat, Chile.

3. Criterios de rendimiento

Mejorar la gestión de la deserción en la institución resulta de vital importancia, dado que esto repercutirá en una mejora de la gestión financiera.

B. Valoración de la situación

1. Inventario de recursos

Recursos de la universidad
Recursos perdidos por el abandono escolar
Recursos destinados al objetivo
Recursos de los alumnos

2. Requisitos, supuestos y restricciones minería de datos para comprender

Minería de datos para comprender e intentar dar explicación al fenómeno de la deserción estudiantil, encontrándose que las principales tendencias en este tema van por el camino de la utilización de una metodología que guíe el proceso de minería.

3. Riesgos y contingencias

cada estudiante que deserta paga alrededor de \$2 000 000 (pesos chilenos) al año, la universidad tiene una pérdida por concepto de arancel anual de \$510 000 000, la cual representa un 29% de los ingresos totales.

4. Terminología

5. Análisis de costo/beneficio

Plantea el objetivo de tomar como herramienta principal la minería de datos para poder estudiar las variables que inciden en la deserción de primer año y generar un modelo que permita clasificar a los estudiantes con mayor riesgo de deserción, para de esta forma enfocar de mejor manera los esfuerzos y recursos asociados a la disminución de esta.

C. Objetivos de la Minería de datos

1. Metas de la minería de datos

Intentar resolver el problema de la deserción estudiantil, partiendo de los datos del usuario asociados a inducirlos a esta situación.

2. Criterio de rendimiento

Comprensión del negocio, Comprensión de los datos, Muestra, Preparación de los datos, Eliminación de valores nulos, Valores perdidos, Transformaciones de atributos y Selección de variables.

D. Plan de proyecto

1. Plan del proyecto

Dadas las características del problema planteado y los resultados encontrados en las investigaciones antes descritas, es que la solución de la problemática se define como una tarea de clasificación, en el caso del problema planteado esta variable será la deserción, que identificará el estado de un estudiante como: "Deserta" o "Permanece". Algunas de las mejores técnicas para el estudio de la deserción de nivel de pregrado en el contexto de los algoritmos de clasificación corresponden a los árboles de decisión, métodos bayesianos y redes neuronales.

2. Valoración de herramientas técnicas

Uso de la metodología como es CRISP-DM, acompañada por el uso de técnicas de clasificación (dependiendo del tipo de solución buscada), como son los árboles de decisión, las redes neuronales y los métodos bayesianos.