

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA EN INFORMÁTICA

**ANÁLISIS DE LA DESERCIÓN DE ALUMNOS DE UNA
UNIVERSIDAD, MEDIANTE ÁRBOLES DE CLASIFICACIÓN**

ANDRÉS ALEJANDRO RAMOS MAGNA

TESIS DE GRADO
MAGÍSTER EN INGENIERÍA INFORMÁTICA

Diciembre 2011



PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
FACULTAD DE INGENIERÍA
ESCUELA DE INGENIERÍA EN INFORMÁTICA

ANÁLISIS DE LA DESERCIÓN DE ALUMNOS DE UNA UNIVERSIDAD, MEDIANTE ÁRBOLES DE CLASIFICACIÓN

ANDRÉS ALEJANDRO RAMOS MAGNA

Profesor Guía: **JOSE LUIS MARTÍ LARA**

Programa: **Magíster en Ingeniería en Informática**

Diciembre 2011

Resumen

La deserción de la educación superior en Chile es un tema recurrente debido a las problemáticas que causa. Su evaluación e investigación ha sido realizada principalmente, por medio del uso de métodos tradicionales (encuestas), donde los resultados no han sido del todo satisfactorios, pese al impacto social y organizacional que tiene el fenómeno.

La minería de datos es una herramienta que podría proporcionar el conocimiento de las acciones de los alumnos, que han sido almacenados por años en las bases de datos institucionales, permitiendo estudiar la información generada por los alumnos desertores.

Este trabajo realiza un análisis de minería de datos utilizando los registros académicos y arancelarios de los alumnos y las causas que pudiesen provocar que un alumno deserte a sus estudios, proponiendo un indicador inicial de riesgo, que permita evitar esto.

Palabras Claves: Minería de Datos, Inteligencia de Negocios, Educación Superior, Deserción Estudiantil.

Abstract

University desertion in Chile is a recurring issue, because of to the problems it causes. Its evaluation and research has been done mainly by using traditional methods (surveys), where the results have not been fully satisfying, despite the social and organizational impact of the phenomenon.

Data mining is a tool that could provide knowledge of the actions of the students, which have been stored for years in corporate databases, allowing to study the information generated by deserters students.

This paper performs a data mining analysis using academic records and tariffs of students and the reasons that could cause in a student deserts to their studie, proposing an initial indicator of risk, which avoids this.

Keywords: Data Mining, Business Intelligence, University, Student Desertion.

Índice general

| | |
|---|-----------|
| 1. Introducción | 1 |
| 2. Marco Referencial | 3 |
| 2.1. La educación superior en Chile | 3 |
| 2.2. El problema de la Deserción Universitaria | 6 |
| 2.3. Consecuencias de la repitencia y la deserción | 7 |
| 2.4. Evaluaciones realizadas en el área | 8 |
| 2.4.1. Evaluaciones con Métodos Tradicionales | 9 |
| 2.4.2. Evaluaciones Basados en Técnicas de Minería de Datos. | 10 |
| 2.5. Definición del Problema | 12 |
| 2.6. Objetivos | 13 |
| 2.6.1. Objetivo General | 13 |
| 2.6.2. Objetivos Específicos | 13 |
| 2.7. Metodología | 13 |
| 3. Estado del Arte | 15 |
| 3.1. Inteligencia de Negocios | 15 |
| 3.1.1. Componentes de una Implementación de Inteligencia de Negocios | 16 |
| 3.1.2. Beneficios de la Inteligencia de Negocios | 17 |
| 3.2. <i>Data Warehouse</i> | 17 |
| 3.3. Minería de Datos | 19 |
| 3.3.1. Tipos de conocimiento | 19 |
| 3.3.2. Proceso de búsqueda del conocimiento | 21 |
| 3.3.3. Técnicas de Minería de Datos | 23 |
| 3.3.4. Definición de algoritmos Utilizados | 25 |
| 3.3.5. Herramientas para la Minería de Datos | 31 |
| 3.4. Proceso de Análisis Jerárquico | 33 |
| 4. Preprocesamiento de los Datos | 35 |
| 4.1. Reconocimiento del Dominio y los Usuarios | 35 |
| 4.2. Análisis Exploratorio General de los Datos de Deserción | 40 |
| 4.3. Conclusiones | 50 |
| 5. Análisis de las Carreras | 52 |
| 5.1. Introducción | 52 |
| 5.2. Análisis de la deserción en la Carrera de Ingeniería en Construcción | 53 |
| 5.2.1. Análisis Exploratorio | 53 |
| 5.2.2. Minería de Datos | 61 |
| 5.3. Análisis de la Deserción en otras carreras | 81 |
| 5.3.1. Carrera de Arquitectura | 82 |

| | |
|---|------------|
| 5.3.2. Carrera de Auditoria (Diurno) | 86 |
| 5.3.3. Carrera de Enfermería | 89 |
| 5.4. Conclusiones | 93 |
| 6. Propuesta de Indicador de Riesgo de Deserción | 95 |
| 6.1. Carrera de Ingeniería en Construcción | 95 |
| 6.2. Carrera de Arquitectura | 97 |
| 6.3. Carrera de Auditoria (Diurno) | 98 |
| 6.4. Carrera de Enfermería | 100 |
| 7. Conclusiones | 102 |
| Bibliografía | 105 |
| A. Anexo 1 | 107 |
| A.1. Análisis Carrera de Arquitectura | 107 |
| A.1.1. Análisis Exploratorio | 107 |
| A.1.2. Clasificación | 109 |
| A.1.3. Evaluación de la Carrera con la aplicación de Tasa en Semestres | 112 |
| A.1.4. Minería de datos en las evaluaciones de los alumnos de Arquitectura. | 114 |
| B. Anexo 2 | 119 |
| B.1. Análisis Carrera de Auditoria (Diurno) | 119 |
| B.1.1. Análisis Exploratorio | 119 |
| B.1.2. Evaluación de la Carrera con Algoritmos | 122 |
| B.1.3. Minería de datos en las evaluaciones de los alumnos de Auditoría. | 125 |
| C. Anexo 3 | 128 |
| C.1. Análisis Carrera de Enfermería | 128 |
| C.1.1. Análisis Exploratorio | 128 |
| C.1.2. Evaluación de la Carrera con Algoritmos | 130 |

Índice de tablas

| | |
|---|----|
| 2.1. Crecimiento de la oferta académica en Chile [4]. | 4 |
| 2.2. Estado de la acreditación de instituciones de educación en Chile [13]. | 5 |
| 2.3. Comparativa de consultas utilizadas en el sector privado que podrían aplicarse en la educación superior | 12 |
| 4.1. Atributos de alumnos desertores | 37 |
| 4.2. Atributos del archivo de alumnos desertores y sus notas | 38 |
| 4.3. Atributos del archivo de alumnos desertores y sus beneficios | 38 |
| 4.4. Descripción atributos proceso de postulacion alumnos desertores | 39 |
| 5.1. Atributos del conjunto de datos con tasas | 66 |
| 5.2. Atributos para el set de datos de clasificacion de la deserción con las evaluaciones de Ingeniería en construcción | 71 |

Índice de figuras

| | |
|--|----|
| 2.1. Cobertura de la educación superior | 6 |
| 2.2. Comparativa entre alumnos matriculados Vs. alumnos titulados | 9 |
| 3.1. Modelo integral de una solución BI [20]. | 16 |
| 3.2. Ejemplo de modelo estrella de almacén de datos | 18 |
| 3.3. Curva costo/beneficio vs Características de una implementación de <i>Data Warehouse</i> [15]. | 19 |
| 3.4. Taxonomía de Técnicas de Minería de Datos | 20 |
| 3.5. Etapas de la extracción de conocimiento | 21 |
| 3.6. Etapas de la extracción de conocimiento | 22 |
| 3.7. Aprendizaje y Clasificación en la generación de los árboles | 26 |
| 3.8. Primera Partición | 28 |
| 3.9. Segunda Partición | 28 |
| 3.10. Tercera Partición | 29 |
| 3.11. Representación de los nodos en el árbol | 29 |
| 3.12. PseudoCódigo del Algoritmo C4.5 [14] | 30 |
| 3.13. Listado de Herramientas por cantidad de usuarios | 31 |
| 3.14. Listado de Herramientas Por cantidad usuarios menor | 32 |
| 3.15. Interfaz de la aplicación WEKA | 32 |
| 3.16. Interfaz de la aplicación Orange Canvas | 33 |
| 4.1. Cantidad de Años Cursados por los alumnos de la carrera de construcción | 41 |
| 4.2. Cantidad de Alumnos por Años Cursados por los alumnos | 41 |
| 4.4. Cantidad de alumnos por promedio de notas | 41 |
| 4.3. Número de alumnos por periodo de deserción | 42 |
| 4.5. Cantidad de alumnos por promedio de nota | 42 |
| 4.6. Grafico de distribución de datos y atributos principales | 43 |
| 4.7. Estado de asignación de becas alumnos desertores | 43 |
| 4.8. Estado de asignación de becas alumnos desertores | 44 |
| 4.9. Porcentaje de asignación total de credito asignado | 44 |
| 4.10. Cluster beneficios utilizando Simple K-Means | 45 |
| 4.11. Dispersión de los datos de las postulaciones | 45 |
| 4.12. Cantidad de Alumnos por Puntaje ponderado | 46 |
| 4.13. Cantidad de Alumnos por Promedio NEM | 46 |
| 4.14. Rama y grupo de dependencia de los colegios de procedencia | 47 |
| 4.15. Cluster con simple K-Means a las postulaciones | 48 |
| 4.16. Outlier postulación alumnos | 48 |
| 4.17. Outliers de los beneficios de los alumnos | 49 |
| 4.18. Caso anómalo en la cantidad de años cursados por los alumnos | 49 |
| 5.1. Cantidad de Desertores por Semestre | 54 |

| | |
|---|----|
| 5.2. Cantidad de alumnos que NO desertan a la carrera por semestre | 54 |
| 5.3. Cantidad de desertores por años | 55 |
| 5.4. Cantidad de desertores por estados | 56 |
| 5.5. Promedio de Notas Alumnos Desertores | 56 |
| 5.6. Cantidad de asignaturas aprobadas | 57 |
| 5.7. Cantidad de desertores por máxima vez cursada una asignatura | 57 |
| 5.8. Cantidad de asignaturas reprobadas de los alumnos desertores | 58 |
| 5.9. Tasa de avance total de los alumnos de la carrera A | 60 |
| 5.10. Tasa de avance según año de los alumnos de la carrera A | 60 |
| 5.11. Comparativa tasa de aprobación alumnos | 60 |
| 5.12. Créditos asignados a alumnos desertores de la Ingeniería en Construcción | 61 |
| 5.13. Ranking de atributos más relevantes de Ingeniería en Construcción | 62 |
| 5.14. Estructura árbol ID3 de la clasificación | 62 |
| 5.15. Clasificación desertores con máxima asignatura por tercera | 63 |
| 5.16. Matriz de Confusión del clasificador por árbol CART | 63 |
| 5.17. Comparativa entre matrices de confusión de la clasificación CART y C4.5 | 64 |
| 5.18. Cantidad de asignaturas cursadas por primera | 64 |
| 5.19. Semestre que cursa el alumno al ser desertor | 65 |
| 5.20. Árbol de clasificación de alumnos desertores sin alumnos regulares | 65 |
| 5.21. Raiz arbol CART carrera A | 67 |
| 5.22. Clasificación de avance alumnos desertores | 68 |
| 5.23. Descripción con algoritmo CN2 de Ingeniería en Costrucción | 68 |
| 5.24. Casos en que se presenta baja tasa de avance total carrera en alumnos desertores. | 69 |
| 5.25. Clasificacion con tasa de avance y credito bajo el 10 %. | 69 |
| 5.26. Clasificacion con tasa de avance y credito bajo el 10 % caso dos. | 69 |
| 5.27. Clasificacion arbol J48 con WEKA. | 70 |
| 5.28. Comparativa de las Matrices de Confusión | 71 |
| 5.29. Cantidad de asignaturas cursadas alumnos desertores. | 72 |
| 5.30. Cantidad de desertores por Profesores dictan asignaturas. | 73 |
| 5.31. Distribución de registros según descripción de las evaluaciones. | 73 |
| 5.32. Clasificación de alumnos Ingeniería en Construcción y la reprobación de asignaturas | 74 |
| 5.33. Clasificación de alumnos en la carrera A y la descripción de la deserción por el profesor | 75 |
| 5.34. Cantidad de alumnos desertores por semestre en la carrera de Ingeniería en Construcción. | 76 |
| 5.35. tasa de avance en semestre de los alumnos de la carrera Ingeniería en Construcción. | 76 |
| 5.36. Clasificación con tasa de avance y crédito del alumno. | 77 |
| 5.37. Clasificación con tasa de avance y crédito del alumno. | 77 |
| 5.38. Poda clasificacion alumnos desertores y la tasa de avance segun semestre | 78 |
| 5.39. Poda clasificacion alumnos desertores y la tasa total de avance | 78 |
| 5.40. Poda clasificacion alumnos desertores y la tasa de avance | 79 |
| 5.41. Poda clasificacion alumnos desertores y la tasa de avance | 80 |
| 5.42. Poda clasificacion alumnos desertores y la tasa de aprobación | 80 |
| 5.43. Poda clasificacion alumnos desertores y la tasa de avance | 81 |
| 5.44. Ajuste atributos clsificación y poda | 82 |
| 5.45. Poda clasificacion alumnos desertores y la tasa de avance total | 83 |
| 5.46. Poda clasificacion alumnos desertores y la tasa de avance total carrera | 83 |
| 5.47. Poda clasificacion alumnos desertores y la tasa de avance segun año mejor al 10 % | 84 |
| 5.48. Poda clasificacion alumnos desertores y asignacion de crédito menor al 10 % | 84 |
| 5.49. Poda clasificacion alumnos desertores y que tienen asignaturas por segunda | 85 |
| 5.50. Poda clasificacion alumnos desertores y tasa de avance del alumno | 85 |
| 5.51. Poda clasificación alumnos desertores y tasa de aprobación por primera | 86 |
| 5.52. Alumnos desertores de carrera c y el porcentaje de asignación de créditos | 86 |

| | |
|---|-----|
| 5.53. Alumnos desertores de carrera c y el porcentaje de asignación de créditos | 87 |
| 5.54. Poda clasificacion alumnos desertores y tasa de Aprobación del alumno | 87 |
| 5.55. Poda clasificacion alumnos desertores y tasa de Avance del alumno | 88 |
| 5.56. Poda clasificación alumnos desertores y tasa de Avance según semestre del alumno | 88 |
| 5.57. Poda clasificación alumnos desertores y tasa de Aprobación por primera del alumno | 89 |
| 5.58. Reevaluación Enfermería con nuevas reglas en alumnos con créditos | 90 |
| 5.59. Reevaluación Enfermería con nuevas reglas en alumnos segun año de ingreso | 90 |
| 5.60. Reevaluación Enfermería con nuevas reglas en alumnos segun avance total. | 91 |
| 5.61. Reevaluación Enfermería con nuevas reglas en alumnos según Tasa de Aprobación. | 91 |
| 5.62. Reevaluación Enfermería con nuevas reglas en alumnos según Tasa de Aprobación. | 92 |
| 5.63. Reevaluación Enfermería con nuevas reglas en alumnos con Tasa avance según año. | 92 |
| 5.64. Influencia de los atributos en la deserción. | 94 |
| | |
| 6.1. matriz de pesos atributos Construcción | 96 |
| 6.2. matriz de ajuste de pesos atributos Construcción y promedios obtenidos | 96 |
| 6.3. matriz de pesos atributos Arquitectura | 97 |
| 6.4. matriz de ajuste de pesos atributos Arquitectura y promedios obtenidos | 97 |
| 6.5. Algoritmo aplicado a la carrera | 98 |
| 6.6. matriz de pesos atributos Auditoría | 99 |
| 6.7. matriz de ajuste de pesos atributos Auditoria y promedios obtenidos | 99 |
| 6.8. matriz de pesos atributos Enfermería | 100 |
| 6.9. matriz de ajuste de pesos atributos Enfermería y promedios obtenidos | 100 |
| | |
| A.1. Distribucion de datos según atributos | 108 |
| A.2. Alumnos desertores según tasa de avance | 108 |
| A.3. Crédito asignados a los alumnos desertores | 109 |
| A.4. Clasificacion alumnos desertores debido a diferencias en el promedio de notas. | 110 |
| A.5. Clasificacion alumnos desertores baja tasa de avance. | 110 |
| A.6. Clasificacion alumnos desertores usando ID3 con WEKA | 111 |
| A.7. Comparación matrices de confusión. | 111 |
| A.8. Curva ROC árbol CART | 112 |
| A.9. Cantidad de semestres que cursa un alumnos desertor en la carrera de Arquitectura. | 112 |
| A.10.Cantidad de alumnos por estados en la carrera de Arquitectura. | 113 |
| A.11.Ranking atributos carrera B. | 113 |
| A.12.Clasificacion importancia avance total carrera carrera de Arquitectura. | 114 |
| A.13.Distribucion atributos de las notas de arquitectura carrera B. | 115 |
| A.14.Cantidad de desertores por asignatura en la carrera de Arquitectura. | 115 |
| A.15.Cantidad de desertores por asignatura en la carrera de Arquitectura. | 116 |
| A.16.Cantidad de desertores por profesor en la carrera de Arquitectura. | 116 |
| A.17.Cantidad de alumnos en cátedra carrera de Arquitectura. | 117 |
| A.18.Tasa de aprobación por asignatura carrera de Arquitectura. | 117 |
| A.19.Caso Clasificación deserción dependiendo de la asignatura en Arquitectura. | 118 |
| | |
| B.1. Cantidad de alumnos desertores según tasa total de avance en la carrera | 120 |
| B.2. Cantidad de alumnos desertores según según porcentaje de crédito asignado | 120 |
| B.3. Cantidad de alumnos desertores según el tipo de dependencia de los colegios. | 121 |
| B.4. Cantidad de alumnos desertores por tramo de selección. | 121 |
| B.5. Ranking de los atributos utilizados en la carrera de Auditoria (Diurno). | 122 |
| B.6. Clasificación importancia de los semestres cursados. | 122 |
| B.7. Clasificación importancia tasa de avance según año. | 123 |
| B.8. Clasificación aplicacndo algoritmo C4.5. | 123 |

| | |
|--|-----|
| B.9. Clasificación Efecto de la tasa de reprobación | 124 |
| B.10. Clasificación Efecto de la tasa de reprobación | 124 |
| B.11. Matrices de confusión de la clasificación de Auditoría (Diurno) | 125 |
| B.12. Cantidad de desertores por asignaturas | 125 |
| B.13. Cantidad de desertores por profesor | 126 |
| B.14. Clasificación alumnos desertores y la tasa de aprobación. | 126 |
| B.15. Clasificación notas alumnos desertores con algoritmo C4.5. | 127 |
| | |
| C.1. Cantidad de alumnos desertores por estados | 129 |
| C.2. Cantidad de alumnos desertores por avance total carrera | 129 |
| C.3. Distribución de datos por atributos relevantes | 130 |
| C.4. Ranking atributos relevantes para la clasificación | 130 |
| C.5. Clasificación alumnos desertores y el avance total Carrera | 131 |
| C.6. Clasificación alumnos desertores y la tasa de aprobación | 132 |
| C.7. Clasificación Alumnos desertores y la tasa de aprobación | 132 |
| C.8. Clasificación Alumnos desertores utilizando algoritmo ID3 en WEKA | 133 |
| C.9. Matrices de clasificación de la carrera de Enfermería | 133 |
| C.10. Cantidad de alumnos desertores por semestre asignatura | 134 |
| C.11. Cantidad de alumnos desertores por cantidad de alumnos en una asignatura | 134 |
| C.12. Clasificación con CART y la cantidad de alumnos en cátedra | 135 |

Capítulo 1

Introducción

Los inicios de la educación superior en Chile se remontan al año 1622, con la creación de la Universidad de Santo Tomás, la cual mantuvo sus puertas abiertas hasta 1747. Desde aquella época, se realizaron múltiples intentos en la evolución de la educación superior, donde principalmente se enseñaba arte y teología, fundamentado en las estructuras impuestas por la colonia española [13].

No fue hasta la independencia de nuestro país, donde nacen nuevos proyectos de educación libre y se crea la Universidad Real de San Felipe, quien mantiene sus puertas abiertas hasta que se refunda en el año 1843 como Universidad de Chile. Pronto la idea dio lugar a nuevas iniciativas, como la creación de la Pontificia Universidad Católica de Chile en el año 1888 y las primeras universidades en provincias como por ejemplo la Universidad Católica de Valparaíso, actualmente Pontificia, fundada en 1928 y la Universidad Católica del Norte, fundada en 1956 [13].

La estructura inicial mantuvo el concepto de Universidades Nacionales, que permiten mantener un control de calidad en las estructuras de los distintos programas impartidos; no fue hasta el año 1981 que una nueva legislación cambió el perfil de la educación superior del país, cuando se decretó la división de Universidades Nacionales en Regionales.

La Ley de 1981, junto con la regionalización de las universidades, dio la libertad para crear establecimientos postsecundarios y abrir nuevas instituciones de educación privada, lo que generó en el futuro una gran expansión y crecimiento en el negocio con una múltiple oferta de carreras profesionales y distintos grados educacionales.

Por lo anterior, existe una gran diversidad de formas y estilos para realizar o entregar la educación superior, lo que llevó al gobierno a discutir en el Congreso Nacional el proyecto de ley que establece un Sistema Nacional de Aseguramiento de la Calidad. La calidad de la educación superior es un tema recurrente en las discusiones del Consejo Superior Educación (CSE)[7] [11] [1], que han descrito una serie de indicadores necesarios para evaluar la calidad de cada programa de educación, siendo uno de ellos la deserción de los estudiantes [7].

Lo anterior se hace más relevante cuando estos factores son evaluados por las entidades de acreditación, donde además de evaluar aquellos que permitan estudiar la calidad de un programa, se identifican los riesgos y problemas (como son la deserción de los estudiantes) y cuáles son las medidas que se están aplicando para enfrentar la problemática.

Existen múltiples estudios de la deserción de los estudiantes de la educación superior, basados en encuestas académicas con los alumnos e integrantes de las comunidades educativas, que permiten aproximar los principales motivos de la ocurrencia de la deserción. Pero son pocos los intentos que se han realizado en evaluar los registros de los estudiantes e intentar identificar ciertas asociaciones o grupos que puedan describir un indicador de riesgo de los alumnos, ante el hecho de la deserción de la educación superior [27] [26].

Realizar una evaluación de los datos de los estudiantes, podría llegar a ser una tarea titánica utilizando técnicas tradicionales, debido a la gran cantidad de información que poseen la mayoría de las instituciones; para esto se propone utilizar técnicas de minería de datos en los registros académicos de la universidad.

La minería de datos se define como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos [19]. Lo anterior permite estudiar el comportamiento de los alumnos en sus propios registros identificando como se describen los alumnos desertores de la universidad en distintas carreras.

A lo largo de esta Tesis se muestra la evaluación de la deserción de la educación superior en 4 carreras de una universidad tradicional de la zona de Valparaíso, utilizando las técnicas de minería de datos, para describir el comportamiento del alumno y detectar de forma anticipada con un rango de riesgo a quienes pueden llegar a ser posibles futuros desertores.

Este documento se encuentra dividido en los siguientes capítulos:

- **Capítulo 1:** se presenta la introducción del trabajo y el documento.
- **Capítulo 2:** se presenta un marco referencial que permite entender la expansión de las universidades, el acceso mayor que existe hoy en día a la educación superior y al mismo tiempo la problemática inherente a la deserción de los estudiantes.
- **Capítulo 3:** se entrega el estado del arte frente a las tecnologías existentes para realizar un análisis de inteligencia de negocios en esta área, sus herramientas y explicar los fundamentos principales del análisis utilizando minería de datos.
- **Capítulo 4:** se presenta el análisis exploratorio de los datos, con el que se ha realizado una evaluación preliminar de la información de los alumnos y los atributos principales que describen a los alumnos desertores de la Universidad.
- **Capítulo 5:** se realiza un análisis aplicando técnicas de minería de datos para evaluar la deserción de los alumnos en 4 carreras seleccionadas.
- **Capítulo 6:** se lleva a cabo una evaluación de los resultados obtenidos identificando el peso de cada uno de los atributos de las carreras según la importancia en la descripción de las reglas que permiten identificar un polinomio para identificar a los desertores con la aplicación de un algoritmo.

Capítulo 2

Marco Referencial

2.1. La educación superior en Chile

Desde el año 1842 y hasta el año 1973 el Gobierno Chileno jugó un rol central en la coordinación y promoción de la función y crecimiento de la educación superior. Esto estructuró a las instituciones de educación superior en un formato de organizaciones públicas en su gran mayoría [13].

La estructura y organización de las universidades se vio modificada en el año 1973 con el ingreso del gobierno militar, donde la cobertura disminuyó desde un 11 % en el mismo año y a un 7,5 % a principios de los 80. La nueva estructura estableció un control de las instituciones y provocó una reducción en algunas áreas, las que comenzaron a ser trasladadas a centros de investigación independientes.

Debido a las necesidades del país, y producto de una serie de cambios en la legislación de 1981, se genera un cambio de perfil de la educación superior, donde se realiza una división de las universidades nacionales en universidades regionales; además se crean nuevas estructuras con instituciones post-secundarias no universitarias: los centros de formación técnica (CFT) y los institutos profesionales (IP). Además se entregó la autorización amplia para la apertura de organizaciones privadas, así bajo este permiso se crearon tres modelos de universidades, los cuales son:

1. Universidades Tradicionales: es el conjunto de Universidades Estatales y particulares de carácter público que reciben aporte fiscal y conforman el Consejo de Rectores de las Universidades Chilenas. Se dividen en dos grandes grupos:
 - a) Universidades Tradicionales antiguas: son las 8 instituciones que existen desde antes de 1980; dos de ellas son estatales y seis privadas.
 - b) Universidades Derivadas: Son instituciones creadas desde 1981, a partir de las sedes regionales de las Universidades de Chile, Técnica del Estado actual USACH y la Pontificia Universidad Católica de Chile.
2. Universidades Privadas Autónomas: son aquellas universidades privadas creadas a partir de 1981, que han alcanzado su plena autonomía a través de los sistemas de examinación o acreditación; no reciben aporte fiscal para financiar sus actividades.

3. Universidades Bajo Supervisión: son aquellas universidades privadas que se encuentran actualmente supervisadas mediante los sistemas de examinación o acreditación. No han obtenido su plena autonomía y no reciben aporte fiscal.

No fue hasta fines de los 80, con la crisis económica de Chile ya superada, que las nuevas estructuras iniciaron su proceso de crecimiento, donde se incrementó la oferta existente de universidades privadas. Esto se refleja en la Tabla 2.1, que evidencia el crecimiento en la oferta de la educación superior entre los años 1980 y 2001.

| | 1980 | | 2001 | |
|-----------------------------------|------|---------|------|---------|
| | Nº | Matr. | Nº | Matr. |
| Universidades Consejo de Rectores | 8 | 119.000 | 25 | 213.663 |
| Universidades Privadas | 0 | 0 | 35 | 107.570 |
| Institutos Profesionales | 0 | 0 | 51 | 86.392 |
| Centros de Formación Técnica | 0 | 0 | 111 | 53.895 |

Tabla 2.1: Crecimiento de la oferta académica en Chile [4].

En base a este nuevo estilo de educación superior, se inició una etapa que se denominaría como “la nueva universidad”, que fue ratificada legalmente en marzo de 1989 mediante la Ley Orgánica Constitucional de Enseñanza (LOCE). Esta ley refuerza el concepto de “Libertad de Enseñanza”, donde el derecho para crear nuevas instituciones educativas se puede ejercer con la mínima restricción de no atentar contra la moral y las buenas costumbres, y la facultad de las personas para escoger la institución que deseen.

Por otra parte, a través de esta ley, el Estado cumple un rol de fomentar el desarrollo de la ciencia y la tecnología y velar por la fe pública, en si el gobierno describe los nuevos fines y funciones de la Universidad en los siguientes términos: “La Universidad es una Institución de Educación Superior de investigación, raciocinio y cultura que, en el cumplimiento de sus funciones, debe atender adecuadamente a los intereses y necesidades del país, al más alto nivel de excelencia”

Todo este cambio abrió nuevas perspectivas de la educación superior, donde se explica que su producto es la oferta de servicios en la formación de profesionales y técnicos en función de las demandas del mercado laboral, y su regulación se establece por los mismos mecanismos que rigen el mercado. Así, el principal problema que las Universidades tienen es que no quedaban sujetas a ningún tipo de regulación y la oferta se crea tomando en cuenta la demanda existente.

Debido a la gran libertad de las universidades, que ha permitido crear un sin número de carreras, en años recientes el Estado ha incorporado la evaluación y acreditación de las instituciones y carreras, lo que no estaba contemplado en la ley de los años 80. Frente a todo este esquema, el Estado le otorga a las universidades los siguientes deberes y/u obligaciones:

- Promover la investigación, creación, preservación y transmisión del saber universal y el cultivo de las artes y las letras;
- Contribuir al desarrollo espiritual y cultural del país, de acuerdo con los valores de su tradición histórica;
- Formar graduados y profesionales idóneos, con la capacidad y conocimientos necesarios para el ejercicio de sus respectivas actividades;

- Otorgar grados académicos y títulos profesionales reconocidos por el Estado; y
- En general, realizar funciones de docencia, investigación y extensión que son propias de la tarea universitaria.

El proceso de acreditación despertó en las distintas instituciones de educación una necesidad y un nuevo desafío por lo cual, pese a que se denomina como voluntaria, pasó a ser en los tiempos de hoy una necesidad, como lo demuestra la Tabla 2.2, donde se presenta el estado de acreditación institucional al año 2004, en el que un 74% de las instituciones acreditables participan del proceso.

| Estado de Participación | Institución | | | |
|----------------------------------|-------------|-----------------------|-----------------------------|-------|
| | Universidad | Instituto Profesional | Centro de formación Técnica | Total |
| Acreditadas | 11 | 2 | 0 | 13 |
| En Proceso de Acreditación | 21 | 1 | 0 | 22 |
| Presentadas a acreditación | 12 | 5 | 7 | 24 |
| No Participan | 10 | 9 | 2 | 21 |
| Aún en proceso de licenciamiento | 9 | 31 | 106 | 146 |
| Total | 64 | 48 | 117 | 229 |

Tabla 2.2: Estado de la acreditación de instituciones de educación en Chile [13].

En la actualidad existe una gran oferta de educación superior, según los indicadores de Ministerio de Educación (MINEDUC); para el año 2009 la oferta se encuentra compuesta por un total de 180 instituciones oficialmente reconocidas de naturaleza variada, donde 61 de estas corresponden a universidades, 45 a institutos profesionales y 74 a centros de formación técnica.

El crecimiento en la oferta de la educación superior ha permitido otorgar una mayor oportunidad a los jóvenes que pretenden estudiar alguna carrera específica: desde la reforma de los años 80, la cobertura de la educación superior ha logrado un fuerte incremento, de 189.151 alumnos en el año 1984 a 876.243 en el año 2009 según las fuentes del MINEDUC, siendo un incremento paulatino y constante cada año como se puede ver en la Figura 2.1.

El aumento en la cobertura y las matriculas de la educación superior, trae consigo una serie de necesidades, como es el financiamiento de las instituciones y sus alumnos. Para esto el país destina aproximadamente el 0,7% del producto geográfico bruto (PGB) en la educación superior. En 1990 el Estado invertía ciento diez mil millones de pesos a doscientos treinta y nueve millones de pesos, aumentando el aporte fiscal a la educación superior en este período de un 117%.

Frente a toda esta inversión y la libertad existente a principios de los 80 para crear instituciones y carreras, la preocupación de las personas y el Estado se centra hoy en la calidad de la educación superior. Para lo anterior, el 18 de febrero de 1999 se estableció, por Decreto del Ministerio de Educación, una comisión asesora encargada de proponer un diseño institucional para el aseguramiento de la calidad de la educación superior y la realización de procesos de acreditación. La función de aseguramiento de la calidad en este momento se encuentra en manos de la Comisión Nacional de Acreditación de Pregrado (CNAP) [4] - [1].

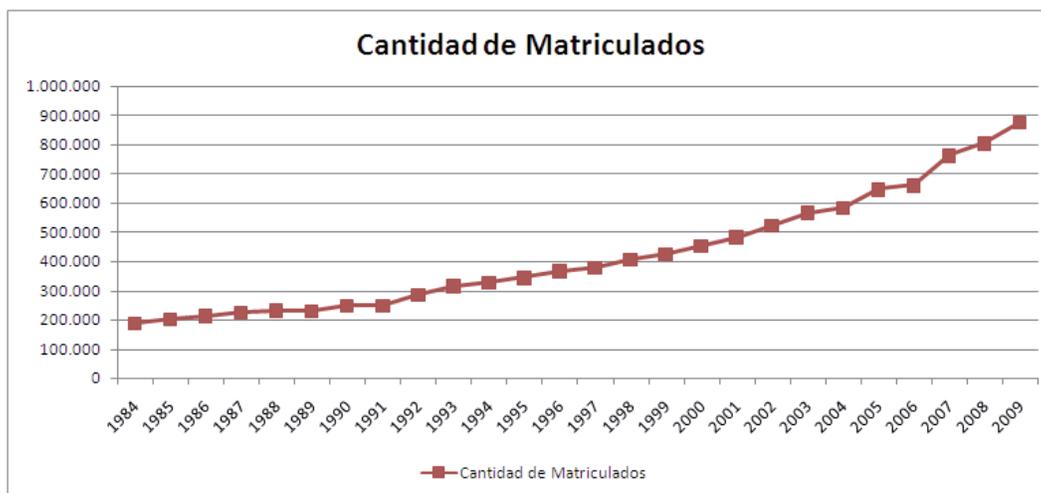


Figura 2.1: Cobertura de la educación superior

2.2. El problema de la Deserción Universitaria

Tal como se ha nombrado anteriormente, la calidad de la educación es un punto importante a la hora de evaluar una institución de educación superior; para esto el Estado, por medio de la CNAP, ha implementado una serie de normas, dentro de las cuales se encuentra la acreditación institucional.

Tal como lo explica la CNAP en su documento que transcribe las normas y procedimientos dispuestos en la Ley 20129 [6], la acreditación tiene como objeto fomentar y dar garantía pública de la calidad de las instituciones de educación superior. La misma Ley explica que, para cada una de las áreas de acreditación, existen deberes que las instituciones deben cumplir, siendo uno de ellos el que “Debe contar con políticas y mecanismos de aseguramiento de la calidad que se apliquen sistemáticamente en los diversos niveles institucionales, de modo eficiente y eficaz”.

Frente a la acreditación en el documento denominado Criterios Genéricos de Evaluación de Carreras Profesionales [6], se definen los principales criterios que debiesen cumplir las carreras para un aseguramiento de la calidad, siendo uno de ellos la efectividad del proceso de enseñanza; y de manera importante, define un segundo criterio de calidad como es el resultado del proceso formativo. Ambos se enfocan en la capacidad de la institución de lograr el éxito en el cumplimiento del proyecto educativo.

Por otra parte el Consejo Superior de Educación define como un criterio de la calidad de la educación, la progresión y logros de los estudiantes en universidades e institutos profesionales, donde indica que los criterios de admisión deben estar claramente establecidos, los que deben ser conocidos y apropiados para que los estudiantes tengan una razonable oportunidad de alcanzar un grado o título universitario. Además explica que la institución debe proporcionar adecuados servicios de apoyo académico a los alumnos, preservando el nivel de la enseñanza que imparte.

Frente a estos términos la deserción es un punto relevante para los procesos educativos; hoy, debido al aumento de la cobertura, son mas jóvenes que ingresan a una institución de educación superior, permitiendo que los estratos de ingresos medios-bajos y bajos representen ya a casi un tercio del total de los alumnos que

cursan una carrera universitaria. En muchos casos ellos pasan a ser la primera generación de su familia en llegar a la educación superior.

El mayor acceso a la educación superior, medido en términos absolutos o en tasas de cobertura, no garantiza el éxito de las políticas públicas, ya que no todos los estudiantes efectivamente concluyen sus programas de estudios. La problemática de esta decisión se encuentra en los costos privados y sociales, que se expresan en endeudamiento; incluso algunas universidades asumen pérdidas cuando han otorgado o avalado créditos [9].

Frente a lo anterior el Consejo Superior de Educación explica que el costo directo por deserción sería de \$47 mil millones de pesos anuales, que corresponde al 23,5 % de la inversión que realiza el Estado en educación superior, lo que se explica como una **falta de eficiencia** en el uso de los recursos.

2.3. Consecuencias de la repitencia y la deserción

De acuerdo a la UNESCO en su documento “Estudio sobre la Repitencia y la Deserción en las Universidades Chilenas”, las consecuencias se pueden reflejar en tres ámbitos:

- En lo personal, implica una condición de fracaso que afecta emocionalmente.
- En lo institucional, conlleva una disminución del rendimiento de la universidad y un aumento innecesario del número de estudiantes.
- En lo social, contribuye a generar inequidad y desequilibrios sociales, y desvirtúa los objetivos que la sociedad le ha entregado a la educación superior.

Por otra parte en el mismo estudio, al encuestar a cada uno de los involucrados en el proceso educativo, identifica que las consecuencias de la deserción son:

1. **Expresada por los Desertores:** incide directamente en su salud psíquica, en si los estudiantes expresan que los niveles de estrés fueron tan altos que prefirieron desertar. Además muchos alumnos indican que han requerido terapia posterior a la deserción para poder superar el problema; algunos ejemplos de los comentarios extraídos desde la Estudio realizado por la UNESCO son:
 - a.- “La carrera me produjo un estrés demasiado grande y fue una de las razones por las cuales me llevaron a tomar la decisión de abandonar”.
 - b.- “El haber desertado me afectó en la parte psíquica, estuve con depresión y como consecuencia estuve con licencia y terapia”.

Existe un grupo menor que se encuentra satisfecho de haber desertado de una carrera, pero el perfil de estos alumnos se encuentra en aquellos que cambiaron de carrera o institución, donde sus principales comentarios se centran en que les fue favorable haber renunciado a la carrera y haber optado por lo que deseaban estudiar; otros explican que el cambio de universidad les favoreció para un mejor estilo de vida, por lo cual ven como un gran beneficio el cambio de institución.

Un grupo menor de la investigación realizada por la UNESCO, visualiza a la deserción como un problema a futuro en su vida, debido a consecuencias laborales, al no contar con una carrera profesional o técnica.

2. **Expresada por las autoridades académicas:** el punto más aludido por las autoridades es el de los costos, debido a la necesidad de crear nuevos cursos para los repitentes o las implicancias directas que tiene en el presupuesto de la carrera, la deserción como tal. Por otra parte las autoridades aluden a temas como la imagen institucional, el daño a los indicadores en el plan estratégico institucional y la migración de estudiantes a otras casas de estudios donde finalmente terminan sus carreras.

Además expresan, que no sólo incide en la disminución del rendimiento académico de la universidad y en un incremento innecesario del número de alumnos, sino incide en la concepción de los procesos de enseñanza-aprendizaje.

Frente a lo anterior el Centro de Microdatos de la Universidad de Chile [9], identifica una serie de factores que pueden determinar la causa de deserción de los estudiantes a sus programas, dentro de los cuales se identifican los siguientes:

1. Falta de claridad vocacional de los estudiantes:
 - Alumno no ingresa a la carrera de su preferencia
 - Mal acceso a información y orientación
2. Situación económica
3. Rendimiento Académico:
 - Baja motivación, en general y debido a problemas vocacionales
 - Debilidades académicas previas
 - Dificultades en metodologías de enseñanza y aprendizaje
 - Insatisfacción con la carrera

Al evaluar la literatura del Consejo Superior de Educación se identifica que no existen muchos estudios que hagan referencia a la evaluación de la deserción de los estudiantes de la educación superior, por lo cual no hay claridad de los motivos que llevan a un alumno a desertar a una carrera. Con los estudios realizados por el Centro de Microdatos de la Universidad de Chile, expresan en su informe que dos de los principales motivos son el problema económico y la falta de claridad vocacional de los estudiantes, lo que se repite en algunos autores del Consejo Superior de Educación sin contar con información del avance académico u otras variables que puedan afectar a los alumnos.

2.4. Evaluaciones realizadas en el área

En torno a la deserción de los estudiantes en la educación superior, existen muy pocas evaluaciones; principalmente se han realizado trabajos utilizando como herramientas descriptivas encuestas hechas a los participantes del proceso educativo [9] [22] [26], como son los estudiantes y los académicos. Otras evaluaciones se han fundamentado en el estudio de los indicadores de titulación, lo que no permite evaluar posibles causas de la deserción [27].

Por otra parte, solo se han identificado algunos textos internacionales que han identificado a la minería de datos como una herramienta que permita describir las causales de la deserción de los estudiantes, donde al menos dos autores han entregado esfuerzos en estas áreas.

Los principales estudios realizados con distintas estrategias y herramientas tanto nacionales como internacionales, se presentan a continuación.

2.4.1. Evaluaciones con Métodos Tradicionales

2.4.1.1. Estimaciones sobre la Repitencia y Deserción en la Educación

Evaluación realizada por el Consejo Superior de Educación donde sus autores, Luis Eduardo Gonzalez y Daniel Uribe [27], realizan un estudio en los sistemas de selección de las universidades y la capacidad de retención de los alumnos; se centra en las estadísticas de los procesos, evaluando la eficiencia de titulación de los alumnos de pregrado.

El principal aporte de este documento se encuentra en los costos que el Estado pierde con la deserción de los alumnos de la educación superior, lo que asciende al 23,5% del total invertido en educación superior.

El estudio demuestra que la eficiencia en esta área por parte de las universidades es muy baja. Esto debido a las diferencias que ocurren en la cantidad de alumnos matriculados versus la cantidad de alumnos titulados, donde en el primer caso la cantidad de alumnos matriculados tiene un importante ascenso en los últimos años, aunque la cantidad de alumnos titulados se mantiene prácticamente constante. Esta situación se puede visualizar en la Figura 2.2, donde se presenta el gráfico de evolución de las matrículas y el total de titulados a nivel nacional.

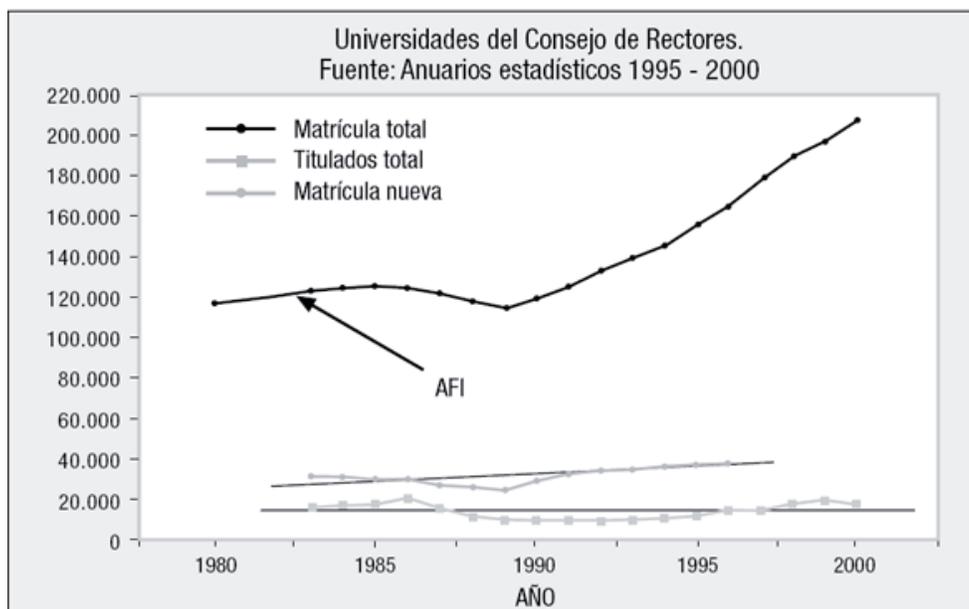


Figura 2.2: Comparativa entre alumnos matriculados Vs. alumnos titulados

2.4.1.2. Estudio sobre Causas de la Deserción Universitaria

Estudio realizado, utilizando entrevistas y encuestas como herramientas, por el Centro de Microdatos de la Universidad de Chile; es un estudio más directo con los participantes del proceso educativo, como los alumnos y las autoridades institucionales de las carreras universitarias, se encuentra más cercano a las causales de la deserción, en el describen las principales causales por las cuales el alumno puede desertar, siendo algunas de las principales: la falta de información previa al elegir una carrera y los problemas vocacionales.

Dentro de los resultados más importantes de este documento se encuentran la evaluación de propuestas existentes en universidades del país, como las enfocadas al mejoramiento académico de los alumnos.

2.4.1.3. Estudio sobre la Repitencia y Deserción en la Educación Superior Chilena

El estudio es realizado por IESALC-UNESCO, en el marco de su Programa “Observatorio de la Educación Superior en América Latina y el Caribe” [26], sobre el problema de la deserción de la educación superior en Chile; el proceso de recolección de la información se realiza por medio de encuestas realizadas a una muestra de alumnos de distintas carreras, profesores y autoridades en distinto nivel de las universidades participantes de la investigación.

El trabajo tiene como objetivo dimensionar la magnitud de la repitencia y deserción en Chile, analizando los factores que inciden en ella, sus implicancias y establecer posibles propuestas para enfrentarla, donde las consecuencias detectadas en esta evaluación, en los tres niveles personal, académico e institucional, pasan a ser el resultado más relevante del informe.

El informe describe los principales efectos de la deserción de la educación superior en los estudiantes, que destaca los altos niveles de frustración que involucran la deserción y repitencia. Por otra parte, muestra la forma en que se recepciona por las autoridades la deserción, principalmente causando un efecto de desmedro económico y de la calidad institucional.

2.4.2. Evaluaciones Basados en Técnicas de Minería de Datos.

Pese a la alta calidad de los trabajos realizados y presentados con anterioridad, ninguno de ellos muestra resultados fundamentados en un análisis desde el punto de vista de la información de los alumnos, en qué situación se encuentran y cómo se podría tomar una alerta temprana de su estado de posible desertor, para esto una alternativa es realizar un trabajo utilizando técnicas de minerías de datos.

Bajo esta estructura los intentos a nivel institucional son prácticamente nulos, a nivel internacional tan sólo se pueden identificar dos trabajos relacionados con la educación superior, pero ninguno de ellos realiza una evaluación de la deserción de los estudiantes utilizando las potencialidades de la herramientas de minería de datos.

Distinto es el caso a nivel nacional, donde al menos se han logrado identificar dos trabajos para la evaluación de la deserción utilizando minería de datos e identificar un indicador que permita mostrar el nivel de riesgo de un alumno.

2.4.2.1. Análisis del rendimiento académico en los estudios de Informática de la Universidad Politécnica de Valencia [29]

El estudio realiza una relación entre el rendimiento con las características socioeconómicas y académicas de los alumnos, que se obtienen al momento de su matrícula. La motivación principal del estudio realizado es identificar qué factores podrían explicar el éxito o fracaso académico de un estudiante.

Para realizar la evaluación se aplicaron métodos basados en minería de datos, siendo CRISP-DM el proceso utilizado durante la evaluación [29], fundamentalmente aplicado al programa de estudio de la carrera de informática en sus distintos niveles.

Para realizar la evaluación se utilizan principalmente dos técnicas; la primera de ellas basa en árboles de decisión y la segunda utiliza regresión multivariante, ambas para predecir el comportamiento de las variables existentes. Los modelos se generaron utilizando la herramienta SPSS Clementine v9.0. La evaluación permitió identificar algunas relaciones de interés para la carrera como por ejemplo:

- Los mejores rendimientos se obtienen por alumnos con estudios anteriores, donde no se evaluó si la procedencia correspondía a la misma Universidad o no.
- A mayor edad de ingreso, peor es el rendimiento.
- Alumnos que trabajan menos de 15 horas semanales tienen mejor rendimiento.
- Alumnos que estudian carreras de ingeniería tienen mejor rendimiento que las carreras técnicas.

Estos resultados son algunos ejemplos del potencial que presenta la minería de datos para la búsqueda del conocimiento, que han sido aplicados a una carrera en particular con una cantidad de información limitada.

2.4.2.2. *Data Mining Applications in Higher Education* [18]

Realizado por Jing Luan, artículo realizado para la empresa SPSS Inc, presenta las capacidades de la minería de datos y su aplicación en la educación superior. El documento muestra tres casos de estudios asociados.

El documento identifica que la minería de datos es una herramienta poderosa para la intervención académica, donde una universidad podría predecir si el 85 % de los estudiantes va a graduarse o no, además de identificar los alumnos con mayor riesgo para otorgar algún tipo de asistencia académica.

En si identifica que la minería de datos en los tiempos de hoy es una herramienta fundamental, donde muchas técnicas, existentes en el sector, privado, podrían ser utilizadas en el análisis de la educación superior, respondiendo a preguntas similares como las presentadas en la Tabla 2.3, donde muestra la equivalencia de las preguntas de inteligencia de negocios, en el sector privado vs la educación superior.

De forma general, el documento presenta las potencialidades que puede tener la minería de datos en la educación superior para la gestión institucional como también a la gestión académica.

2.4.2.3. Aplicar Minería de Datos a la Predicción de la Deserción Retención de Estudiantes [28]

Tesina realizada por Nadja Starocelsky, para optar al grado de magister en tecnologías de la información de la UTFSM, donde se realiza una evaluación de la deserción de los estudiantes, principalmente utilizando

| Preguntas en el sector privado | Equivalentes en la educación superior |
|--|--|
| ¿cuáles son mis mejores clientes? | ¿qué alumnos han tomado mayor cantidad de créditos? |
| ¿qué clientes son propensos a desertar a los competidores? | ¿qué alumnos se encuentran más propensos a cambiarse de universidad? |
| ¿quiénes son mis clientes más leales? | ¿quiénes son alumnos persistentes en la universidad? |
| ¿quiénes aumenten probablemente más sus compras? | ¿qué ex-alumnos podrían realizar grandes donaciones? |
| ¿quiénes son los visitantes frecuentes en mi sitio web? | ¿qué tipos de cursos atraen más estudiantes? |

Tabla 2.3: Comparativa de consultas utilizadas en el sector privado que podrían aplicarse en la educación superior

los antecedentes de ingreso y de notas de la Universidad Austral.

La evaluación entrega antecedentes iniciales para determinar las causas de la deserción estudiantil y que permitirían determinar de alguna forma el perfil de los alumnos con mayor riesgo de deserción. en especial en sus primeros años.

Se destaca de este trabajo el resultado obtenido donde ya se identifica que las causas de la deserción de los estudiantes no son genéricas y que depende mucho del perfil del alumno y de la carrera asociada.

2.5. Definición del Problema

En si todos los autores coinciden en que no existen muchos trabajos en torno al tema de la deserción pese al impacto y la relevancia que tiene, en las implicancias directas a las personas, instituciones y el Estado mismo. En la gran mayoría de los estudios y evaluaciones, los autores explican que no existen antecedentes para identificar las reales causas de la deserción [9].

Frente a esta situación existen estudios que intentan identificar la causa del problema y algunas medidas que puedan evitarlo, como son los estudios realizados por IESALC-UNESCO [26] y la Universidad de Chile [9], que han permitido identificar factores iniciales que inciden en la deserción de los estudiantes como son los problemas socioeconómicos.

Por otra parte, se debe destacar que la deserción de la educación superior es un punto relevante en los procesos de acreditación, según lo expuesto en el documento extraído de la Biblioteca del Congreso Nacional “estudio de los indicadores de calidad” [4], donde explican los criterios del Consejo Superior de Educación, específicamente el sexto criterio que hace referencia al proceso de enseñanza-aprendizaje, que dice en el punto 6.6. “La unidad debe realizar un análisis sistemático de las causas de deserción de los estudiantes y definir acciones tendientes a su disminución progresiva” [6].

Frente a este problema, no existen evaluaciones relevantes que permitan identificar la progresión de los alumnos en base a sus datos, registrados en los múltiples sistemas informáticos y bases de datos disponibles en las distintas instituciones de educación, permitiendo responder preguntas como:

- ¿es factible identificar el riesgo de la deserción de un estudiante previamente?

- ¿existe una correlación entre los factores de ingreso y la continuidad de los alumnos en el programa?
- ¿existen asignaturas que destaquen la factibilidad de deserción de los alumnos en una carrera?
- ¿realmente los alumnos con mal comportamiento de pago o aquellos que tienen muchos beneficios están más expuestos a la deserción en la educación superior?

2.6. Objetivos

2.6.1. Objetivo General

Evaluar y explicar la deserción de los alumnos de la educación superior en distintos programas de estudios de una universidad tradicional de la zona de Valparaíso, identificar las causas de este fenómeno y ayudar a determinar el riesgo de deserción de un alumno en su carrera.

2.6.2. Objetivos Específicos

Dentro de los objetivos específicos a cumplir en este trabajo se encuentran:

1. Identificar los programas de estudios que permiten describir a los tipos de estructuras existentes en la universidad, para evaluar las posibles combinaciones genéricas existentes en la institución.
2. Definir el conjunto de atributos que describan la deserción de los alumnos de las carreras, para evaluar los modelos a usar para identificar los atributos que impactan directamente en la deserción de los estudiantes.
3. Realizar pruebas y experiencias, utilizando técnicas de minería de datos, en base a distintos algoritmos existentes, que permitan definir patrones de la deserción de los estudiantes.
4. Realizar interpretación de resultados y entregar por medio de una representación visible los resultados obtenidos de la experiencias, para identificar claramente las causas de la deserción en cada carrera.
5. Definir un predictor inicial de riesgo de deserción del alumno, utilizando los resultados obtenidos en la evaluación de minería de datos, para contar a futuro con un medidor de alerta temprana de la deserción.

2.7. Metodología

La metodología a utilizar para este trabajo de título será una investigación descriptiva, la cual estará seguida de una experiencia de evaluación utilizando las técnicas de minería de datos; al finalizar se realizará una evaluación e interpretación de los resultados obtenidos. Para las distintas etapas se describe lo siguiente.

- **Definición del Problema:** se realizará un análisis preliminar de los datos para evaluar las características de la información y el set de atributos relevantes que podrían describir la deserción en cada una de las carreras.

- **Diseño de la solución:** enfocado principalmente para explicar el proceso, arquitectura y algoritmos a utilizar.
- **Implementación:** destinado principalmente a la explicación de las herramientas utilizadas para realizar los experimentos y la ejecución de éstos.
- **Pruebas y resultados:** esta sección contendrá los resultados más relevantes de los experimentos, donde se realizara una evaluación y análisis de los resultados obtenidos que permita obtener el conocimiento esperado utilizando proceso de análisis Jerárquico para una propuesta de algoritmo por medio de la jerarquización de los atributos más relevantes.

Capítulo 3

Estado del Arte

3.1. Inteligencia de Negocios

Desde la década del 70, las organizaciones han dirigido sus esfuerzos para crear principalmente sistemas que permitan administrar sus negocios, al igual que todas las organizaciones, las instituciones de educación también han adquirido o desarrollado una gran cantidad de sistemas que han generado un gran volumen de información.

En el documento “Data Warehousing and Business Intelligence” de la Universidad de Puerto Rico [12] se predijo lo siguiente: “En 1996, el uso de las soluciones de inteligencia de negocio cambiará dramáticamente de analistas dedicados a todos los gerentes y profesionales como la manera preferida de entender el negocio. En lugar de un pequeño número de analistas invirtiendo el cien por ciento de su tiempo analizando datos, todos los gerentes y profesionales invertirán el diez por ciento de su tiempo utilizando herramientas de inteligencia de negocio”.

La predicción no ha tardado mucho en ser una realidad, hoy en día las herramientas para la inteligencia de negocios son cada vez más necesarias para las nuevas gerencias, al momento de la toma de decisiones. La pregunta es qué es la Inteligencia de Negocios o Business Intelligence (BI).

La Universidad Nacional de La Rioja en su documento “Business Intelligence: Negocios Inteligentes para Empresas Inteligentes” [21] define la inteligencia de negocios como el proceso de analizar los bienes o datos acumulados en la empresa y extraer una cierta inteligencia o conocimiento de ellos. Gartner Group define la inteligencia de negocios, como la transformación de los datos de la compañía en conocimiento para obtener una ventaja competitiva. Por otro lado, Iberoamerica en su documento “*Business Intelligence*” [20] describe la Inteligencia de Negocios, desde un punto de vista más pragmático y asociado a las tecnologías de la información, como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales en información estructurada, para su explotación o análisis, permitiendo obtener conocimientos que dan soporte a la toma de decisiones sobre el negocio. La Figura 3.1 muestra el modelo integral de una solución BI.

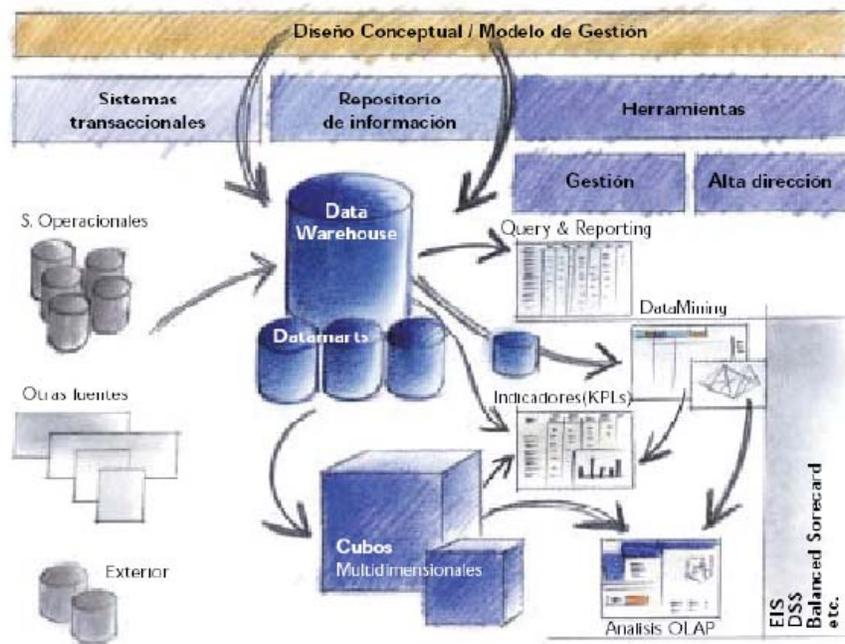


Figura 3.1: Modelo integral de una solución BI [20].

3.1.1. Componentes de una Implementación de Inteligencia de Negocios

Una solución de inteligencia de negocios se compone de las siguientes etapas [20]:

1. **Diseño Conceptual de los Sistemas:** Contiene los modelos a utilizar para contestar a las interrogantes existentes, formato y composición de los datos y procedencia de los datos más disponibilidad de la información.
2. **Construcción y alimentación del datawarehouse y/o datamarts:** se deben determinar las herramientas e implementaciones, donde se debe definir la depuración de la información en los data warehouse o en los data mart. Es aquí donde no siempre se obtienen buenos resultados y donde se pueden originar situaciones no deseadas.
3. **Herramientas de explotación de la información:** es el área con más avances, donde se pueden identificar algunas herramientas básicas que integra, combina y potencia cada etapa, dentro de las cuales se pueden nombrar:
 - *Query & Reporting*
 - Cuadro de mando analítico
 - Cuadro de mando integral o estratégico
 - OLAP
 - Datamining

Para la selección de las herramientas se deben considerar varios parámetros como qué información se necesita, para qué y para quién está dirigida y aspectos técnicos que permita ayudar a responder a las interrogantes o que sean restrictivos de la aplicación.

3.1.2. Beneficios de la Inteligencia de Negocios

Una solución de inteligencia de negocios otorga a la empresa un punto de acceso a la información crítica y de interés, respuestas oportunas ante las preguntas de la empresa y un máximo aprovechamiento de internet para ampliar el negocio. En si los principales beneficios de una implementación BI según la empresa CorpSolution [23] son:

1. Mejora la eficiencia corporativa.
2. Elimina los retrasos en los reportes.
3. Negocia mejores tratos con los clientes y proveedores.
4. Permite encontrar las causas de posibles problemas.
5. Identifica recursos desperdiciados y reduce costos de inventario.
6. Aprovecha la inversión del ERP y los Data Warehouse.
7. Mejora la satisfacción de los clientes.

3.2. *Data Warehouse*

Un *Data Warehouse* o almacén de datos es la colección de datos extraídos desde varios sistemas operacionales, donde los datos son transformados para hacer data consistente y que pueden ser leídos para el análisis [15]. Los almacenes de datos basados en tecnologías de bases de datos, proporcionan los mecanismos para gestionar el cúmulo de información que existe en una empresa[2].

El concepto original de almacenes de datos fue desarrollado por IBM, denominándolo “Almacén de Información” y se presentó como una solución para acceder a datos no relacionales [2]; en si se describen los almacenes de datos como una colección de datos clasificada por temas, integrada, variable en el tiempo y no volátil, que se utiliza como ayuda en la toma de decisiones:

- Clasificada por temas: dado que se organizan por medio de los temas de relevancia de la empresa, en vez de organizarse por áreas de aplicación.
- Integradas: debido a que los datos ingresan al sistema desde muchas fuentes, donde se deben realizar procesos de transformación y limpieza que permita obtener información coherente.
- Variable en el tiempo: los datos solo son precisos y validos en algún instante de tiempo.
- No Volátiles: los datos no se actualizan en tiempo real, por el contrario se refrescan por medio de procedimientos periódicos.

El objetivo de un *data warehouse* dista mucho de las bases de datos de los sistemas transaccionales debido a marcadas diferencias en su método de operación y almacenamiento. Una de las preguntas que muchas empresas se realizan es si realmente ¿requiere un Data Warehouse?, la respuesta es que todas las organizaciones que requieran competir, necesitan de un *data warehouse*, que les permita realizar la toma de decisiones de forma ordenada, para lograr obtener los beneficios de la inteligencia de negocios.

Los modelos de los *data warehouse* son representados por modelos específicos denominados como modelos estrella o modelos copo de nieve como el que se presenta en la Figura 3.2. Estos modelos consisten en la descripción de una tabla de hechos que se encuentra rodeado por una serie de tablas de dimensiones, que son aquellas que permiten describir los componentes de éste.

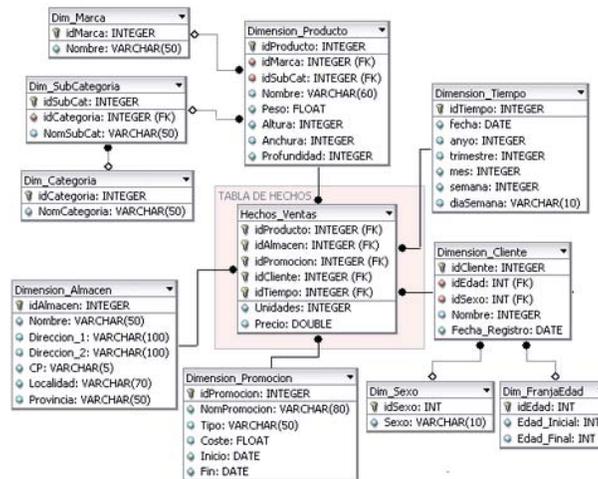


Figura 3.2: Ejemplo de modelo estrella de almacén de datos

La implementación de un *data warehouse* requiere de una infraestructura básica entre software y hardware dentro de la cual se encuentra:

- El servidor sobre el cual corre el software de bases de datos.
- El servidor donde corre el software ETL y los procesos de carga de datos.
- El servidor web que provee el punto de acceso de los usuarios.
- El servidor de BI que procesa las queries, mantiene el panel de control (dashboards) y los reportes.

Es parte de toda infraestructura, contar con múltiples servidores que permita mantener *mirror* de seguridad u otra infraestructura adicional para rendimiento, balanceo de carga, soporte de fallas y la infraestructura de red que permite la comunicación entre los usuarios y la implementación.

Pese a los altos costos de infraestructura, los sistemas de inteligencia de negocios reportan grandes beneficios a la organizaciones que los utilizan, pero el dimensionamiento debe estar basado en las necesidades de la organización de tal forma de no superar la curva de inversión o costos, con el beneficio que pudiese reportar, como lo muestra el gráfico de la Figura 3.3.

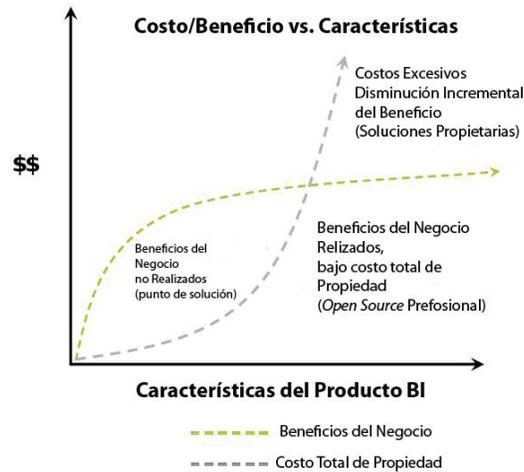


Figura 3.3: Curva costo/beneficio vs Características de una implementación de *Data Warehouse* [15].

3.3. Minería de Datos

Tal como se ha nombrado anteriormente, las instituciones hoy en día cuentan con una gran cantidad de información, que les puede ser útil para la toma de decisiones, pero la manipulación de ésta se hace prácticamente imposible con métodos tradicionales y en muchos casos el conocimiento que pueden otorgar se encuentra oculto entre la gran cantidad de atributos. Frente a esta necesidad es que surge el proceso de Descubrimiento de Conocimiento en Bases de Datos o KDD (Knowledge Discovery in Databases). El proceso se encuentra dividido en tres partes principales que se definen como procesamiento de datos, minería de datos y procesamiento de los resultados, siendo el más importante de ellos el de minería de datos [16]. La minería de datos se define como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. De otra forma la minería de datos es encontrar modelos intangibles a partir de los datos [19]. Por otro lado [2] complementa la definición indicando que la información permite a las organizaciones tomar decisiones de negocios cruciales.

La minería de datos como se ha descrito permite obtener conocimiento desde los datos almacenados, que puede ser aplicado a la toma de decisiones, en área de la educación según Hernandez [19], puede ser aplicada para evaluar la selección o captación de estudiantes, detección de abandonos y fracasos, estimación del tiempo de estancia en la institución.

3.3.1. Tipos de conocimiento

Tal como se ha descrito la minería de datos, permite obtener conocimiento útil para el apoyo a la toma de decisiones. Se identifican dos tipos principales de conocimientos a obtener:

- **Basados en Modelos Predictivos:** También denominadas como técnicas supervisadas, los modelos predictivos utilizan la variable de interés como variable objetivo o dependiente y se usan los otras

variables desde los datos como variables independientes o predictivas. De forma genérica lo que se pretende obtener por estos modelos, es permitir que a partir de una serie de predictores evaluar la pertenencia a una clase. Estos modelos se dividen en dos categorías.

- **Clasificación:** la clasificación se busca encontrar un modelo que pueda predecir el comportamiento de una variable a partir de sus características donde los valores de la clase son discretos. para esto se utilizan algoritmos como árboles de decisión, reglas, redes neuronales.
 - **Categorización:** se estima una correspondencia, para este caso los valores de las clases son continuos y en algunos casos se podrían solapar, también se denomina como regresión, algoritmos comunes son regresión lineal, arboles de regresión.
- **Basados en Modelos Descriptivos:** son modelos que identifican patrones que explican o resumen los datos, sirven para explorar las propiedades de los datos examinados, para esto se utilizan técnicas como:
- **Análisis de Asociaciones:** estas técnicas pretenden sacar patrones de las relaciones que hay entre diferentes rasgos de los datos.
 - **Detección de Anomalías:** el objetivo de la detección de anomalías es encontrar aquellos elementos o características que son significativamente diferentes del resto de los datos.

En ambos casos los principales algoritmos utilizados son visualización, asociaciones y detección de outliers.

Los tipos de conocimientos describen en forma general la búsqueda que se intenta realizar y al mismo tiempo los posibles algoritmos que pueden llegar a ser utilizados para la búsqueda del conocimiento dentro de las bases de datos, como se puede observar en la Figura 3.4, donde se aprecian las tareas principales que se realizan en la minería de datos y las técnicas más comunes utilizadas.

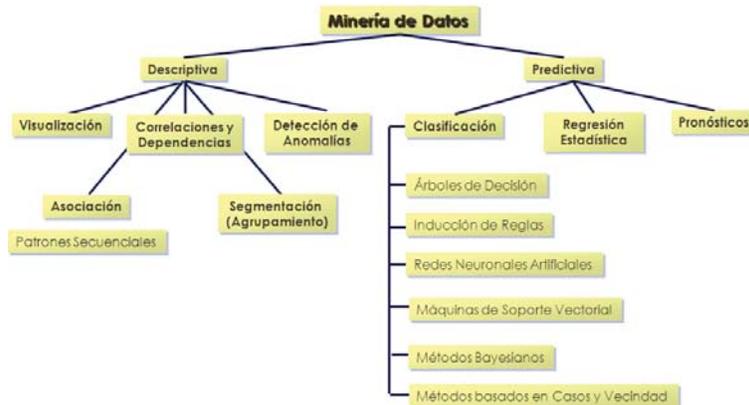


Figura 3.4: Taxonomía de Técnicas de Minería de Datos

3.3.2. Proceso de búsqueda del conocimiento

El proceso de búsqueda de conocimiento corresponde a una serie de etapas que deben ser realizadas y que permiten obtener el conocimiento esperado, tal como lo muestra la Figura 3.5. El proceso que tiene un punto de partida en la selección de los datos y finaliza posterior a una serie de fases entre las que se encuentran la extracción de datos, transformación de datos, minería de datos e interpretación. El proceso que muestra la Figura 3.5, permite describir de forma simple, por medio de un modelo en cascada, las etapas que son utilizadas para la aplicación de la minería de datos.

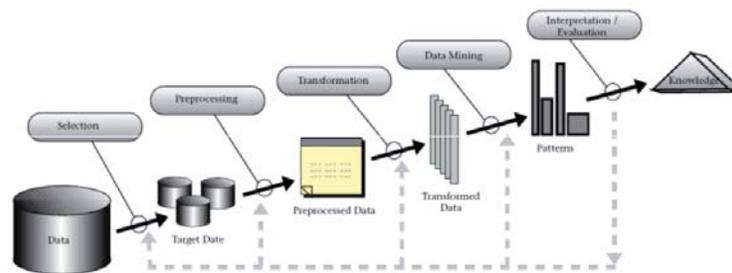


Figura 3.5: Etapas de la extracción de conocimiento

El método anterior es muy útil para explicar de forma visual, el procedimiento desde la extracción al modelado de los datos utilizado en la búsqueda de conocimiento. Para realizar el proceso de extracción de conocimiento, se utilizan principalmente dos metodologías las que corresponden a: la metodología SEMMA (Sample, Explore, Modify, Model, Assess) asociada a productos SAS y la metodología libre, utilizada en este trabajo, CRISP-DM (Cross Industry Standard Process for Data Mining) [24], es una metodología creada en 1996 y que entrega las pautas básicas que permiten realizar un proceso de minería de datos. La Figura 3.6 muestra las etapas del proceso que son descritas como necesarias para la obtención del conocimiento por medio de la minería de datos, donde se visualizan seis etapas, las cuales son:

1. **Comprensión del negocio:** corresponde a las etapas necesaria que permitan describir el negocio y los objetivos que se esperan cumplir con la minería de datos y que permitan formar parte del conocimiento de la empresa, las acciones de esta etapa son:
 - a) Determinación de objetivos de negocio
 - b) Evaluación de la situación
 - c) Determinación de los objetivos de la minería de datos
 - d) Generación del plan del proyecto
2. **Comprensión de datos:** corresponde a la etapa en la cual se deben trabajar más directamente con los datos identificando atributos existentes, cantidad de éstos, sus tipos y si es que estos permitirán otorgar el conocimiento deseado, para esta etapa se requiere trabajar con un experto en los datos del negocio o con los modelos de las bases de datos existentes. Las actividades de esta fase son:

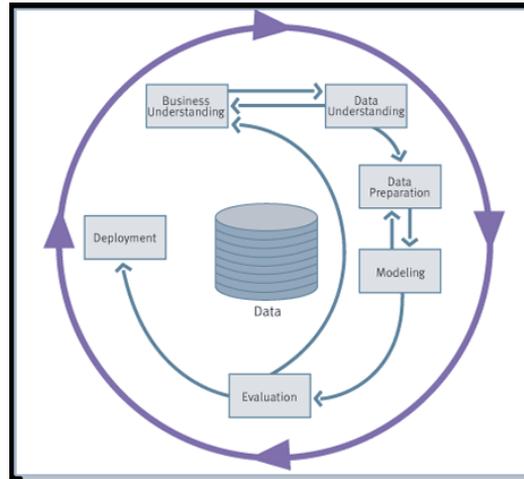


Figura 3.6: Etapas de la extracción de conocimiento

- a) Recolección de datos iniciales.
 - b) Descripción de los datos.
3. **Preparación de datos:** esta fase implica todas las actividades para construir el conjunto de datos final al que puedan aplicarse directamente las herramientas de modelado, las cuales son:
- a) Selección de los datos
 - b) Limpieza de los datos
 - c) Construcción de los datos
 - d) Integración de los datos
 - e) Formateo de los datos
4. **Modelado:** esta fase es la operación de la minería de datos e implica seleccionar las técnicas y parámetros de modelado y evaluar el modelo creado; las tareas de esta fase son:
- a) Selección de la técnica de modelado
 - b) Generación de la prueba de diseño
 - c) Construcción del modelo
 - d) Evaluación del modelo
5. **Evaluación:** valida el modelo desde el punto de vista del análisis de los datos. El modelo y las etapas seguidas durante el modelado se verifican dentro del contexto del cumplimiento de los objetivos. Tareas de esta fase son:
- a) Evaluación de los resultados

b) Determinación de los próximos pasos

6. **Implantación:** el conocimiento obtenido y reflejado en el modelo tiene que organizarse y presentarse de una forma que sea comprensible por parte de los usuarios de la organización. La fase de implantación puede ser tan simple como presentar un informe con el análisis, o tan compleja como implantar procesos repetidos de minería de datos para toda la empresa. Tareas de esta fase son:

- a) Desarrollo del plan de implantación
- b) Desarrollo plan de supervisión y mantenimiento
- c) Construcción del informe definitivo del producto
- d) Revisión del proyecto

3.3.3. Técnicas de Minería de Datos

La minería de datos presenta una serie de técnicas, cada una de las cuales contiene una serie de algoritmos que son utilizados para búsqueda de conocimiento. La extracción de conocimiento, tiene como objetivo la generación de patrones que deben ser validos, novedosos, interesantes y comprensibles [19].

Se ha nombrado anteriormente que existen dos grandes grupos de conocimiento o tareas de minería de datos, como el conocimiento predictivo y el descriptivo; cada uno de ellos envuelve a un grupo de tareas que se realizan dependiendo de la necesidad de extracción de conocimiento o del tipo de información con la que se cuenta. Los tipos de tareas, según lo descrito en [19], son:

- **Predictivas:** se trata de problemas que buscan predecir uno o más valores entre uno o más ejemplos. Los valores de salida van acompañados por una clases que es de tipo discreta, dependiendo de la correspondencia entre ejemplos se definen la siguientes tareas:
 - **Clasificación:** Corresponde a un conjunto de pares de elementos de dos conjuntos, descritos por $\delta = \{\langle e, s \rangle : e \in E, s \in S\}$, donde S es el conjunto de los valores de salida. Los ejemplos e , al estar acompañados por S se denominan ejemplos etiquetados. El objetivo es aprender una función $f : E \rightarrow S$, denominada clasificador, que represente la correspondencia entre los valores. Ejemplos: clasificar medicamentos cual es mejor para cierta patología o clasificar un correo electrónico como spam o no.
 - **Categorización:** Se trata de aprender una correspondencia; es decir, cada ejemplo de $\delta = \{\langle e, s \rangle : e \in E, s \in S\}$, así como la correspondencia a aprender $\lambda : E \rightarrow S$, pueden asignar varias categorías de e , a diferencia de la clasificación, que sólo asigna una. Dicho de otra forma, un ejemplo podría tener varias categorías asociadas. El clasificador consiste en utilizar la clase de mayor probabilidad y seleccionar las K mejores categorías.
 - **Preferencias:** El aprendizaje de preferencias consiste en determinar a partir de dos o más ejemplos, un orden de preferencia. Cada ejemplo es en realidad una secuencia: $\langle e_1, e_2, \dots, e_k \rangle, e_i \in E, k \geq 2$, donde el orden representa la secuencia. Lo más característico de esta tarea es la presentación de los datos.

- **Regresión:** Es quizás la tarea más sencilla de definir. El conjunto de evidencias son correspondencias entre dos conjuntos $\delta : E \rightarrow S$, donde S es el conjunto de valores de salida. Al igual que la clasificación los ejemplos, al ir acompañados de un valor S , se denominan valores ejemplos etiquetados y δ es un conjunto de datos etiquetados. La Diferencia respecto de la clasificación es que S es numérico, es decir puede ser un valor entero o real.
- **Descriptivas:** Los ejemplos se presentan sin orden ni etiqueta, el objetivo central está en describir los ejemplos existentes. Existen muchas maneras de realizar esta tarea, las más comunes son:
 - **Agrupamiento (*clustering*):** El objetivo es obtener grupos o conjuntos entre los elementos de δ , de tal manera que los elementos asignados al mismo grupo sean similares.
Lo importante del agrupamiento respecto a la clasificación son, precisamente, los grupos y es lo que se quiere determinar; a priori, no se conocen ni como son los grupos ni cuantos hay. En algunos casos se proporcionan el número de grupos que se desean obtener, otras veces este número se determina por el algoritmo de agrupamiento, según las características de los datos. Muchos autores consideran el agrupamiento con el objetivo de una nueva tarea, llamada sumarización.
 - **Correlaciones:** Tienen el objetivo de ver la relevancia de atributos, detectar atributos redundantes o dependencias entre atributos, o seleccionar un subconjunto. Los estudios correlacionales y factoriales se centran en los atributos numéricos; el objetivo es ver, dados los ejemplos de conjunto $E = A_1x A_2x \dots x A_n$, si dos o más atributos numéricos A_i y A_j están correlacionados linealmente o relacionados de algún modo, las relaciones son bidireccionales y no orientadas.
 - **Reglas de asociación:** Es una de las tareas principales de la minería de datos, reciben el nombre de análisis de asociaciones o de vínculos, aunque este término también se utiliza en el agrupamiento jerárquico. Dados los ejemplos de conjunto $E = A_1x A_2x \dots x A_n$, una regla de asociación se define de la forma: “ si $A_i = a \wedge A_j = b \wedge \dots \wedge A_k = h$, entonces $A_r = u \wedge A_s = v \wedge \dots A_z = w$ ”, donde todos los atributos son nominales y las igualdades se definen utilizando algún valor de los posibles para cada atributo. La regla anterior está orientada, es decir es una regla de asociación direccional; se llaman dependencias de valor, También existen reglas bidireccionales, donde por cada implicación se tienen reglas de complicación.
 - **Dependencias Funcionales:** Las dependencias funcionales se suelen incluir dentro de la variedad de las tareas conocidas con el nombre de “ reglas de asociación”. Se definen de la manera “ dados los valores de A_i, A_j, \dots, A_k se puede determinar el valor de A_r ”. Las dependencias funcionales, en particular cuando hay un atributo a cada lado, pueden ser orientadas y no orientadas, al igual que las reglas de asociación.
 - **Detección de valores o instancias anómalos:** Son utilizadas para limpieza de los datos, pero al mismo tiempo son muy útiles para detectar comportamientos anómalos, que pueden sugerir fraudes, fallos, intrusos o comportamientos diferenciados. La definición de instancias anómalas es más general debido a que no se considera un atributo, sino que la totalidad de ellos. La tarea se define con el objetivo de encontrar aquellas instancias que no son similares.

A lo largo del informe se han presentado principalmente dos tareas que corresponden a clasificación,

donde se han utilizado técnicas basadas en árboles y técnicas bayesianas y agrupamiento y detección de valores anómalos en la segunda.

3.3.4. Definición de algoritmos Utilizados

Como muestra la Figura 3.4, existen muchos tipos de algoritmos que pueden ser aplicados para realizar las tareas de clasificación o agrupamiento. Durante el proceso de evaluación y búsqueda de patrones, realizado en esta tesis, se han aplicado una serie de algoritmos en respuesta a la tarea que se requiere realizar. Donde principalmente, la tarea desarrollada ha sido una tarea predictiva, aunque muchos de los valores se han validado por medio de tareas descriptivas. Para explicar de mejor forma se ha dividido los algoritmos aplicado según la tarea realizada en la tesis y tan sólo serán explicados aquellos algoritmos utilizados en este trabajo.

3.3.4.1. Tareas Descriptivas

Durante el trabajo se han aplicado dos algoritmos que han permitido describir las correlaciones existentes entre los atributos y aquellos que permiten describir de mejor manera a los alumnos desertores, son:

1. **Agrupamiento (*clustering*) [5]:** Para realizar las tareas de agrupamiento se ha utilizado el algoritmo K-means. Es uno de los más simples y conocidos algoritmos de agrupamiento, sigue una forma fácil y simple para dividir una base de datos dada en K grupos (fijados a priori).

La idea principal se encuentra en definir K centroides (uno para cada grupo) y luego tomar a cada una de las muestras y ubicarlas con respecto a la clase de su centroide mas cercano. El algoritmo se ejecuta por medio de un conjunto de objetos $D_n = (X_1, X_2, \dots, X_n)$, para todo i, X_i reales y k, V_1 , los centros de los K cluster. El algoritmo del K-means se desarrolla en 4 etapas:

- **Etapa 1:** Elegir aleatoriamente K objetos que forman así los K clusters iniciales. Para cada cluster k , el valor inicial del centro es $= x_i$, con los x_i únicos objetos de D_n pertenecientes al cluster.
- **Etapa 2:** Reasignar los objetos del cluster. Para cada objeto x , el prototipo que se le asigna es el que es más próximo al objeto, según una medida de distancia, (habitualmente la medida euclidiana).
- **Etapa 3:** Una vez que todos los objetos son colocados, recalculan los centros de K grupos (los centroides).
- **Etapa 4:** Repetir las etapas 2 y 3 hasta que no se hagan más reasignaciones. Aunque el algoritmo termina siempre, no se garantiza el obtener la solución óptima. En efecto, el algoritmo es muy sensible a la elección aleatoria de los K centros iniciales; ésta es la razón por la que, se utiliza el algoritmo del K-means numerosas veces sobre un mismo conjunto de datos para intentar minimizar este efecto, sabiendo que a centros iniciales lo mas espaciados posibles dan mejores resultados.

3.3.4.2. Tareas Predictivas

Las tareas principales realizadas en el proceso de búsqueda de patrones se han ejecutado utilizando árboles de clasificación, los cuales son uno de los métodos inductivos supervisados no paramétrico más

utilizado. Como forma de representación, los árboles de clasificación destacan por su sencillez y simpleza en la interpretación de los resultados. Los árboles en general, son una forma de representar el conocimiento obtenido en el proceso de aprendizaje inductivo.

Pueden verse a los árboles como la estructura resultante de la partición recursiva y jerárquica del espacio de representación a partir de un conjunto de prototipos. Cada nodo contiene una pregunta, con tantos hijos como respuestas pudiese haber, y cada nodo hoja refiere a una decisión de clasificación. En resumen cada nodo está etiquetado con un par atributo-valor y las hojas con una clase, de forma que la trayectoria que describe desde la raíz hasta las hojas determina la clase de clasificación.

Entre los clasificadores basados en árboles utilizados en esta tesis se encuentran ID3, CART y C4.5; para la gran mayoría de los árboles la metodología de clasificación que se presenta en la Figura 3.7 y se puede resumir como:

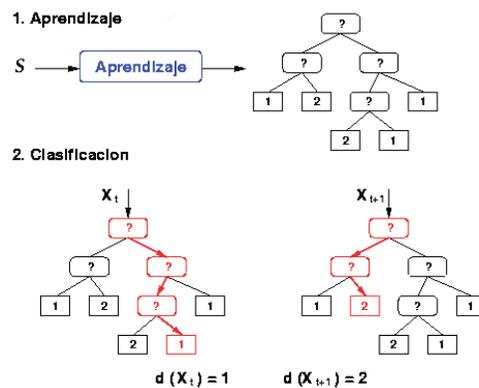


Figura 3.7: Aprendizaje y Clasificación en la generación de los árboles

- **Aprendizaje:** consiste en la construcción de un árbol a partir de un conjunto de prototipos, por medio de un subconjunto de datos, constituye la fase más compleja y la que determina el resultado de la clasificación.
- **Clasificación:** consiste en el etiquetado de un patrón X , independiente del conjunto de aprendizaje. Se trata de responder a las preguntas asociadas a los nodos interiores utilizando los valores de los atributos del patrón X . Este proceso se realiza desde la raíz hasta llegar a las hojas del árbol.

Cada uno de los algoritmos utilizados tiene sus propias características, que permiten cambiar la calidad de los resultados y de las reglas encontradas en el sistema. A continuación se presentan brevemente los algoritmos principales:

1. Algoritmos CART

CART: (Classification and Regression Trees [25]):

a Descripción:

El árbol de decisión CART es un proceso de peticiones recursivas binarias de procesamientos continuos y atributos nominales como objetivo y predictor. Los datos son obtenidos desde su forma nativa, no es necesario o recomendado un enlace. Los árboles son creados hasta un tamaño máximo (cuando la regla lo detiene) y a continuación realiza la poda hacia la raíz. La poda se realiza por niveles al evaluar la menor contribución del rendimiento del árbol sobre los datos de entrenamiento.

El mecanismo de CART permite producir no uno, sino una secuencia de árboles podados anidados, todos árboles candidatos óptimos. El "tamaño adecuado" "esperado" del árbol es identificado por la evaluación del desempeño predictivo de cada árbol, en la secuencia de la poda. Si no existen datos de prueba y no se ha realizado una validación cruzada, CART permanecerá agnóstico respecto a qué árbol de la secuencia es mejor. Esto está en contraste con métodos tales como C4.5 que generan modelos preferidos en la base de datos de entrenamiento medidos.

El mecanismo de CART incluye un balanceo de carga automático (opcional) y una administración también automática de valor faltante; permite una búsqueda de costo-sensibilidad del aprendizaje, construcción característica dinámica, y la estimación de árbol de probabilidad. Los informes finales incluyen un atributo nuevo ranking de importancia.

b Reglas de División

Son siempre expresadas en la forma: una instancia va a la izquierda si cumple la CONDICION, y va a la derecha en cualquier otro caso.

Cuando la CONDICION es expresada como atributo $X_i \leq C$ para atributos continuo. Para atributos nominales la CONDICION es expresada como miembro en una lista de valores. Los autores de CART argumentan que la división binaria es realizada preferentemente porque, en primer lugar hay fragmentos de datos más pequeños que múltiples divisiones y dos divisiones repetidas sobre el mismo atributo son realizadas y si son seleccionadas, podría eventualmente podría generar muchas particiones para un atributo como requerido.

El algoritmo de CART se centra en el proceso de división sobre la medida del índice de la regla Gini, la cual es mejor conocida como entropía o criterio de ganancia de información. Para un criterio binario (0/1) el valor Gini de un nodo t es presentado en la ecuación 3.1.

$$G(t) = 1 - p(t)^2 - (1 - p(t))^2 \quad (3.1)$$

donde $p(t)$ es la frecuencia relativa de la clase 1 en el nodo, y la ganancia generada por una división del nodo padre P en hijos izquierdo y derecho (L y R) es presentada en la ecuación 3.2.

$$I(P) = G(P) - qG(L) - (1 - q)G(R) \quad (3.2)$$

Además, q es la fracción de casos que van a la izquierda. Los autores de CART están a favor de del criterio de Gini sobre ganancia de información porque Gini puede ser fácilmente ampliado para incluir los costos sintetizados y se calcula más rápidamente que la ganancia de información. Esto introduce la regla dos elementos modificada, que se basa en una comparación directa de la distribución de atributo de destino en dos nodos secundarios 3.3.

$$I(split) = \left[.25 (q(1-q))^u \sum_k |p_L(k) - p_r(k)| \right]^2 \tag{3.3}$$

En un ejemplo presentado en [8], se aprecia el proceso de segmentación de una muestra, donde se identifica las diferentes evaluaciones; para este ejemplo se muestra una partición resultante de evaluar $X_1 - X_2 < -1$.

La primera partición hace que uno de los nodos resultantes sea puro. En este caso, el izquierdo, que corresponde a la clase 1 y se establece como nodo terminal, según los cálculos de ganancia ya obtenidos, Figura 3.8.

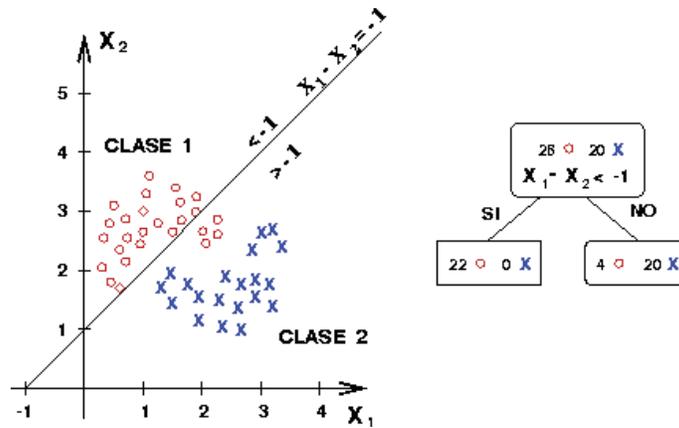


Figura 3.8: Primera Partición

Una segunda partición está basada en la pregunta: $X_1 + X_2 < 4$. La partición resultante se muestra en la Figura 3.9. La que genera un nodo puro para la clase 2.

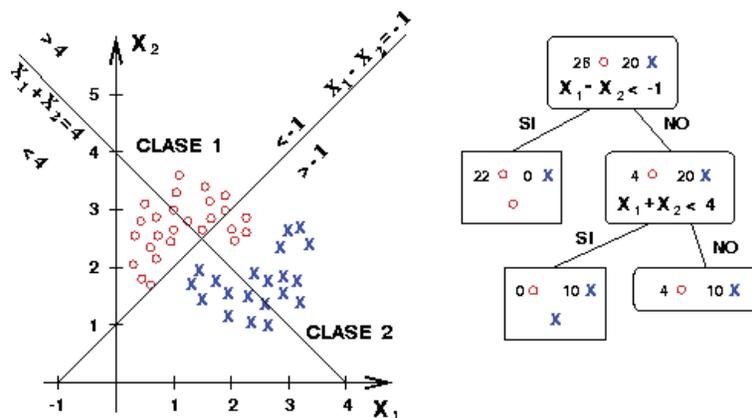


Figura 3.9: Segunda Partición

Finalmente, una tercera partición hace que todos los nodos sean puros, la asociada a la pregunta:

$X_1 - X_2 < 0$, como se muestra en la Figura 3.10.

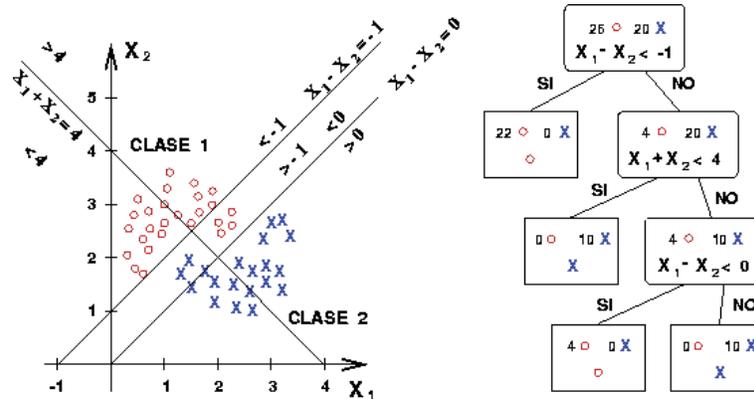


Figura 3.10: Tercera Partición

Ahora, las particiones se hacen con hiperplanos arbitrarios, no necesariamente paralelos a los ejes. La conclusión más importante que podemos sacar de los dos últimos ejemplos es que un árbol de clasificación T representa una partición recursiva del espacio de representación, P , realizada en base a un conjunto de prototipos, S . La Figura 3.11 muestra como cada una de los segmentos del hiperplano son representados en los nodos del árbol.

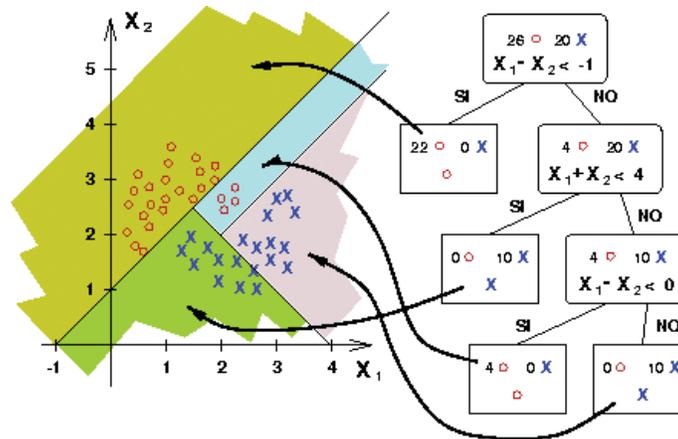


Figura 3.11: Representación de los nodos en el árbol

2. Algoritmo C4.5

El C4.5 es una extensión del ID3, que termina con muchas de sus limitaciones. Por ejemplo, permite trabajar con valores continuos para los atributos, separando los posibles resultados en dos ramas: una para aquellos $A_i \leq N$ y otra para $A_i > N$. Además, los árboles son menos frondosos porque cada hoja no cubre una clase en particular sino una distribución de clases, lo cual los hace menos profundos y menos frondosos [14].

El C4.5 genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente, según la estrategia de búsqueda de primero en profundidad (depth-first). Antes de cada partición de datos, el algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona aquella que resulta en la mayor ganancia de información o en la mayor proporción de ganancia de información.

Para cada atributo discreto, se considera una prueba con N resultados, siendo N el número de valores posibles que puede tomar el atributo. Para cada atributo continuo, se realiza una prueba binaria sobre cada uno de los valores que toma el atributo en los datos.

La Figura 3.12 muestra el pseudocódigo del algoritmo C4.5. Éste recibe como entrada un conjunto de atributos no clasificadores de entrada R , un atributo clasificador C y un conjunto de datos de entrenamiento y como salida, generando un modelo clasificador que puede ser representado como árbol.

PSEUDOCODIGO DE C4.5

```

Función C4.5
R: conjunto de atributos no clasificadores,
C: atributo clasificador,
S: conjunto de entrenamiento, devuelve un árbol de decisión
Comienzo
  Si S está vacío,
    Devolver un único nodo con Valor Falla; 'para formar el nodo raíz
  Si todos los registros de S tienen el mismo valor para el atributo
  clasificador,
    Devolver un único nodo con dicho valor; 'un unico nodo para todos
  Si R está vacío,
    Devolver un único nodo con el valor más frecuente del atributo
    Clasificador en los registros de S [Nota: habrá errores, es decir,
    Registros que no estarán bien clasificados en este caso];
  Si R no está vacío,
    D ← atributo con mayor Proporción de Ganancia (D,S) entre los
  atributos de R;
    Sean {dj | j=1,2,..., m} los valores del atributo D;
    Sean {Sj | j=1,2,..., m} los subconjuntos de S correspondientes a los
  valores de dj respectivamente;
    Devolver un árbol con la raíz nombrada como D y con los arcos
  nombrados d1, d2,...,dm, que van respectivamente a los árboles
    C4.5(R-{D}, C, S1), C4.5(R-{D}, C, S2), C4.5(R-{D}, C, Sm);
Fin

```

Figura 3.12: PseudoCódigo del Algoritmo C4.5 [14]

En cada uno de los nodos el algoritmo C4.5 debe escoger cual prueba realizar para particionar los datos. Existen tres casos:

- a. Prueba estándar para atributos discretos, obteniendo un resultado y una rama para cada valor posible del atributo.
- b. Prueba basada en un atributo discreto, donde los valores son asignados a un número variable de grupos, con un resultado posible para cada grupo, contrario a un resultado para cada valor.

Si el atributo A tiene atributos numéricos continuos, se realiza una prueba binaria con resultados $A < L$ y $A \geq L$, para lo que se debe determinar el valor del límite de L .

Cada una de las pruebas se realiza de la misma forma y se evalúa o compara la ganancia de información obtenida.

3.3.5. Herramientas para la Minería de Datos

Al hablar de las herramientas para la minería de datos es necesario indicar que existen en este espacio dos grandes grupos, que corresponden a las herramientas con licencias comerciales y las herramientas libres.

Cada una de las herramientas presenta ventajas y desventajas y son utilizadas en mayor y menor porcentaje por los usuarios, según se muestra en un estudio realizado por kdnuggets [17] por medio de una encuesta realizada a 1103 votantes se identifica la distribución del uso de las herramientas de minería de datos según licencia, en su mayoría pertenecen a herramientas bajo licencia, seguido por las herramientas de software libre

Dentro de las herramientas más utilizadas se encuentran RapidMiner, R, Excel, KNIME, WEKA y otros [17], como se muestra en la Figura 3.13.

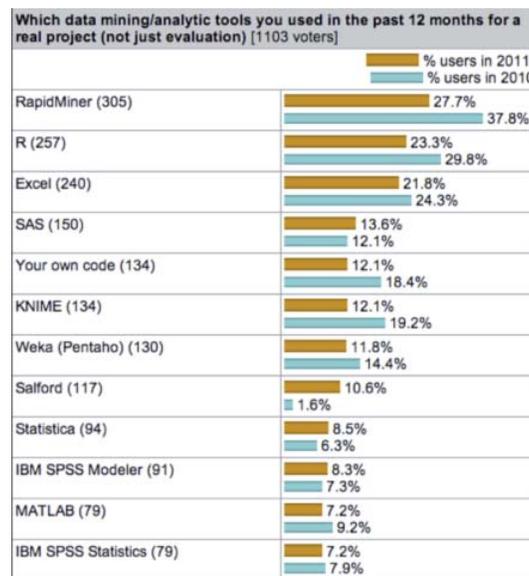


Figura 3.13: Listado de Herramientas por cantidad de usuarios

Con un uso menor se encuentran herramientas como Orange Canvas, que pese a sus cualidades, es menos conocida, como se muestra en la Figura 3.14

Tal como se aprecia en las Figuras 3.13 y 3.14, existe una gran cantidad de herramientas que permiten realizar las tareas de minería de datos; para el trabajo de esta tesis y debido a restricciones de costos se han utilizado sólo herramientas de código abierto o software libre, indicadas a continuación:

1. Weka (Waikato Environment for Knowledge Analysis)

Es una plataforma de software para aprendizaje automático y minería de datos escrito en Java y desarrollado en la Universidad de Waikato. Es un software libre distribuido bajo licencia GNU-GPL.

| | |
|-----------------------------|------|
| | 0.8% |
| Hadoop Map/Reduce (19) | 1.7% |
| Mathematica (18) | 1.6% |
| Revolution Computing (15) | 1.4% |
| | 0.4% |
| KXEN (15) | 1.4% |
| | 2.1% |
| Orange (14) | 1.3% |
| | 2.7% |
| Miner3D (14) | 1.3% |
| | 0.8% |
| XLSTAT (10) | 0.9% |
| NoSQL databases (10) | 0.9% |
| Stata (9) | 0.8% |
| Other cloud-based tools (9) | 0.8% |

Figura 3.14: Listado de Herramientas Por cantidad usuarios menor

El paquete contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo. [10]

Weka soporta varias tareas estándar de minería de datos, especialmente preprocesamiento de datos, *clustering*, clasificación, regresión, visualización, y selección. Todas las técnicas de Weka se fundamentan en la presunción de que los datos están disponibles en un archivo plano, en la que cada registro de datos está descrito por un número fijo de atributos. No puede realizar minería de datos multi relacional, pero existen aplicaciones que pueden convertir una colección de tablas relacionadas de una base de datos en una única tabla que ya puede ser procesada con Weka. Su interfaz de entrada se presenta en la Figura 3.15



Figura 3.15: Interfaz de la aplicación WEKA

2. Orange Canvas

Es una herramienta de tipo *open source* basada en componentes de minería de datos y máquina de aprendizaje. Orange cuenta con gran facilidad de uso y una potente programación, con una presentación visual flexible para el análisis exploratorio de datos y visualización. La programación de la interfaz está en Python, proporcionada con acceso directo a toda su potencia, para la programación rápida de nuevos algoritmos y desarrollo de complejos procedimientos de análisis de datos.

Orange está siendo desarrollado en el Laboratorio de Bioinformática de la Facultad de Ciencias de

la Computación e Informática de la Universidad de Ljubljana, Eslovenia, junto con la comunidad de código abierto y se utiliza activamente en la investigación [10].



Figura 3.16: Interfaz de la aplicación Orange Canvas

3.4. Proceso de Análisis Jerárquico

Diseñada en 1970 por el Dr. Thomas Saaty, es un método de análisis multiobjetivo [3], para problemas discretos, el cual divide o jerarquiza el problema de decisión en criterios y subcriterios. AHP es uno de los métodos de estimación más sencillos, donde la dificultad principal se centra en la identificación de los atributos y su contribución relativa, problema que en el caso de este análisis se encuentra cubierto debido los resultados obtenidos por el proceso de minería de datos. El modelo de AHP permite, de maneras eficiente y gráfica descomponer y analizar por partes la información, además de visualizar los efectos de los cambios en los niveles. AHP se fundamenta en:

- Estructuración de un modelo jerárquico.
- Priorización de los elementos del modelo jerárquico.
- Comparaciones binarias entre los elementos.
- Evaluación de los elementos mediante la asignación de pesos.
- Ranking de las alternativas de acuerdo a los pesos dados
- Síntesis

- Análisis de sensibilidad.

En general AHP es una herramienta metodológica que ha sido utilizada en varios países para realizar toma de decisiones en distintos conflictos desde la calidad de un servicio, preferencias de ventas y otros servicios, algunas ventajas del método son:

- Presenta un sustento matemático.
- Permite desglosar y analizar un problema por partes.
- Permite medir criterios cuantitativos y cualitativos mediante una escala común
- Incluye la participación de diferentes personas o grupos.
- Permite verificar un índice de consistencia y hacer correcciones.
- Permite generar una síntesis y realizar un análisis de sensibilidad
- Es de fácil uso.

Para realizar el proceso se utiliza una escala de preferencias que se encuentra definida por:

- Los elementos tienen la misma importancia valor 1
- El elemento a es ligeramente más importante que el elemento b valor 3
- El elemento a es más importante que el elemento b valor 5
- El elemento a es fuertemente más importante que el elemento b valor 7
- El elemento a es muy fuertemente importante que el elemento b valor 9

Capítulo 4

Preprocesamiento de los Datos

4.1. Reconocimiento del Dominio y los Usuarios

El dominio del análisis se encuentra en la evaluación de la deserción de la educación superior, específicamente en una universidad tradicional de la zona de Valparaíso, que permita a los directores y coordinadores de carrera tomar decisiones frente al inminente riesgo de que un alumno sea candidato a deserción de la carrera, o peor aún que este alumno sea candidato a ser eliminado académicamente de la universidad.

Existen principalmente dos niveles de usuarios al proceso de evaluación; el primer nivel se encuentra en la División Académica que es la unidad principal que evalúa la educación a niveles generales e implanta las directrices base a todas las carreras; para ella tiene una importancia fundamental conocer las causales que llevan a un alumno desertar de la universidad.

Por otra parte están los directores y coordinadores de carrera, para quienes su centro de atención se encuentra en las causas que provocan la deserción de los alumnos de la carrera, y en posibles medidas que podrían ser implementadas para evitar este fenómeno.

Las distintas autoridades encuentran una serie de dudas, frente a las cuales se han aplicado algunos modelos utilizados en otras universidades que intentan evitar que un alumno sea eliminado académicamente en primer año, como es la aplicación de ciclos de reforzamiento previo al primer semestre obligatorio.

Pero las preguntas son ¿es necesario este ciclo de reforzamiento?, ¿existen factores que permitan determinar si un alumno pueda ser eliminado académicamente?, ¿existe algún tipo de alumno que rinde la PSU que pueda ser eliminado académicamente?, ¿existen factores académicos que permitan evaluar las solicitudes de modificación de asignaturas?. Éstas son algunas de las dudas que se plantean comúnmente en reuniones de análisis de la problemática de la deserción.

Para enfrentar estas preguntas hoy se utilizan modelos basados en las experiencias de los académicos, los cuales más que reglas tienen criterios personales. Esto se realiza principalmente en situaciones donde los alumnos solicitan cursar asignaturas sin prerrequisitos, pero no hay información estadística sobre si es factible que el alumno apruebe o repruebe una cierta asignatura por no contar con ellos; por otro lado en la evaluación de las solicitudes de continuidad de estudios, se aplican principalmente criterios de avance del alumno, donde el proceso no incluye decisiones fundamentado en reglas que permitan validar la probabilidad que el alumno continúe o no continúe en la carrera.

La evaluación que realizan los directores de carrera, para los procesos de solicitud de modificación de inscripción de asignaturas o para permitir la continuidad de estudios, se realiza por medio de indicadores básicos de cada alumno, que refieren a cantidad de asignaturas cursadas por primera, segunda, tercera o más veces, asignaturas aprobadas efectivamente versus cuantas debería tener aprobadas a la fecha, o en un estado de avance por semestre que cursa de la carrera; pero estos indicadores no han sido validados para determinar si es factible que los alumnos puedan o no ser desertores, o bien su condición puede permitirles finalizar su carrera.

Por otra parte a nivel central existen muchas dudas con respecto a los alumnos que ingresan a la universidad, las preguntas implican si ¿los resultados obtenidos en la PSU pueden afectar a un cierto perfil de alumno que lo lleve a desertar a la universidad?, o ¿existe un tipo de alumno o procedencia del alumno que identifique debilidades que lo lleven a desertar a la educación superior?.

Entre los factores más relevantes que se nombran en distintas reuniones se encuentran el puntaje ponderado de ingreso, promedio de enseñanza media, tipo de educación, si rindió la PSU inmediatamente posterior a la educación media o egresó en años anteriores, de qué región proviene, cuál es el lugar en que fue seleccionado, tipo de selección (seleccionado o lista de espera) o cuál es la preferencia la carrera, entre otros.

En muchos documentos se valida como causa predominante de la deserción de la educación superior a los problemas económicos, ¿es realmente valido esto?, ¿se trata de alumnos que no tienen ningún tipo de beneficio? o ¿son alumnos que al menos tienen un tipo de ayuda económica para su educación?. Se dice por otro lado que los alumnos becados pueden ser más destacados que los no becados ¿es un factor que predomine en los eliminados?.

En resumen es importante determinar para la División Académica y los directores de carrera si existen atributos que dependan de su avance académico, de la información financiera o de la información de la postulación a la carrera que permitan describir a los alumnos desertores.

Antes de realizar la extracción de los datos es necesario explicar que estos se encuentran distribuidos en distintas bases de datos, las cuales son:

1. **Aranceles:** contiene los datos de las cuentas corrientes de los alumnos, históricos de pagos y beneficios asignados al alumno.
2. **Pregrado:** registra la historia del alumno, con sus notas y registro académico, además de las estructuras curriculares, cátedras y profesores que han dictado éstas en cada semestre.
3. **Postula:** contiene lo antecedentes de los alumnos que han postulado a la universidad, ésta se encuentra dividida por los datos de los alumnos que han rendido la PAA y PSU
4. **Titulados:** guarda todos los alumnos que han realizado el trámite de obtención de su título o grado.

En base a estas cuatro bases de datos se han generado archivos con los datos de los alumnos desertores, extrayendo la información, sin importar carrera o condición. Esto, con la intención de identificar cual es la situación inicial en la cual se encuentran los alumnos eliminados, permitiendo así identificar algunas condiciones genéricas de los alumnos desertores.

Este proceso se realizó principalmente utilizando secuencias de comandos SQL, donde la restricción principal se encuentra en los alumnos que no tienen periodo actual y que no han obtenido su título o grado, sin importar su estado; con este proceso se han generado los siguientes archivos:

1. **Alumnos desertores con periodo actual:** contiene a los alumnos que han desertado de alguna carrera incluyendo su datos básicos como periodos de permanencia, estados y vía de ingreso. Los atributos del archivo se encuentran descritos en la Tabla 4.1

- Cantidad de registros: 13.516
- Cantidad de atributos: 6
- Atributos continuos: 3
- Atributos discretos: 3

| Atributo | Descripción |
|-------------|--|
| nombre | nombre de la carrera y decreto asociado |
| ingreso | año en el cual ingreso el alumno a la carrera |
| periodo | año en el cual el alumno desertó de la carrera |
| annios | cantidad de años que el alumno perteneció a la carrera |
| semestre | semestre en el cual el alumno desertó |
| via_ingreso | vía por la cual ingresó a la carrera o programa |
| estado | estado académico del último periodo cursado |

Tabla 4.1: Atributos de alumnos desertores

2. **Alumnos desertores con sus notas:** este archivo contiene la información de los alumnos desertores con sus notas máxima y mínima y la cantidad de asignaturas rendidas por primera, segunda y más veces. Los atributos se encuentran descritos en la Tabla 4.2.

- Cantidad de registros: 13.101
- Cantidad de atributos: 14
- Atributos continuos: 8
- Atributos discretos: 6

| Atributo | Descripción |
|----------------|---|
| promedio | promedio del alumno incluyendo todas sus notas |
| maximo | máxima nota obtenida en sus estudios |
| minimo | mínima nota obtenida en sus estudios |
| nombre_carrera | nombre de la carrera del alumno |
| primera | cantidad de asignaturas cursadas por primera |
| segunda | cantidad de asignaturas cursadas por segunda |
| tercera | cantidad de asignaturas cursadas por Tercera |
| cuarta | cantidad de asignaturas cursadas por cuarta |
| mas_de_cuatro | cantidad de asignaturas cursadas más de cuatro |
| rpi | cantidad de asignaturas reprobadas por inasistencia |

| | |
|-------------|--|
| rpL | cantidad de asignaturas reprobadas por laboratorio |
| sin_nota | cantidad de asignaturas sin nota |
| postergadas | cantidad de asignaturas con nota postergada por parte del alumno |
| via_ingreso | vía de ingreso del alumno |
| estado | estado en el cual se encuentra al momento de desertar a la carrera |

Tabla 4.2: Atributos del archivo de alumnos desertores y sus notas

3. **Alumnos desertores y sus beneficios:** uno de los puntos que fundamentan muchos autores es que los alumnos principalmente desertan debido a problemas económicos o porque no tienen beneficios; este archivo los representa identificando los pagos realizados, créditos otorgados y becas o rebajas asignadas. Los atributos del archivo utilizado para este proceso se muestran en la Tabla 4.3.

- Cantidad de registros: 13.382
- Cantidad de atributos: 14
- Atributos continuos: 9
- Atributos discretos: 5

| Atributo | Descripción |
|------------------|---|
| beca | tipo de beca obtenida por el alumno |
| porcentaje_beca | porcentaje de beca asignada |
| estado_beca | estado en el cual se encuentra la beca B03 es una beca asignada definitivamente |
| credito | tipo de crédito obtenido por el alumno |
| porcentaje_total | porcentaje de crédito total obtenido por el alumno |
| rebaja | tipo de rebaja asignada al alumno |
| rebaja_total | valor total en dinero asignado como rebaja al alumno |
| atrasos | cantidad de pagos con retraso que ha realizado el alumno |
| cod_carrera | código de la carrera |
| nombre | Nombre de la carrera que estudia el alumno |
| ingreso | año de ingreso del alumno |
| periodo | año en el cual deserta el alumno |
| semestre | semestre en el cual realiza la deserción |
| via_ingreso | vía por la cual ingresa el alumno |
| estado | ultimo estado académico del alumno |

Tabla 4.3: Atributos del archivo de alumnos desertores y sus beneficios

4. **Alumnos desertores y su proceso de postulación:** este archivo contiene al subconjunto de alumnos que ingresaron a la universidad por medio del proceso tradicional de la PSU. El detalle de los atributos del archivo se presenta en la Tabla 4.4.

- Cantidad de registros: 6.581
- Cantidad de atributos: 25
- Atributos continuos: 16
- Atributos discretos: 9

| Atributo | Descripción |
|--------------|---|
| carrera | carrera a la cual ingreso el alumnos vía PSU |
| preferencia | preferencia de postulación a la carrera |
| codsel | tipo de selección seleccionado o lista espera |
| lugar | lugar en el cual quedo seleccionado |
| puntaje | puntaje ponderado de ingreso |
| situ_egreso | situación en la cual finalizó sus estudios de enseñanza media |
| local_educ | local educacional en el cual termino sus estudios de enseñanza media |
| unidad_educ | unidad educacional particular, subvencionado o municipalizado |
| rama_educ | rama de la educación media, científico humanista, técnico profesional |
| grupo_depend | particular, subvencionado o municipalizado |
| region | región en la cual cursa sus estudios secundarios |
| provincia | provincia de procedencia del alumno |
| anoegreso | año de egreso de la enseñanza media del alumno |
| promnota | promedio de enseñanza media |
| ptjeNem | puntaje nem asignado al promedio de notas |
| lenguaje | puntaje en la prueba de lenguaje |
| matematica | puntaje en la prueba de matemática |
| historia_sco | Puntaje obtenido en la prueba específica de Historia y ciencias Sociales |
| ciencias | puntaje obtenido en la prueba de ciencias |
| Pronmlm | promedio PSU matemáticas y leguajes |
| ingreso_brut | ingreso bruto familiar |
| ingreso | año de ingreso a la carrera |
| periodo | año en el cual deserto de la carrera |
| semestre | Semestre en al cual deserta el alumno |
| via_ingreso | vía de ingreso, esto se debe a muchos alumnos sólo rinden la psu para optar a los beneficios estatales. |
| estado | estado académico en el cual finaliza el alumno |

Tabla 4.4: Descripción atributos proceso de postulacion alumnos desertores

Tal como se ha nombrado existe una serie de condicionantes generales utilizadas para la extracción de

los archivos iniciales, las restricciones utilizadas para la extracción de los datos son:

1. Carreras que tengan periodo actual de tal forma de descartar a todas las carreras que ya no continúan o sus programas han sido modificados.
2. Alumnos que no han obtenido o no han realizado el trámite de obtención de su título o grado, pese a que puedan haber completado su programa de estudios.
3. Último periodo cursado por el alumno, del cual se obtienen su estado y su situación final.
4. Finalmente las bases de datos se han posicionado estáticamente en el primer periodo 2011.

En base a estas restricciones se ha realizado una evaluación inicial de forma general a todos los desertores, con la cual se pretende identificar factores genéricos que puedan afectar a todos los alumnos; una vez identificados los atributos más representativos de los alumnos se realizará, utilizando técnicas de minería de datos, una evaluación exhaustiva de las carreras que sean representativas dentro de los programas que se encuentran en curso, donde se han determinado algunos criterios de selección como son cantidad de alumnos eliminados, cantidad de periodos que tiene la carrera o programa y si este programa continúa a futuro, entre otros parámetros.

Para realizar el análisis se han utilizado las mejores potencialidades de las herramientas más destacadas de tipo *open source* y simples de usar en minería de datos como son Weka, Orange y Knime.

4.2. Análisis Exploratorio General de los Datos de Deserción

Al evaluar los datos recopilados, se identifica que en todas las carreras la deserción ocurre en los primeros años, donde los desertores dejan la carrera en su gran mayoría antes del tercer año de la carrera como muestran las Figuras 4.1 - 4.2

Tal como se puede visualizar el 45 % de los alumnos eliminados de la universidad se encuentran entre primer y segundo año, y más del 55 % de los alumnos deserta hasta el tercer año, esto ya revela que es un punto importante aumentar la preocupación por el alumnado en los primeros años de trayectoria, dado que los factores que puedan determinar su deserción son más relevantes al inicio de sus carreras.

Al observar la Figura 4.1, se identifica que el estado de alumno regular (A14), mayoritariamente representa a los alumnos de primer año, que no renuevan su matrícula en dicha carrera, junto a ellos se encuentran aquellos que abandonaron su carrera (A02) y aquellos que desertan en segundo año están representados por los alumnos eliminados académicamente (A06).

Por otra parte los alumnos que se encuentren en un estado de titulado (A12), sin haberlo obtenido o sin encontrarse en los registros de títulos son considerados como desertores por parte de la División Académica. Para los estudios iniciales serán considerados dado que afecta a los indicadores de acreditación para las tasas de titulación efectiva, pero para evaluar un indicador de riesgo no son útiles dado que la probabilidad de deserción académica en este caso es prácticamente nula.

Junto a lo anterior se identifica el efecto que tienen las crisis institucionales podría repercutir directamente en la deserción de los estudiantes; en la Figura 4.3 se observa que en el año 2005, la tasa de deserción subió ostensiblemente, coincidente con el hecho de haber publicado el problema en la administración de la universidad.

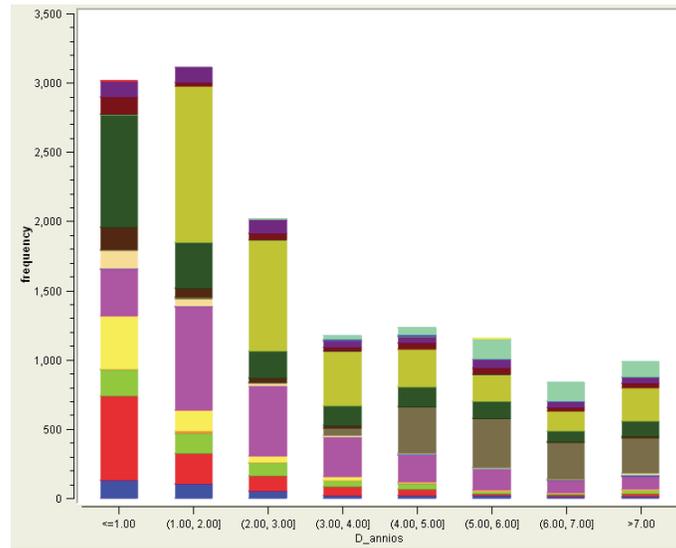


Figura 4.1: Cantidad de Años Cursados por los alumnos de la carrera de construcción



Figura 4.2: Cantidad de Alumnos por Años Cursados por los alumnos

Al visualizar la información académica de los alumnos, se puede apreciar que los desertores son alumnos que su promedio de notas se encuentra entre un 3.0 y un 4.0, Figura 4.4.

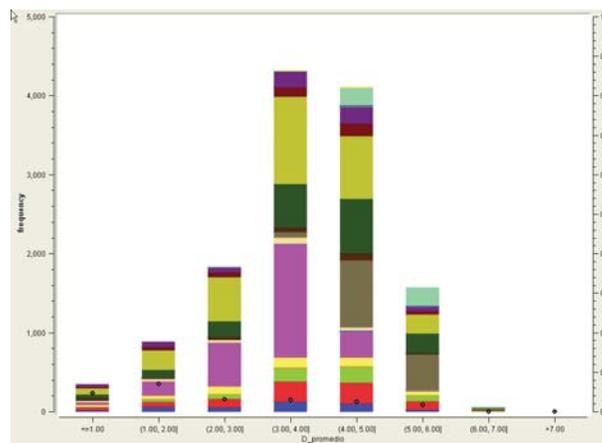


Figura 4.4: Cantidad de alumnos por promedio de notas

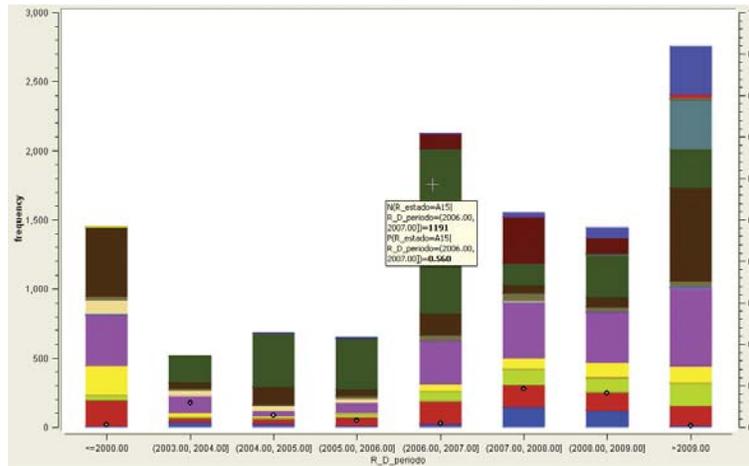


Figura 4.3: Número de alumnos por periodo de deserción

Junto a lo anterior se visualiza que más del 70 % de los alumnos desertores han obtenido una nota mínima inferior a 2.0, interesante sería observar a posterior si estas notas se reflejan en los primeros años de estudio.

Lo mismo se puede ver en los promedios de los alumnos, donde más del 60 % de los promedios se encuentra con notas entre un 3.0 y un 5.0, siendo en su mayoría notas entre un 3.0 y un 4.0, tal como se muestra en la Figura 4.5.

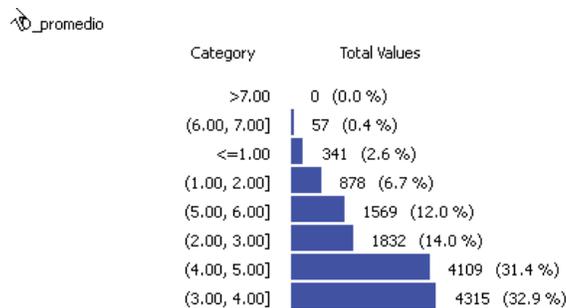


Figura 4.5: Cantidad de alumnos por promedio de nota

Tras una primera visualización, un punto importante que se debe evaluar es que existe un número no menor de alumnos que han desertado manteniendo alguna asignatura sin nota ver Figura 4.6, lo cual sería muy interesante evaluar para verificar cuanto afecta este problema en los registros.

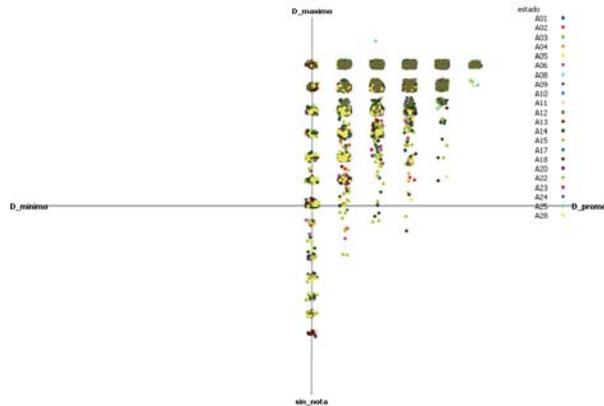


Figura 4.6: Grafico de distribucion de datos y atributos principales

Una de las mayores causas que distintos autores indican de la deserción se encuentra en los problemas económicos de los alumnos, dado que éstos no cuentan con un verdadero apoyo financiero que les permita su continuidad de estudios.

Al realizar un análisis de los componentes se puede visualizar que los alumnos desertores, en general, son alumnos que no tienen asignada algún tipo de beca; como lo muestra la Figura 4.7 donde más del 78 % de los alumnos no tiene asignada una beca y tan sólo un 18,9 % de ellos tiene algún tipo de beca.

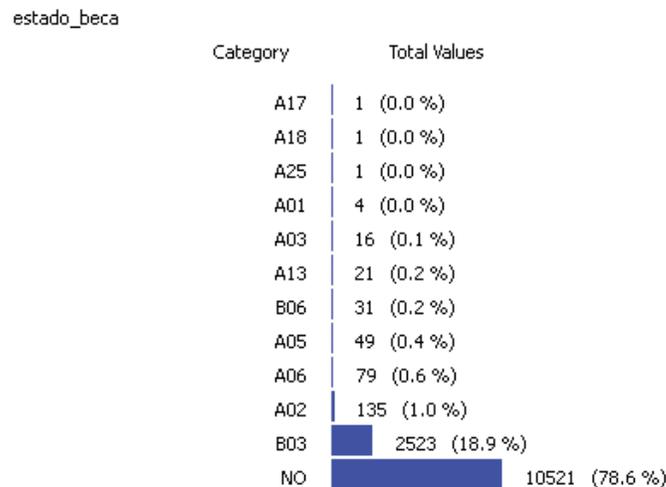


Figura 4.7: Estado de asignacion de becas alumnos desertores

Por otra parte la asignación de los créditos no dista mucho de la asignación de las becas donde apenas un 68 % de los alumnos cuenta con la asignación de un crédito como se puede ver en la Figura 4.8 donde la

categoría 0 corresponde a los alumnos que no tienen crédito y los restantes a diversos tipos.

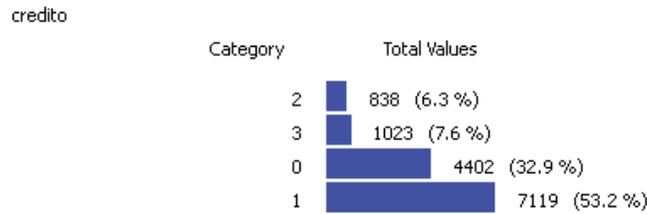


Figura 4.8: Estado de asignacion de becas alumnos desertores

Además, al complementar este análisis con el porcentaje de asignación, los alumnos en su mayoría tienen un muy bajo porcentaje de asignación, sin considerar a los datos $j=1$ que corresponden a aquellos que no tienen crédito alguno o se les asignó sin valor. Por otro lado, el grupo mayoritario de alumnos desertores tienen créditos asignados con porcentajes mayores al 50 % como lo muestra la Figura 4.9, por lo cual el 68 % de los alumnos tiene algún tipo de ayuda económica que les permite solventar el pago de su carrera.

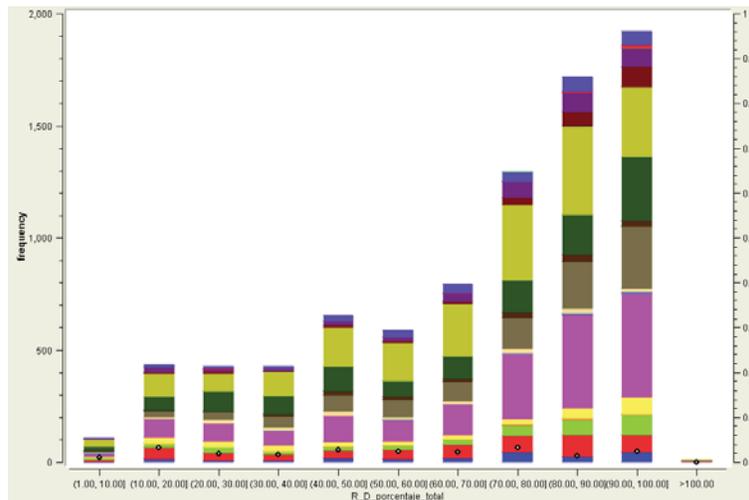


Figura 4.9: Porcentaje de asignación total de credito asignado

Un detalle importante se encuentra que en económicamente hablando, mayoritariamente los alumnos desertores se encuentra descritos por el segmento de A15 o sea alumnos que no volvieron a la universidad simplemente al semestre siguiente, donde el 47 % de estos alumnos no tiene crédito, no tiene beca y al menos se ha retrasado en un pago. Figura 4.10.

| Cluster centroides: | | | | | | |
|---------------------|----------------------|-------------------------|-----------------|----------------|----------------|----------------|
| Attribute | Full Data (12082) | Cluster# 0 (6234) | 1 (3031) | 2 (2130) | 3 (821) | 4 (1166) |
| D_beca | <=9.00 | <=9.00 | <=9.00 | <=9.00 | <=9.00 | <=9.00 |
| D_porce_beca | <=10.00 | <=10.00 | <=10.00 | <=10.00 | <=10.00 | <=10.00 |
| estado_beca | NO | NO | NO | NO | NO | NO |
| credito | 1 | 0 | 1 | 1 | 1 | 1 |
| D_porcentaje_total | <=1.00 | <=1.00 (80.00, 90.00) | (90.00, 100.00) | (40.00, 50.00) | (80.00, 90.00) | |
| D_rebaja | <=1.00 | <=1.00 | <=1.00 | <=1.00 | <=1.00 | (22.00, 25.00] |
| D_rebaja_total | <=189450.00 | <=189450.00 | <=189450.00 | <=189450.00 | <=189450.00 | >=200000.00 |
| D_atrasos | (0.00, 1.00] | (0.00, 1.00] | (1.00, 2.00] | (2.00, 4.00] | >4.00 | >4.00 |
| via_ingreso | Y01 | Y01 | Y01 | Y00 | Y00 | Y01 |
| estado | A15 | A15 | A06 | A15 | A15 | A12 |

| Clustered Instances | |
|---------------------|-------------|
| 0 | 6234 (47%) |
| 1 | 3031 (23%) |
| 2 | 2130 (16%) |
| 3 | 821 (6%) |
| 4 | 1166 (9%) |

Figura 4.10: Cluster beneficios utilizando Simple K-Means

Además, simple Kmeans, también muestra que los alumnos eliminados académicamente (A06), son alumnos que en general tienen un porcentaje de crédito importante (sobre el 80 %) y no tienen becas asignadas; es importante a posterior analizar estos casos de forma independiente. En conclusión es relevante evaluar a los alumnos que no tienen becas y que no tienen créditos debido a que estos alumnos son una deserción sin motivo aparente. Además es importante evaluar las causas de la eliminación académica dado que estos alumnos se encuentran en segmentos que tienen altos porcentajes de créditos.

En las reuniones de análisis realizadas en conjunto con la División Académica, una de las dudas recurrentes es cuánto representan los resultados obtenidos en el proceso de postulación para los alumnos desertores, si es una condicionante el puntaje o promedio de notas que han obtenido para postular. Si se analizan de forma rápida los datos se puede determinar que, en general, los alumnos desertores se encuentran descritos principalmente por el puntaje, lugar, preferencia y tipo de selección, ver Figura 4.11.

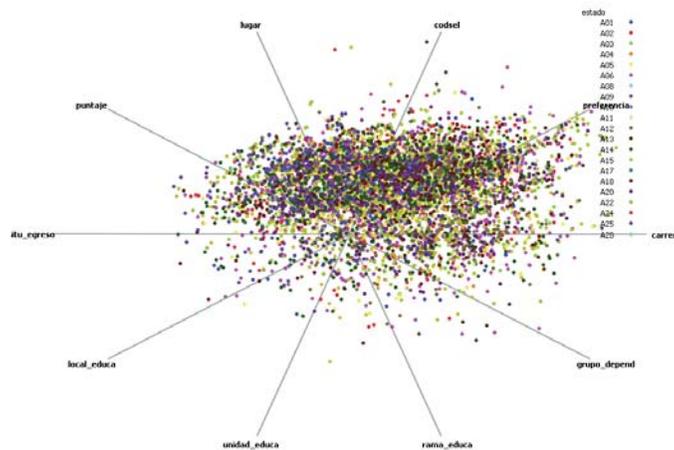


Figura 4.11: Dispersión de los datos de las postulaciones

Una de las primeras inquietudes que despiertan estos atributos es cuáles son sus valores representativos; es válido que los alumnos desertores son alumnos que tienen un promedio bajo y que han ingresado con puntajes inferiores al promedio, donde se tiene una creencia no evaluada que los puntajes de los desertores

son menores a 500 pts. Esto, se revierte al analizar los puntajes de los alumnos, pues como se puede ver en la Figura 4.12 los alumnos desertores mayoritariamente son alumnos que han postulado con puntajes sobre los 500 puntos.

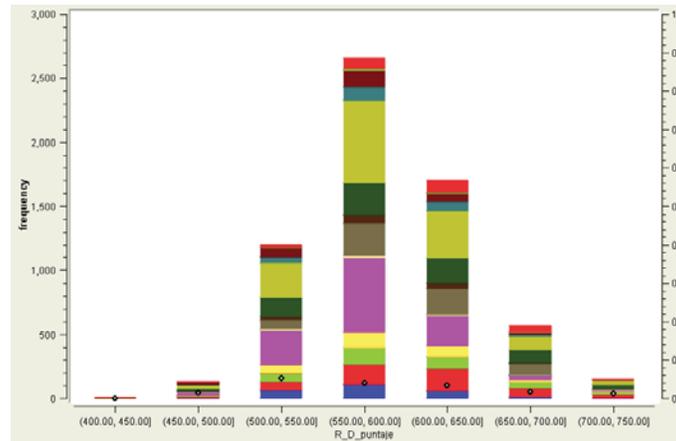


Figura 4.12: Cantidad de Alumnos por Puntaje ponderado

Por otra parte, los alumnos que desertan de la universidad postulan principalmente en las tres primeras prioridades (ver Figura 4.13), con estos resultados se podría concluir que los alumnos que desertan no tienen una plena claridad de las carreras a las cuales ingresan, pero también habría que comparar estos resultados con los promedios de notas obtenidos, ya que se puede deber a un problema de diferencia en su formación obtenida.

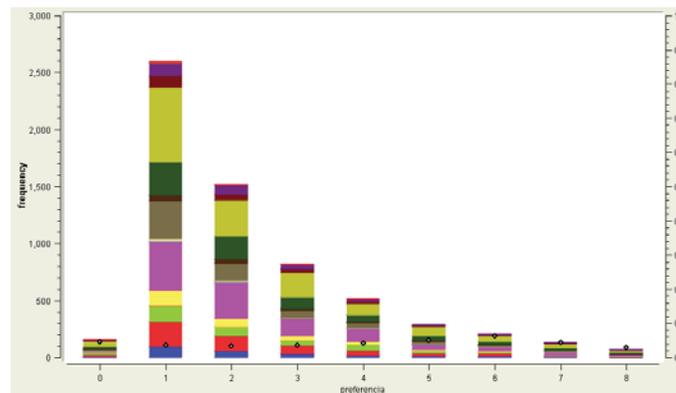


Figura 4.13: Cantidad de Alumnos por Promedio NEM

Frente a estos resultados, importante evaluar de forma independiente a los alumnos que se encuentran en estos segmentos, y en primera instancia sería recomendable mejorar los procesos de admisión y reconocer de forma individual para cada carrera que permita identificar las causales de deserción de los alumnos y aún los motivos por los cuales alumnos de buen rendimiento cumplen con condiciones de ser eliminados o a desertar de las carreras a las cuales ingresan.

Por otra parte también se piensa que los alumnos desertores son alumnos que principalmente no han cursado un segmento científico-humanista y que éstos pertenecen a un grupo de dependencia municipalizado, ambos casos se muestran en la Figura 4.14. En éste ya se puede apreciar que los alumnos desertores son prácticamente alumnos que han estudiado en una estructura científico humanista y en colegios con una dependencia del tipo subvencionado (2).

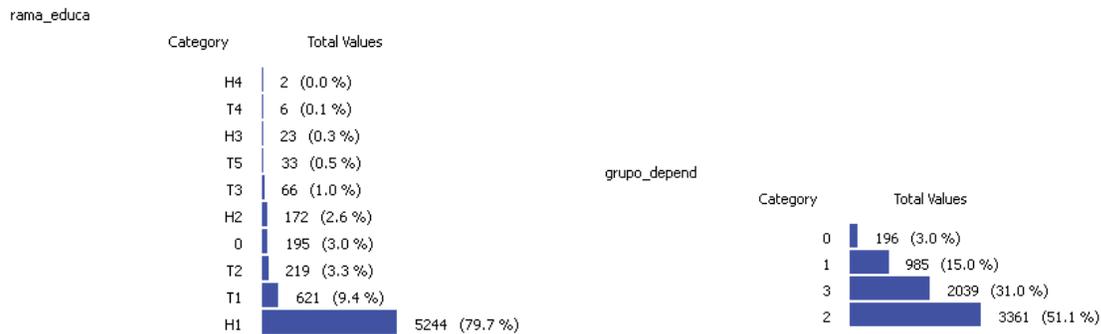


Figura 4.14: Rama y grupo de dependencia de los colegios de procedencia

Por otro lado al realizar un análisis global utilizando el algoritmo de cluster Simple Kmeans, se puede identificar al menos 2 ítemes relevantes en el análisis como se muestra en la Figura 4.15; el primero de ellos es que los alumnos eliminados académicamente (A06), son alumnos que son seleccionados en una posición mayor al puesto 110, con un puntaje entre 550 y 600 puntos, con un buen puntaje en Historia y con un mal puntaje en Ciencias, lo que hace necesario en el análisis a futuro evaluar las carreras a las cuales pertenecen los alumnos eliminados.

En segundo lugar los alumnos que desertan al semestre siguiente, donde simplemente no cancelan o vuelven a matricularse, siendo su último estado "a la espera de cancelación del derecho básico de matrícula (A15)", se podrían considerar como buenos alumnos pues son seleccionados en los lugares superiores al puesto 30 y con un puntaje de ingreso similar a los eliminados académicamente, con la diferencia que obtuvieron buenos resultados en la prueba de Ciencias. Estos puntos hacen relevante evaluar a las carreras científicas, más que a las humanística, debido que la prueba de Ciencias juega un rol importante para los alumnos desertores.

| Cluster centroids: | | | | | |
|--------------------|---------------------|-------------------------|------------------|------------------|------------------|
| Attribute | Full Data (6581) | Cluster# 0 (1680) | 1 (887) | 2 (1223) | 3 (814) |
| preferencia | 1 | 2 | 1 | 1 | 2 |
| code1 | 5 | 5 | 5 | 5 | 5 |
| D_supe | <-10.00 | (10.00, 20.00] | >110.00 | >110.00 | (60.00, 70.00] |
| D_puntaje | (550.00, 600.00] | (550.00, 600.00] | (550.00, 600.00] | (500.00, 550.00] | (550.00, 600.00] |
| estu_egreso | 1 | 1 | 1 | 5 | 5 |
| D_local_educa | <-334.00 | (334.00, 401.00] | <-334.00 | (334.00, 401.00] | >1396.00 |
| unidad_educa | 1 | 1 | 1 | 1 | 1 |
| suma_educa | NI | NI | NI | NI | NI |
| grupo_depend | 2 | 2 | 1 | 2 | 2 |
| D_region | (4.00, 5.00] | (4.00, 5.00] | (4.00, 5.00] | (4.00, 5.00] | (12.00, 13.00] |
| D_provincia | (42.50, 51.50] | (42.50, 51.50] | (42.50, 51.50] | (42.50, 51.50] | >79.50 |
| D_promota | (5.50, 6.00] | (5.50, 6.00] | (5.50, 6.00] | (5.50, 6.00] | (6.00, 6.50] |
| D_prijmas | (550.00, 600.00] | (600.00, 650.00] | (550.00, 600.00] | (550.00, 600.00] | (650.00, 700.00] |
| D_languaje | (550.00, 600.00] | (550.00, 600.00] | (600.00, 650.00] | (500.00, 550.00] | (550.00, 600.00] |
| D_matematica | (550.00, 600.00] | (550.00, 600.00] | (550.00, 600.00] | (500.00, 550.00] | (550.00, 600.00] |
| D_historia_pco | <-400.00 | <-400.00 | (550.00, 600.00] | (500.00, 550.00] | (600.00, 650.00] |
| D_ciencia | <-400.00 | <-400.00 | <-400.00 | <-400.00 | <-400.00 |
| D_Formalia | (550.00, 600.00] | (550.00, 600.00] | (550.00, 600.00] | (500.00, 550.00] | (550.00, 600.00] |
| D_ingreso_bruto | <3186999.00 | <3186999.00 | <3186999.00 | <3186999.00 | <3186999.00 |
| estado | A15 | A15 | A06 | A06 | A15 |

Figura 4.15: Cluster con simple K-Means a las postulaciones

Al evaluar posibles casos anómalos se encuentran datos que claramente son más representativos de un posible error de digitación que un caso real, debido a que se trata de alumnos que superan los máximos lugares de matrículas normales: una posición sobre el puesto 700, lo que está fuera de la normalidad dado que los listados normales llegan a un máximo promedio del lugar 250; este ejemplo se muestra en la Figura 4.16 o valores créditos asignados que corresponden a 1000 % y 945 %, por lo cual serán, eliminados del análisis.

E

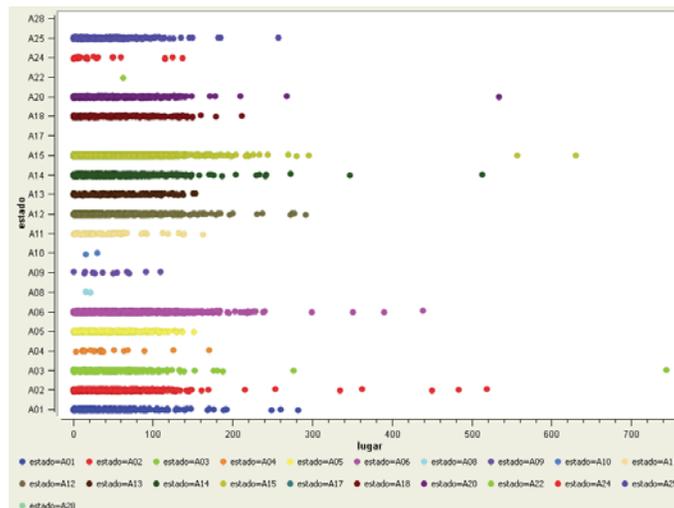


Figura 4.16: Outlier postulacion alumnos

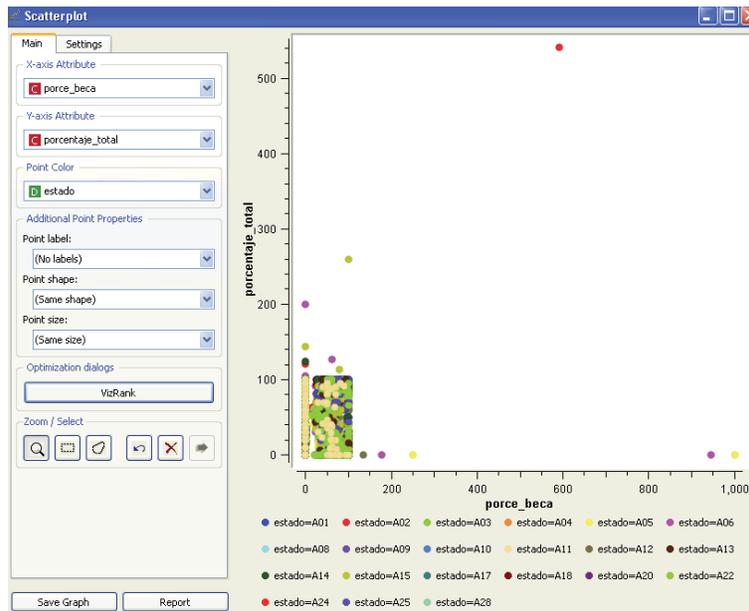


Figura 4.17: Outliers de los beneficios de los alumnos

Otro caso se encuentra en la Figura 4.18 donde se visualiza un caso de un alumno que congelo por mucho tiempo y retorno a finalizar sus estudios pero aún no ha finalizado todo su proceso dado que no ha solicitado el certificado de título o grado.

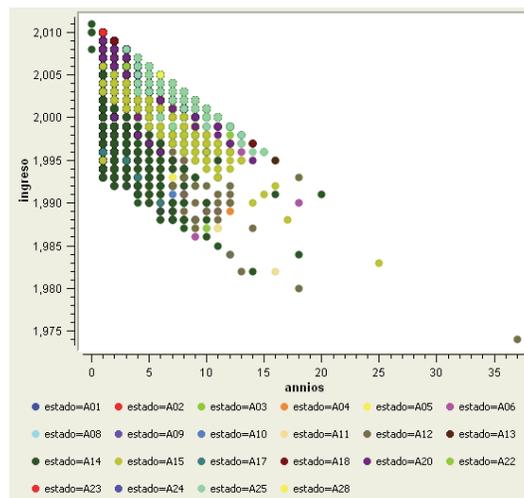


Figura 4.18: Caso anómalo en la cantidad de años cursados por los alumnos

4.3. Conclusiones

En base al análisis exploratorio de los datos se ha identificado las primeras respuestas a las preguntas que se han presentado, como son:

- ¿En qué año de la carrera ocurre la deserción?

Con los resultados se puede definir que los alumnos principalmente desertan en los dos primeros años.

- ¿Cuál es la relevancia de las notas ante los desertores?

Tanto en el proceso de postulación como en el transcurso de la carrera, las notas y puntajes obtenidos describen a los alumnos desertores, donde los promedios de los desertores se encuentran, principalmente, entre un 3.0 y un 4.0. Además de este resultado toma gran relevancia los resultados de las pruebas de Ciencias, debido a que los alumnos que han obtenido puntajes más bajos en estas pruebas han desertado más que los que obtienen puntajes menores en la prueba de Historia. Esto es importante dado que los alumnos eliminados académicamente en mayor número son alumnos de las carreras matemáticas.

- ¿Existe representación de los atributos de los alumnos en el proceso de postulación?

Por medio de Kmeans se pudo determinar que efectivamente los resultados con los que postula el alumno son representativos para describir el avance del alumno; se identifica que los alumnos que son eliminados académicamente en general son aquellos que son seleccionados en lugares superiores al 110, con puntajes bajos.

- ¿Existe una condición económica que represente a los alumnos desertores?

La gran mayoría de los alumnos desertores está representada por medio de aquellos que no tienen becas asignadas y en si toma gran relevancia los atrasos de los alumnos y el porcentaje de créditos obtenidos.

- ¿Los colegios de los cuales provienen los alumnos son determinantes en la deserción?

Más que la rama del colegio, el tipo de dependencia si juega un rol importante, dado que los alumnos desertores están representados principalmente por los colegios de tipo subvencionados y municipalizados.

En base a estas respuestas se puede determinar que los atributos más importantes que permiten describir a los alumnos desertores son:

1. Años Cursados
2. Vía de ingreso
3. Estado Académico Final
4. Porcentaje de Beca
5. Porcentaje de crédito
6. Cantidad de Atrasos

7. Promedio de notas
8. Nota mínima
9. Veces en que cursa una asignatura
10. Puntaje
11. Promedio NEM
12. Lugar
13. Tipo Colegio
14. Rama colegio
15. Preferencia

Los atributos posibilitan definir una vista minable que permita identificar los agentes principales que afectan que un alumno deje de pertenecer a la universidad, desertando de sus estudios o simplemente siendo eliminado académicamente. Además un análisis más en detalle permitirá identificar las causas de la eliminación académica de los alumnos y validar estrategias que permitan mejorar las oportunidades de los alumnos.

Capítulo 5

Análisis de las Carreras

5.1. Introducción

Al realizar el análisis exploratorio de los datos, se identificaron 15 atributos relevantes para el estudio. Al mismo tiempo el análisis ha permitido identificar las carreras relevantes que serán exploradas en profundidad, para lo cual se ha realizado un análisis con respecto a los siguientes atributos:

- Cantidad de eliminados: número representativo de alumnos eliminados que permita realizar el análisis.
- Cantidad de alumnos: la carreras deben contener un número importante de muestras que permitan identificar un clasificado de deserción
- Tipo de carrera: dado que la Universidad tiene una gran gama de carreras, se determinó evaluar una carrera de cada una de las áreas más fuertes como son el Diseño, Ingeniería, Salud y Ciencias Económicas.
- Continuidad de la carrera: la condicionante más fuerte del análisis se centra en contar con una carrera que no se encuentre en proceso de cambio de estructura curricular.

Utilizando estas restricciones se determinó que las carreras más representativas para un análisis en profundidad son:

- Arquitectura: debido a que es una de las más antiguas y no ha sufrido grandes modificaciones en su malla curricular; por otra parte tiene un enfoque mixto entre el arte y la matemática.
- Auditoría (diurno): carrera que tiene una gran cantidad de desertores, pertenece al mundo de las finanzas.
- Ingeniería en Construcción: ingeniería sin cambios estructurales en su malla curricular y que está entre las que tiene la mayor cantidad de desertores en su tipo.
- Enfermería: por lo general las carreras del área de la salud tienen una alta tasa de retención; en este caso, aunque la tasa de deserción no es tan alta, sí está dentro de las mayores en carreras del área de la salud. Por otra parte tiene la ventaja de ser una carrera con menciones, lo que permitirá evaluar este caso.

Para la descripción de cada una de las carreras se utilizará una secuencia de evaluación visual de los datos y a posterior se aplicarán técnicas de evaluación predictiva, fundamentalmente se utilizarán algoritmos basados en árboles que permitan realizar una clasificación de los datos, de tal forma de identificar las principales causas de la deserción que se encuentran descritas en las bases de datos.

Al evaluar los atributos más destacados del análisis visual realizado en el Capítulo 4, se puede identificar la necesidad de contar con los 15 atributos más relevantes, pero al mismo tiempo toman una gran relevancia otros atributos, que podrían permitir identificar al alumno desertor, los que se han identificado como duda de los indicadores institucionales, por lo que se hace necesario incorporarlos al conjunto de datos que se evaluará, estos atributos son:

1. Cantidad de asignaturas cursadas en primera oportunidad.
2. Cantidad de asignaturas cursadas en segunda, tercera, cuarta o más oportunidades.
3. Cantidad de asignaturas aprobadas en primera.
4. Cantidad de asignaturas total aprobadas.
5. Cantidad de asignaturas total reprobadas.
6. Procedencia del alumno ciudad y región.

Con esta nueva estructura se realiza el análisis de cada una de las carreras donde se presenta en profundidad a continuación en análisis de la carrera de Ingeniería en Construcción, para las carreras de Auditoría (Diurna), Arquitectura y Enfermería se presentan los resultados finales, el proceso inicial se encuentra en los anexos.

5.2. Análisis de la deserción en la Carrera de Ingeniería en Construcción

El análisis involucra a todos los alumnos que han ingresado a la carrera entre los años 2004 y 2010. Pese a que existen alumnos en años anteriores a ese período, el inicio de la prueba de selección universitaria PSU se realizó por primera vez en el año 2004, de ahí el conjunto de datos de trabajo.

5.2.1. Análisis Exploratorio

Al revisar la información de la carrera se puede visualizar que el semestre no tiene una gran relevancia debido a que la cantidad de desertores en ambos semestres es muy similar como muestra la Figura 5.1. Además se muestra que el semestre no tiene impacto en los alumnos que no desertan, dado que ellos se encuentran agrupados en el primer semestre, donde en el mejor de los casos culmina su carrera, como se muestra en la Figura 5.2.

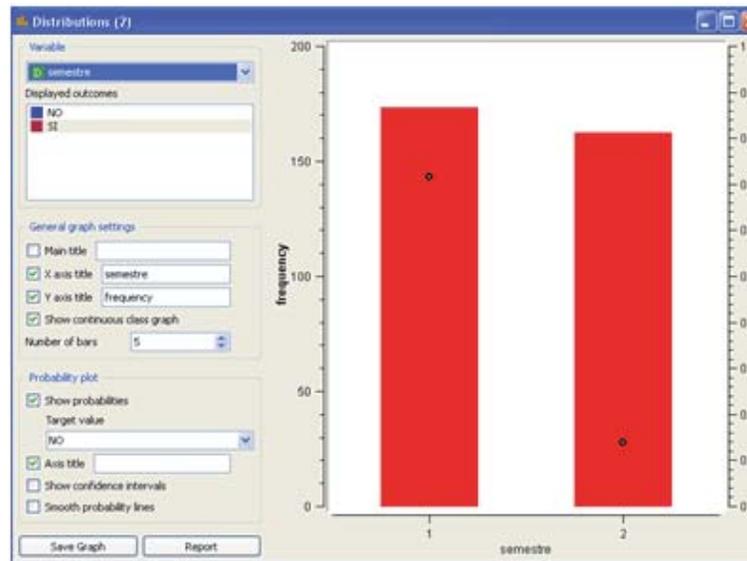


Figura 5.1: Cantidad de Desertores por Semestre

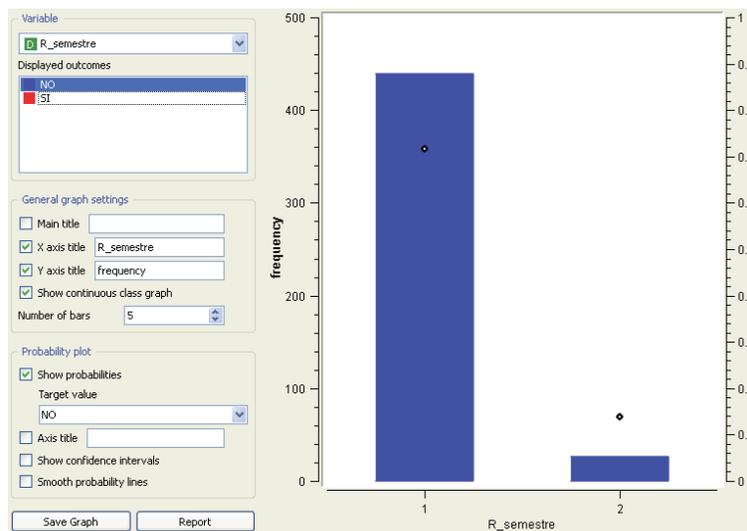


Figura 5.2: Cantidad de alumnos que NO desertan a la carrera por semestre

Un punto importante de los desertores de la carrera se centra en que, mayoritariamente, los alumnos desertan antes del tercer año de permanencia en la carrera, donde se destaca que los alumnos con un año en la carrera superan a todo el grupo de desertores, se puede ver en la Figura 5.3. Junto a lo anterior es importante evaluar a los alumnos que desertan en su primer año o menos sin intentar recuperar su condición de alumno regular. Sería muy interesante a futuro evaluar las causas por las que en particular desertan estos alumnos, dado que muchos de ellos se podrían ser alumnos con problemas vocacionales o con falta de información al momento de postular.

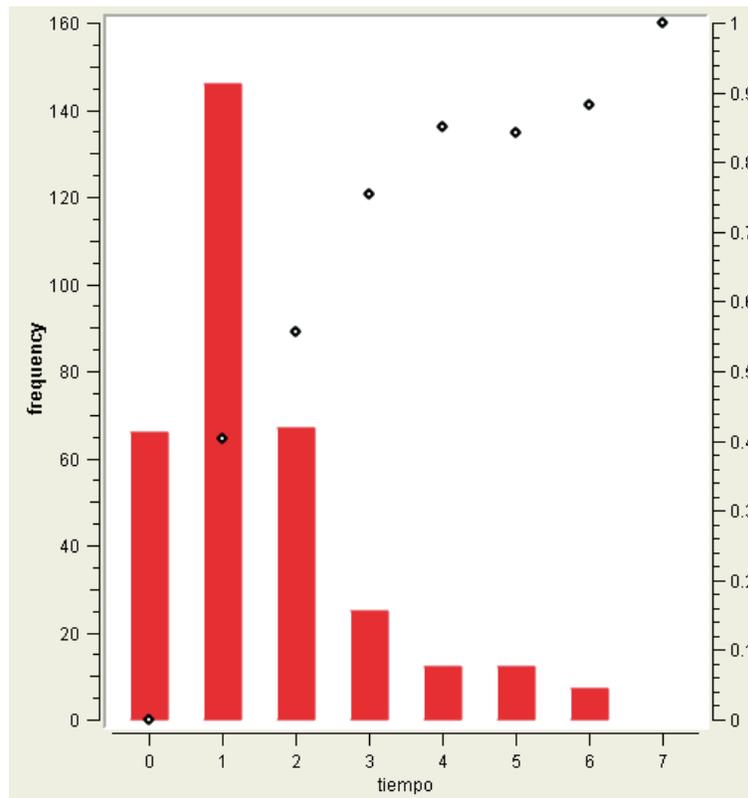


Figura 5.3: Cantidad de desertores por años

La deserción temprana de los alumnos de la carrera, preocupa a la rectoría académica que requiere identificar los factores que afectan a esto, para tomar medidas que lo eviten. En cursos más avanzados esta decisión se encuentra en la mayoría de los casos directamente asociada a la eliminación académica, ver Figura 5.4. La eliminación académica se puede deber a las bajas calificaciones que presentan los alumnos desertores, dado que mayoritariamente, sus promedios se encuentran entre un 3,0 y un 4,0, y si se evalúa la totalidad de los promedios estos se encuentran en valores menores a un 4,0, como muestra la Figura 5.5. Esto da pie a considerar, nuevamente, la pregunta ¿cual asignatura causa este problema?.

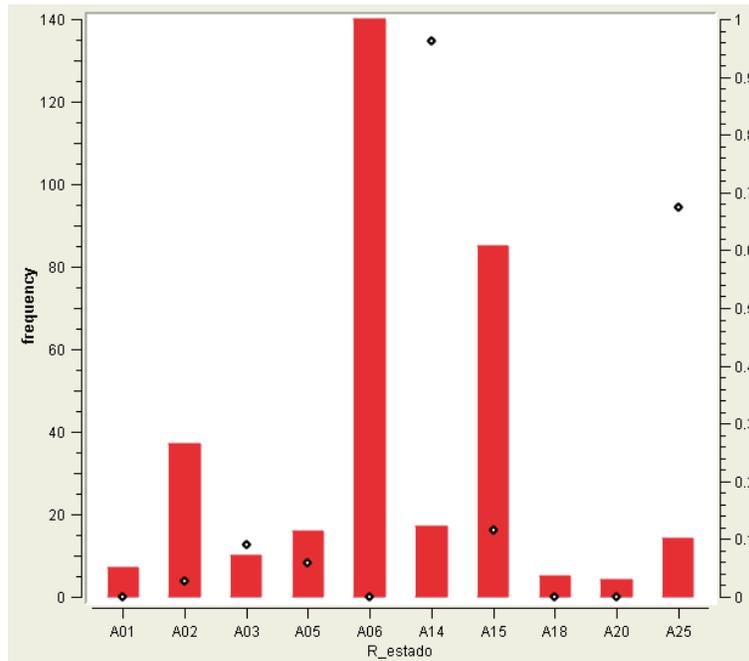


Figura 5.4: Cantidad de desertores por estados

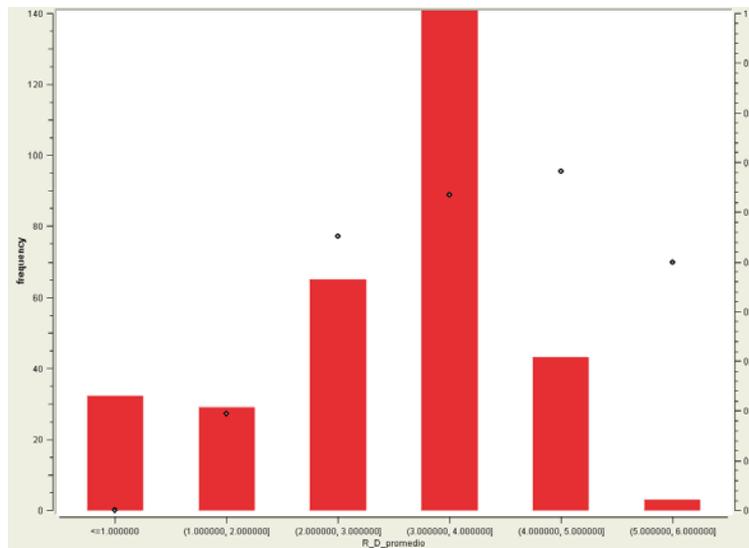


Figura 5.5: Promedio de Notas Alumnos Desertores

Junto a lo anterior se valida que los alumnos desertores, tienen una muy baja cantidad de asignaturas aprobadas, como muestra la Figura 5.6, la que es menor a 5 asignaturas, lo que no debe ser así, en especial en alumnos de curso superior que deberían tener un mejor avance y con mayor complejidad en alumnos de los primeros años.

El problema mencionado anteriormente, se encuentra que en promedio un alumno desertor no aprueba la cantidad de asignaturas que debería en sus primeros semestres, lo que hace interesante estudiar como los alumnos enfrentan las asignaturas o si existe implicancia de las asignaturas en los resultados obtenidos.

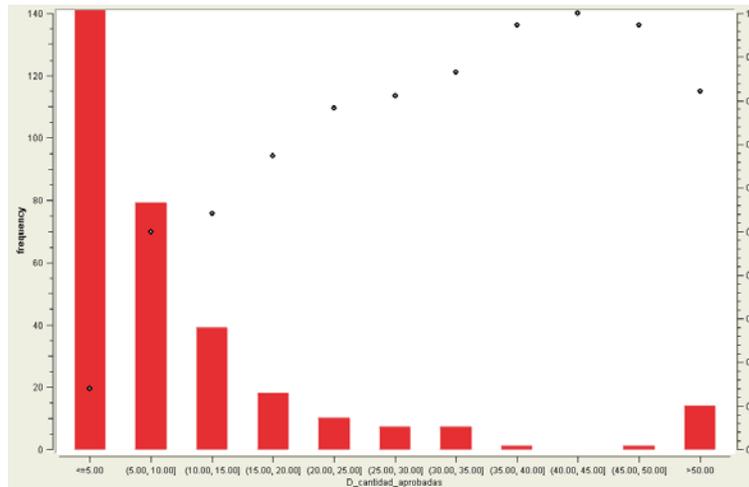


Figura 5.6: Cantidad de asignaturas aprobadas

En resumen, el problema principal se trata de una falencia académica por parte de los alumnos, los que en su mayoría, han cursado asignaturas por tercera oportunidad como muestra la Figura 5.7, que en la Universidad según reglamentos es causal de eliminación académica .

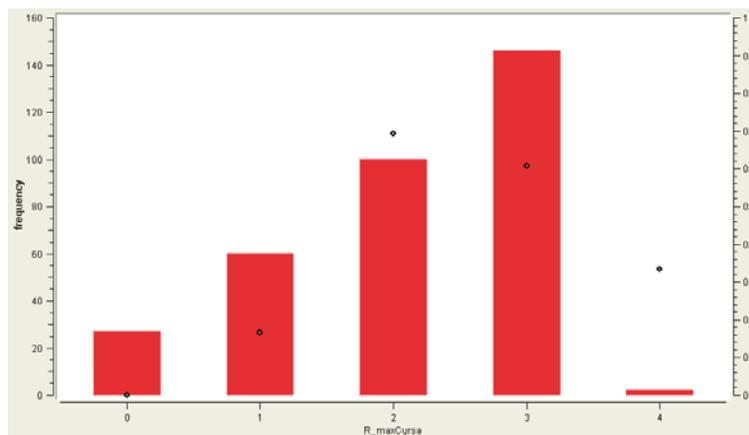


Figura 5.7: Cantidad de desertores por máxima vez cursada una asignatura

Junto a lo anterior, se identifica que los alumnos desertores presentan una gran cantidad de asignaturas reprobadas, donde se encuentran casos que tienen de 8 a 10 reprobaciones de sus asignaturas cursada (ver Figura 5.8), aunque parece un número menor para un alumno de gran avance. Toma relevancia, entonces, visualizar que la mayor cantidad de alumnos que desertan en la carrera son de primer año, por lo cual este análisis inicial se destaca la necesidad de evaluar las asignaturas de la carrera y chequear cuáles son las más

y menos reprobadas.

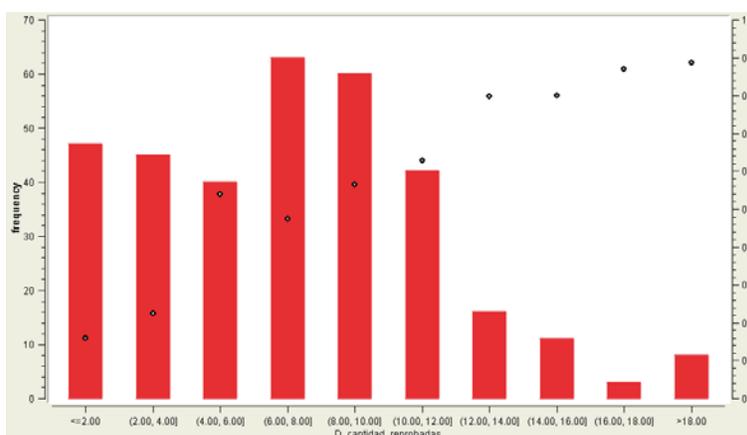


Figura 5.8: Cantidad de asignaturas reprobadas de los alumnos desertores

Debido al bajo aporte de la información obtenida para la carrera, donde principalmente se indica que los alumnos que han reprobado hasta 8 asignaturas y que han cursado menos de 10 asignaturas desertan de la carrera. Se identifica que el avance académico y la referencia del tiempo en la universidad son importante en el momento de desertar. utilizando los datos anteriores se identifican indicadores que hacen referencia al avance académico de los alumnos y que se hacen necesario incorporarlos al análisis en reemplazo de los atributos de conteo, por lo que se excluyen los atributos de cantidad de asignaturas aprobadas, reprobadas y cantidad de asignaturas en segunda y primera, reemplazando estos valores por atributos que permitan posicionar al alumno, en base a sus valores con respecto a toda la muestra.

Los valores considerados se refieren a tasas de avance desde el punto de vista del tiempo y la cantidad de asignatura y tasas de aprobación en distinta oportunidades, las tasas definidas son las siguientes:

1. Tasa de avance en la carrera: que corresponde a la cantidad de asignaturas aprobadas sobre la cantidad total de asignaturas; este valor se encuentra en porcentaje y se ha calculado de la siguiente manera.

$$\frac{\text{cantidad_de_asignaturas_aprobadas}}{\text{cantidad_total_de_asignaturas_carrera}} * 100 \quad (5.1)$$

2. Tasa de avance según año de ingreso: corresponde a la cantidad de asignaturas aprobadas por el alumno, con respecto a la cantidad de asignaturas que debería tener el alumno aprobadas según la cantidad de años que tiene en la carrera.

$$\frac{\text{cantidad_de_asignaturas_aprobadas}}{\text{cantidad_total_de_asignaturas_al_año_cursado}} * 100 \quad (5.2)$$

3. Tasa de aprobación total: corresponde a la cantidad de asignaturas aprobadas versus la cantidad de asignaturas cursadas total.

$$\frac{\text{cantidad_de_asignaturas_aprobadas}}{\text{cantidad_total_de_asignaturas_cursadas}} * 100 \quad (5.3)$$

4. Tasa de reprobación: corresponde a la cantidad de asignaturas reprobadas con respecto a la cantidad de asignaturas cursadas.

$$\frac{\text{cantidad_de_asignaturas_reprobadas}}{\text{cantidad_total_de_asignaturas_cursadas}} * 100 \quad (5.4)$$

5. Tasa de aprobación en primera oportunidad: muy relevante según la evaluación, hecha corresponde a la cantidad de asignaturas aprobadas en primera vez con respecto al total de asignaturas cursadas.

$$\frac{\text{cantidad_de_asignaturas_aprobadas_en_primera_oportunidad}}{\text{cantidad_total_de_asignaturas_cursadas}} * 100 \quad (5.5)$$

6. Tasa de aprobación en segunda oportunidad, aunque se implementó para evaluar el impacto de las asignaturas inscritas por segunda vez con respecto a la cantidad total de asignaturas cursadas.

$$\frac{\text{cantidad_de_asignaturas_aprobadas_en_segunda_oportunidad}}{\text{cantidad_total_de_asignaturas_cursadas}} * 100 \quad (5.6)$$

Al realizar este cambio en el análisis de la carrera se puede identificar la correlación que tienen las asignaturas con respecto a avance académico y la aprobación total real del mismo, lo que permite identificar de mejor forma el verdadero impacto de los antecedentes académicos de los alumnos. Con estos nuevos antecedentes se ha logrado identificar que la tasa de avance total de la carrera toma una gran relevancia en la descripción de los alumnos desertores, como se puede identificar en la Figura 5.9. En esta se puede apreciar que la deserción de los alumnos ocurre principalmente en aquellos con un avance en su carrera menor al 20 % del total de ésta; al considerar que dicha carrera tiene un total de 62 asignaturas, los alumnos desertores son quienes no sobrepasan las 12 primeras de ellas. Al contrastar esta información con respecto a la malla curricular se puede identificar que los alumnos no superan el tercer semestre de la carrera, situación que se puede resaltar con la Figura 5.10.

Se puede identificar que los alumnos desertores principalmente no superan el 20 % del avance personal, en contraste con los alumnos regulares que tienen una tasa de avance de entre un 30 % y un 80 %. Un caso muy particular a ser destacado es que existe un porcentaje no menor de alumnos que se encuentran sobre el 90 % de avance y no terminan su carrera; muchos de estos casos se pueden encontrar en aquellos que pese a que son alumnos de últimos años no terminan su plan de estudios. Lo que impacta directamente en la tasa de titulación de la carrera y los indicadores universitarios presentes en las acreditaciones.

Por otra parte los alumnos que desertan de la carrera no superan el 60 % de tasa de aprobación de sus asignaturas, donde la mayor cantidad de estos alumnos no supera el 50 % de tasa de aprobación, siendo esto un contraste con los alumnos que no desertan de la carrera que se tienen tasas de aprobación sobre el 70 % de sus asignaturas. Como se puede ver en la Figura 5.2.1.

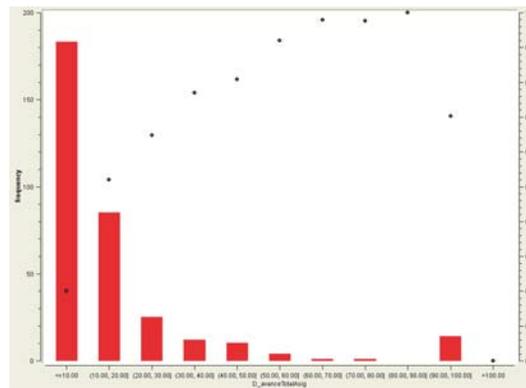


Figura 5.9: Tasa de avance total de los alumnos de la carrera A

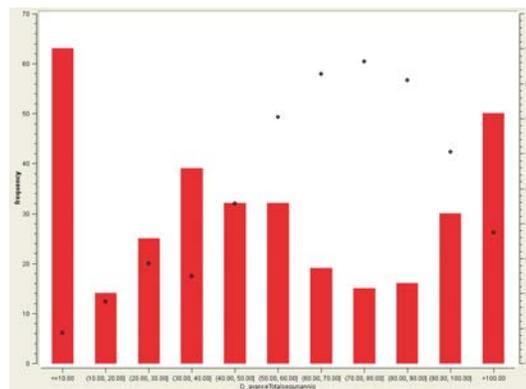


Figura 5.10: Tasa de avance según año de los alumnos de la carrera A

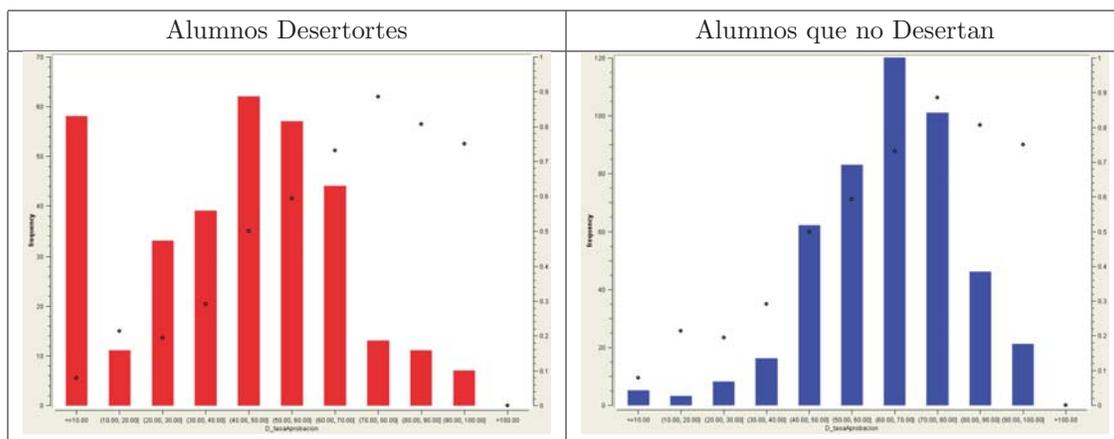


Figura 5.11: Comparativa tasa de aprobación alumnos

Un punto importante en este nuevo análisis se encuentra en las problemáticas económicas que presentan

los alumnos con respecto a su apoyo en financiamiento, debido a que prácticamente todos los alumnos desertores tienen créditos y becas menores al 10 %, como se observa en la Figura 5.12. El financiamiento se contraponen en gran medida con los alumnos que no desertan donde el gran porcentaje de ellos tiene entre el 90 % y 100 % de crédito, lo que hace destacar las problemáticas económicas de los alumnos al momento de estudiar una carrera en cursos superiores.

Aunque no se refleja este mismo caso, las becas son un ítem muy importante para los alumnos de la carrera de Ingeniería en Construcción al momento de tomar la decisión de desertar. Como último punto se puede identificar que los alumnos desertores se encuentran seleccionados principalmente en lugares mayores al 100, lo que refleja que sus puntajes son bajos o están dentro de la escalas finales de selección, lo que permitiría suponer que su preparación previa a la universidad en el ámbito educativo es menor a lo esperado.

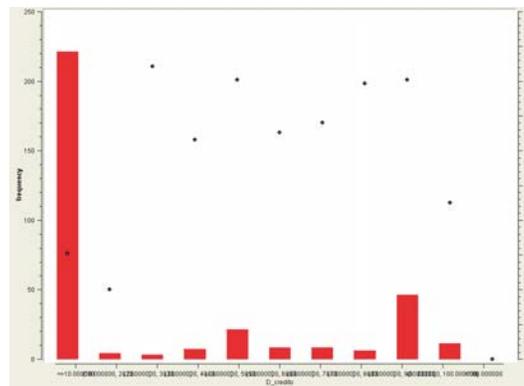


Figura 5.12: Créditos asignados a alumnos desertores de la Ingeniería en Construcción

5.2.2. Minería de Datos

5.2.2.1. Clasificación

El proceso de clasificación se ha realizado de forma incremental, los primeros patrones obtenidos, aunque interesantes ante las características de los alumnos, son muy obvios, pero han reflejado la necesidad de crear nuevos atributos basados en las tasas de avance y aprobación. Esto último, ha entregado nuevos resultados destacando algunos de estos atributos que se pueden visualizar en el estudio, pese a esto en cada una de las evaluaciones presentadas en esta sección se han identificado patrones con un soporte bajo, pero que dan lugar a la descripción de los atributos que se han aplicado en la poda de los árboles, donde se han obtenido las reglas finales, que dan lugar al patrón de descripción de la deserción en cada carrera.

La primera tarea de esta evaluación en cada una de las etapas se centra en reducir la cantidad de atributos iniciales y que permiten obtener los atributos representativo o de mayor aporte ante la clase

Para seleccionar el grupo de atributos en la aplicación Orange se utiliza la propiedad estadística de ganancia de información, que permite, por medio de la clase, definida por deserta si o no) clasificar a los ejemplos, utilizando este caso se determino utilizar los 15 atributos más relevantes, con respecto al valor de la clase, los que se muestran en la Figura 5.13.

La selección de estos atributos se realiza en base a aquellos que tienen una ganancia de información mayor a 0,03, permitiendo obtener 15 atributos representativos, seleccionados por aquellos que reflejan

| Attribute | # | ReliefF | Inf. gain | Gain ratio | Gini | og OR (for 'SI') |
|----------------------|----|-----------|-----------|------------|----------|------------------|
| 1 estado | 10 | 0.787292 | 0.718435 | 0.347973 | 0.195221 | NA |
| 2 D_primera | 13 | 0.183417 | 0.322290 | 0.097064 | 0.093576 | NA |
| 3 D_cantidad_apro... | 10 | 0.167625 | 0.266571 | 0.098085 | 0.081374 | NA |
| 4 tiempo | 8 | 0.188428 | 0.233244 | 0.085743 | 0.066269 | NA |
| 5 semestre | 3 | 0.327540 | 0.187195 | 0.237972 | 0.060197 | NA |
| 6 D_credito | 6 | 0.142498 | 0.154378 | 0.082089 | 0.050776 | NA |
| 7 D_cantidad_repr... | 10 | 0.043875 | 0.135151 | 0.042360 | 0.042416 | NA |
| 8 D_segunda | 10 | 0.058102 | 0.117086 | 0.046896 | 0.037097 | NA |
| 9 maxCursa | 5 | 0.040398 | 0.107238 | 0.069206 | 0.032570 | NA |
| 10 D_promedio | 8 | 0.088876 | 0.081070 | 0.045297 | 0.023703 | NA |
| 11 D_minimo | 9 | -0.033613 | 0.053604 | 0.033203 | 0.017002 | NA |
| 12 strasos | 10 | 0.066739 | 0.038870 | 0.014553 | 0.012713 | NA |
| 13 via_ingreso | 8 | 0.012959 | 0.036787 | 0.031471 | 0.011236 | NA |
| 14 D_lugar | 15 | 0.006207 | 0.031876 | 0.008511 | 0.010652 | NA |
| 15 tipocolegio | 9 | 0.013903 | 0.018540 | 0.017331 | 0.005796 | NA |
| 16 D_region | 17 | 0.016516 | 0.017662 | 0.010069 | 0.005310 | NA |
| 17 D_puntaje | 10 | 0.016697 | 0.016906 | 0.011824 | 0.005555 | NA |
| 18 D_puntajenem | 10 | 0.042827 | 0.012785 | 0.004670 | 0.003933 | NA |
| 19 D_beca | 6 | 0.038740 | 0.010286 | 0.014711 | 0.002468 | NA |
| 20 pref | 9 | 0.044355 | 0.009601 | 0.003796 | 0.003155 | NA |
| 21 estadoseleccion | 3 | 0.015481 | 0.007658 | 0.010803 | 0.002472 | NA |
| 22 grupo_dependen... | 4 | 0.047005 | 0.007606 | 0.004485 | 0.002479 | NA |
| 23 tercera | 5 | -0.125099 | 0.002685 | 0.001545 | 0.000898 | NA |
| 24 terceraCursa | 5 | -0.125099 | 0.002685 | 0.001545 | 0.000898 | NA |
| 25 cuarta | 2 | 0.000000 | 0.000676 | 0.018983 | 0.000231 | 1.024889 (1) |
| 26 mas_de_cuatro | 1 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | NA |

Figura 5.13: Ranking de atributos más relevantes de Ingeniería en Construcción

mejores resultados de aporte de ganancia de información. En base a estos atributos se construyen árboles que permite realizar la clasificación de los atributos.

Al aplicar el algoritmo ID3 para obtención de un árbol de clasificación, en una primera instancia se puede evidenciar que existen al menos, tres ramas que resultan interesantes. En la Figura 5.14 se observa que la estructura está compuesta en su raíz por los estados académicos de los alumnos, lo que es coherente dado que ellos establecen la condición de un alumno, donde los estados de no continuidad como abandono, suspensión, postergación y eliminación académica, clasifican a los alumnos como desertores.

| Classified information | Class | P(Class) | P(Target) | # Inst. | Distribution (rel) | Distribution (abs) |
|----------------------------|-------|----------|-----------|---------|--------------------|--------------------|
| estado = A01 | SI | 1.000 | 0.000 | 7 | 0.000:1.000 | 0:7 |
| estado = A02 | SI | 0.974 | 0.026 | 38 | 0.000:0.998 | 1:37 |
| estado = A03 | SI | 0.909 | 0.091 | 11 | 0.001:0.909 | 1:10 |
| estado = A04 | SI | 0.941 | 0.059 | 17 | 0.000:0.941 | 1:16 |
| estado = A06 | SI | 1.000 | 0.000 | 140 | 0.000:1.000 | 0:140 |
| estado = A14 | NO | 0.961 | 0.961 | 439 | 0.961:0.979 | 422:17 |
| semestre = 1 | NO | 0.995 | 0.995 | 420 | 0.995:0.995 | 419:1 |
| semestre = 2 | SI | 0.789 | 0.211 | 19 | 0.211:0.789 | 4:15 |
| D_promedio <= (1.00, 2.00) | SI | 1.000 | 0.000 | 7 | 0.000:1.000 | 0:7 |
| D_promedio <= (2.00, 3.00) | SI | 1.000 | 0.000 | 3 | 0.000:1.000 | 0:3 |
| D_promedio <= (3.00, 4.00) | NO | 1.000 | 1.000 | 4 | 1.000:0.000 | 4:0 |
| via_ingreso = 100 | SI | 0.685 | 0.115 | 96 | 0.115:0.676 | 13:83 |
| via_ingreso = 100 | SI | 1.000 | 0.000 | 62 | 0.000:1.000 | 0:62 |
| maxCursa = 1 | SI | 0.900 | 0.100 | 10 | 0.000:0.900 | 1:9 |
| maxCursa = 2 | NO | 0.612 | 0.612 | 13 | 0.612:0.300 | 9:4 |
| maxCursa = 3 | NO | 0.500 | 0.500 | 2 | 0.500:0.500 | 1:1 |
| estado = A01 | SI | 1.000 | 0.000 | 5 | 0.000:1.000 | 0:5 |
| estado = A02 | SI | 1.000 | 0.000 | 4 | 0.000:1.000 | 0:4 |
| estado = A03 | NO | 0.874 | 0.874 | 43 | 0.674:0.526 | 29:14 |
| D_promedio <= (1.00, 2.00) | SI | 0.538 | 0.462 | 13 | 0.462:0.538 | 6:7 |
| semestre = 1 | NO | 1.000 | 1.000 | 3 | 1.000:0.000 | 3:0 |
| semestre = 2 | SI | 0.700 | 0.300 | 10 | 0.300:0.700 | 3:7 |
| maxCursa = 2 | NO | 0.600 | 0.600 | 5 | 0.600:0.400 | 2:3 |
| maxCursa = 3 | SI | 1.000 | 0.000 | 5 | 0.000:1.000 | 0:5 |
| D_promedio <= (2.00, 3.00) | NO | 0.714 | 0.714 | 14 | 0.714:0.586 | 10:4 |
| D_promedio <= (3.00, 4.00) | NO | 0.917 | 0.917 | 12 | 0.917:0.083 | 11:1 |
| D_promedio <= (3.00, 4.00) | NO | 0.500 | 0.500 | 4 | 0.500:0.500 | 2:2 |

Figura 5.14: Estructura árbol ID3 de la clasificación

Casos importantes y de análisis son los alumnos regulares que en su último semestre han desertado. Este caso se puede deber a dos causas que pueden ser alumnos que en el segundo semestre que han obtenido notas menores a 4.0 no continúan en la carrera o son alumnos que pese a terminar no ha finalizado su proceso de titulación, por que sería recomendable un seguimiento por parte de las autoridades respectivas.

Lo anterior preocupa en este análisis en especial con los alumnos que se encuentra en una estado de "A la espera de examen de título". debido a que estos alumnos se encuentran con sus cursos terminados pero no han realizado el proceso terminal de la carrera. Un caso particular se encuentra en alumnos que tiene máxima asignatura cursada por tercera encontrándose con su carrera terminada como muestra el árbol obtenido mediante el algoritmo CART de la Figura 5.15

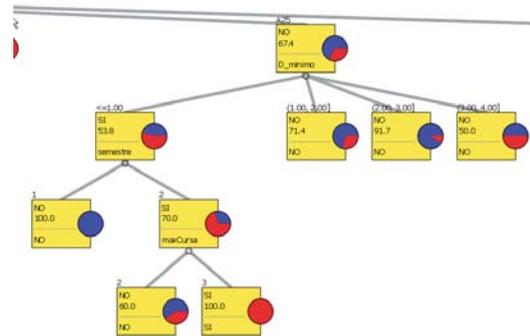


Figura 5.15: Clasificación desertores con máxima asignatura por tercera

Aunque entrega puntos importantes, este clasificador permite evaluar en primera instancia a los alumnos de la carrera, donde se presentan casos complejos y genéricos que permiten ayudar en la observación de los alumnos y tomar medidas que permitan evitar la deserción de los alumnos. Los resultados de la clasificación tras aplicar el algoritmo CART, como se muestra en la matriz de confusión en la Figura 5.16, son muy buenos dado que presenta bajos valores de mal clasificados.

| | NO | SI | |
|----|------------|------------|------------|
| NO | 96.9 % | 5.9 % | 465 |
| SI | 3.1 % | 94.1 % | 335 |
| | 459 | 341 | 800 |

Figura 5.16: Matriz de Confusión del clasificador por árbol CART

Al evaluar los resultados de la clasificación se identifica que las reglas obtenidas en esta primera aproximación, son mejor mediante el árbol CART, lo que se puede ver en la matriz de confusión de la aplicación de los algoritmos CART y C4.5, tal como se presenta en la Figura 5.2.2.1. Se debe destacar que el árbol CART tiene una mayor profundidad en comparación al árbol C4.5, lo que permite obtener mejores patrones o reglas de clasificación.

El problema del clasificador C4.5 está en que solamente entrega como resultados un nivel que corresponde estados de los alumno, por lo que entrega resultados muy obvios; en base a estos resultados se identifica que la variable estados no tiene un gran aporte a la descripción de la condición de los alumnos. Por esto se retira de los datos, para volver a analizar nuevos resultados.

| CART | | | | C4.5 | | | |
|------|------------|------------|------------|------|------------|------------|------------|
| | NO | SI | | | NO | SI | |
| NO | 96.9 % | 5.9 % | 465 | NO | 93.6 % | 4.4 % | 465 |
| SI | 3.1 % | 94.1 % | 335 | SI | 6.4 % | 95.6 % | 335 |
| | 459 | 341 | 800 | | 482 | 318 | 800 |

Figura 5.17: Comparativa entre matrices de confusión de la clasificación CART y C4.5

La mayor ventaja de la nueva clasificación se encuentra en que se obtuvo un árbol con mayor profundidad donde se obtiene para CART un árbol mejor definido.

La clasificación presenta como resultado para el atributo raíz la cantidad de asignaturas por primera vez que ha cursado en alumno, y obtiene en primer lugar que los alumnos que han cursado en primera vez menos de dos asignaturas ya son casos que desertarán inmediatamente, tal como muestra la Figura 5.18.

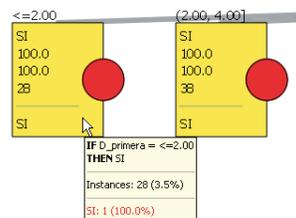


Figura 5.18: Cantidad de asignaturas cursadas por primera

Un caso importante se encuentra en los alumnos que han cursado entre 20 y 40 asignaturas por primera vez desertan de la carrera, aunque los casos no son muchos (cobertura de un 2,9%) tienen un 100 % de confianza (Figura 5.19), lo que hace relevante evaluar una referencia entre la cantidad de asignaturas aprobadas del total que debería tener como una tasa de aprobación .

Lo anterior entrega reglas que deben ser validadas al momento de realizar seguimiento de los alumnos. Al evaluar los casos se puede identificar se ha visto la gran injerencia que tienen los estados, lo que puede deber a que el gran porcentaje de los alumnos se encuentran en estado de "alumno regular" representando el 54.8% del total de los alumnos, lo que hace necesario evaluar la clasificación de estos casos por separado.

Al evaluar la clasificación, utilizando los alumnos que no se encuentra en estado de "alumno regular" se observa que las variables puntaje y región y dejan de ser parte del ser las variables de atrasos y lugar.

Con los resultados, se obtienen como reglas importantes la descripción del árbol de clasificación que se muestra en la Figura 5.20. En base a estos resultados se identifica que los alumnos desertores han cursado por primera vez entre 6 y 8 asignaturas, ya han cursado algunas asignaturas por segunda y que su puntaje de ingreso es menor a 550.

Además van a ser desertores la totalidad de los alumnos que habiendo cursado entre 8 y 10 asignaturas aprueben menos de 8 de ellas, un caso importante se encuentra en los alumnos que han cursado más de 50 asignaturas y han rendido por segunda 12 de ellas y tienen al menos una asignatura por tercera cursada,

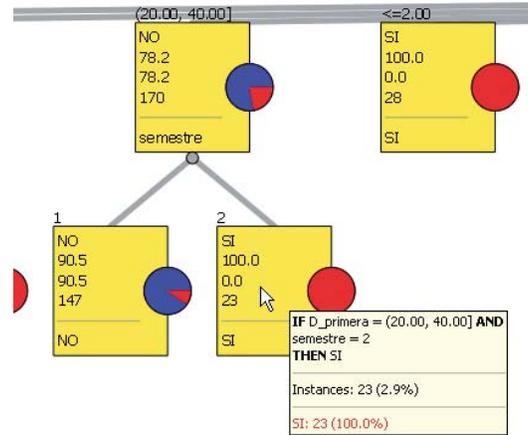


Figura 5.19: Semestre que cursa el alumno al ser desertor

tienen una alta probabilidad de desertar, esto es complejo dado que son alumnos de cursos superiores que han avanzado bastante en su carrera.

| Classification Tree | Class | P(Class) | P(Target) | # Inst | Distribution (ref) | Distribution (abs) |
|--------------------------------------|-------|----------|-----------|--------|--------------------|--------------------|
| <root> | SI | 0.881 | 0.119 | 361 | 0.119-0.881 | 43:318 |
| D_primera <= 2.00 | SI | 1.000 | 0.000 | 28 | 0.000-1.000 | 0:28 |
| D_primera = (2.00, 4.00] | SI | 1.000 | 0.000 | 30 | 0.000-1.000 | 0:30 |
| D_primera = (4.00, 6.00] | NO | 0.500 | 0.500 | 4 | 0.500-0.500 | 2:2 |
| D_primera = (6.00, 8.00] | SI | 0.850 | 0.150 | 20 | 0.150-0.850 | 3:17 |
| maxCurso = 1 | SI | 1.000 | 0.000 | 3 | 0.000-1.000 | 0:3 |
| maxCurso = 2 | SI | 0.786 | 0.214 | 14 | 0.214-0.786 | 3:11 |
| R_D_puntaje = (500.00000, 55.00000] | NO | 0.500 | 0.500 | 4 | 0.500-0.500 | 2:2 |
| R_estado = A06 | SI | 1.000 | 0.000 | 2 | 0.000-1.000 | 0:2 |
| R_estado = A15 | NO | 1.000 | 1.000 | 2 | 1.000-0.000 | 2:0 |
| R_D_puntaje = (950.00000, 60.00000] | SI | 0.900 | 0.100 | 10 | 0.100-0.900 | 1:9 |
| maxCurso = 3 | SI | 1.000 | 0.000 | 3 | 0.000-1.000 | 0:3 |
| D_primera = (8.00, 10.00] | SI | 0.951 | 0.049 | 61 | 0.049-0.951 | 4:57 |
| D_cantidad_aprobadas <= 2.00 | SI | 1.000 | 0.000 | 11 | 0.000-1.000 | 0:11 |
| D_cantidad_aprobadas = (2.00, 4.00] | SI | 0.966 | 0.034 | 29 | 0.034-0.966 | 1:28 |
| D_cantidad_aprobadas = (4.00, 6.00] | SI | 1.000 | 0.000 | 30 | 0.000-1.000 | 0:30 |
| D_cantidad_aprobadas = (6.00, 8.00] | SI | 0.889 | 0.111 | 9 | 0.111-0.889 | 1:8 |
| D_cantidad_aprobadas = (8.00, 10.00] | NO | 1.000 | 1.000 | 2 | 1.000-0.000 | 2:0 |
| D_primera = (10.00, 12.00] | SI | 1.000 | 0.000 | 36 | 0.000-1.000 | 0:36 |
| D_primera = (12.00, 14.00] | SI | 1.000 | 0.000 | 30 | 0.000-1.000 | 0:30 |
| D_primera = (14.00, 16.00] | SI | 0.960 | 0.040 | 25 | 0.040-0.960 | 1:24 |
| D_primera = (16.00, 18.00] | SI | 1.000 | 0.000 | 10 | 0.000-1.000 | 0:10 |
| D_primera = (18.00, 20.00] | SI | 1.000 | 0.000 | 4 | 0.000-1.000 | 0:4 |
| D_primera = (20.00, 40.00] | SI | 0.944 | 0.056 | 36 | 0.056-0.944 | 2:34 |
| D_primera = (40.00, 50.00] | SI | 0.750 | 0.250 | 4 | 0.250-0.750 | 1:3 |
| D_primera > 50.00 | NO | 0.667 | 0.667 | 45 | 0.667-0.333 | 30:15 |
| R_D_puntaje <= 1.00 | SI | 0.538 | 0.462 | 13 | 0.462-0.538 | 6:7 |
| maxCurso = 2 | NO | 0.667 | 0.667 | 6 | 0.667-0.333 | 4:2 |
| maxCurso = 3 | SI | 0.714 | 0.286 | 7 | 0.286-0.714 | 2:5 |
| R_D_segunda = (6.00, 8.00] | NO | 0.667 | 0.667 | 3 | 0.667-0.333 | 2:1 |
| R_D_segunda = (10.00, 12.00] | SI | 1.000 | 0.000 | 4 | 0.000-1.000 | 0:4 |
| R_D_minimo = (1.00, 2.00] | NO | 0.688 | 0.688 | 16 | 0.688-0.312 | 11:5 |
| R_D_minimo = (2.00, 3.00] | NO | 0.917 | 0.917 | 12 | 0.917-0.083 | 11:1 |
| R_D_minimo = (3.00, 4.00] | NO | 0.500 | 0.500 | 4 | 0.500-0.500 | 2:2 |

Figura 5.20: Árbol de clasificación de alumnos desertores sin alumnos regulares

Esta primera evaluación identifica que la deserción de los alumnos de la carrera de Ingeniería en Construcción se produce principalmente por causas académicas, las cuales son:

- Se encuentran en los dos primeros años de la carrera.
- Tienen créditos asignados en todos sus periodos menores a un 10
- Han cursado menos de 10 asignaturas en su carrera.
- Han reprobado hasta 8 de éstas.
- Su puntaje de ingreso es menor a 550 puntos.

Estos factores iniciales permiten declarar a los alumnos con estas características, presentan un alto riesgo de deserción y que la mayor cantidad de condicionantes se fundamenta en procesos académicos y no económicos y menos de del proceso de selección. Así se concluye que es importante realizar la evaluación de las asignaturas de los dos primeros años, al menos, para identificar cuáles son las que deben ser reforzadas para evitar que los alumnos deserten y al mismo tiempo se requiere implementar y evaluar tasas de avance y aprobación que permitan reflejar la condición del alumno con respecto al total de la muestra.

| Atributo | Tipo | Descripción |
|---------------------------|----------|---|
| D_avanceTotalAsig | Continuo | Tasa de avance en la carrera |
| D_avanceTotalsegunannio | Continuo | Tasa de avance según añotiempo en la carrera |
| D_beca | Continuo | Porcentaje de beca |
| D_credito | Continuo | Porcentaje de Credito |
| D_lugar | Continuo | Lugar de selección |
| D_minimo | Continuo | Mínima nota obtenida |
| D_promedio | Continuo | Promedio general de notas |
| D_puntaje | Continuo | Puntaje de selección |
| D_puntajenem | Continuo | Puntaje notas de enseñanza media |
| D_region | Discreto | Región de Procedencia |
| D_tasaAprobacion | Continuo | Tasa de aprobación general de asignaturas cursadas |
| D_tasaAprobacionPrimera | Continuo | Tasa de aprobación de asignaturas cursadas en primera oportunidad |
| D_tasaAprobacionenSegunda | Continuo | Tasa de aprobación de asignaturas en segunda oportunidad |
| D_tasaReprobacion | Continuo | Tasa de reprobación de las asignaturas cursadas |
| atrasos | Continuo | Cantidad de veces que se atrasa en los pagos |
| grupo_dependencia | Discreto | Grupo de dependencia del colegio de procedencia particular subencionado, particular, municipalizado, Etc. |
| preferencia | Discreto | Preferencia en la cual realiza postulación a la carrera |
| tercera | Continuo | Cantidad de asignaturas en tercera |
| tiempo | Continuo | Cantidad de años en la carrera |
| tipocolegio | Discreto | Tipo de colegio humanista, científico. Técnico u otro |
| via_ingreso | Discreto | Via por la cual ingresa a la universidad. |

Tabla 5.1: Atributos del conjunto de datos con tasas

Al incorporar al análisis los atributos de los indicadores identificados anteriormente, se ha implementado

un conjunto con los nuevos valores, Este nuevo grupo considera las variables de avance académico por medios de las tasas de avance y aprobación, junto a los aspectos de ingreso como la región de procedencia, puntaje, preferencia entre otros, los que se pueden ver en Tabla 5.1.

Posterior a la discretización de los valores continuos se realizó un ranking para evaluar el aporte de los distintos atributos, donde se han seleccionado para el siguiente trabajo de clasificación todos los atributos que tienen valores de aporte mayores a cero. Al realizar este primer análisis se puede identificar que los antecedentes con los cuales ingresa un alumno no tienen gran aporte o relevancia para la clasificación de los alumnos desertores.

En primer lugar, al aplicar el algoritmo CART se obtiene un árbol de clasificación con una mayor cantidad de niveles que las presentes en el árbol de la sección 5.2.2.1.

En primer lugar este árbol ubica en la raíz el avance total del alumno en la carrera, lo que refleja este atributo para identificar a los alumnos desertores de la carrera (ver en la Figura 5.21). En cuanto a reglas de clasificación, la primera de ellas se encuentra en alumnos con un avance total entre un 30 % y un 40 % con un crédito alto y una tasa de aprobación por primera vez entre un 40 % y 50 %, los cuales son desertores, lo que implica que es importante el avance que el alumno tiene y el cómo aprueba sus asignaturas. Se piensa que los alumnos dejan sus estudios principalmente por problemas económicos pero a la luz de los resultados cabe pensar si es así o no (ver Figura 5.22).

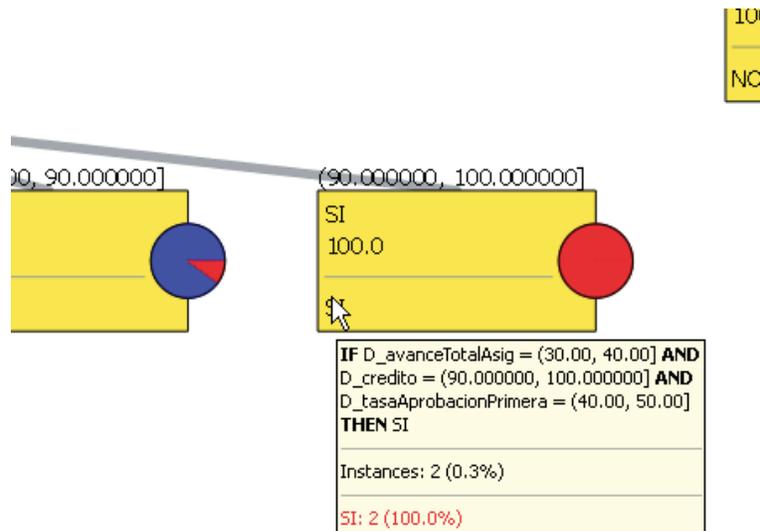


Figura 5.21: Raiz arbol CART carrera A

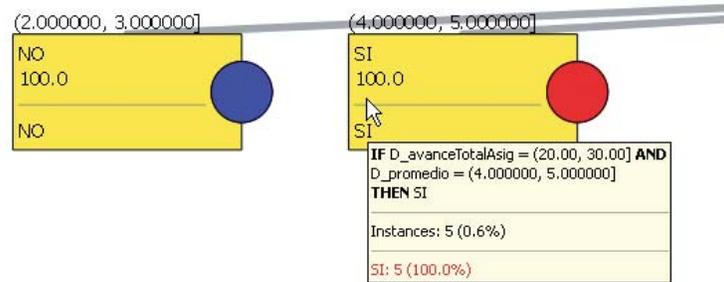


Figura 5.22: Clasificación de avance alumnos desertores

Algunos piensan que el promedio de notas de los alumnos es una causal de deserción, cuando es inferior a un 4.0, pero se dan casos particulares que debe ser visualizado por parte de las autoridades de la carrera debido que alumnos con promedios no de excelencia pero que permiten continuidad, toman la decisión de desertar a la carrera cuando su avance está entre un 20 % y 30 %.

Si se considera que la carrera tiene un total de 62 asignaturas, todo alumno que no ha aprobado más de 19 asignaturas tiene un alto riesgo de deserción; tomando en cuenta aproximadamente 6 asignaturas por semestre, todo alumno que no ha aprobado el segundo año de su carrera está con una alta probabilidad de deserción. Este caso también es reforzado al utilizar el clasificador CN2, que determina con una calidad del 95 % y con una cobertura de un 6,8 %, que los alumnos con un avance menor al 10 % de la carrera pese a estar cursando todas sus asignaturas por primera vez deserta de la carrera. Lo anterior puede indicar problemas en la presentación de la carrera a los alumnos al momento de postular o al momento de presentar su preferencia. Esto se presenta en la Figura 5.23.

| Rule length | Rule quality | Coverage | Predicted class | Distribution | Rule |
|-------------|--------------|----------------|-----------------|--------------|--|
| 13 | 2 | 0.974027454853 | 55.0 | SI | <0,0,55.0> IF D_avanceTotalAsig=['<=10.00'] AND maxCursa=['1'] THEN clase=SI |
| 14 | 2 | 0.787417590618 | 86.0 | SI | <16.0,70.0> IF tiempo=['1'] AND maxCursa=['3'] THEN clase=SI |
| 15 | 2 | 0.901957452297 | 49.0 | NO | <47.0,2.0> IF maxCursa=['2'] AND D_avanceTotalsegunanno=['(60.00, 70.00]'] THEN clase=NO |

Figura 5.23: Descripción con algoritmo CN2 de Ingeniería en Construcción

Lo expuesto en el párrafo anterior es reforzado con otra regla presentada por CN2, con una cobertura de un 8,2 % y una calidad de un 98 %, existe un grupo de alumnos que deserta a la carrera cuando el alumno tiene una tasa de avance menor al 10 %. Basados en el tiempo, el árbol de clasificación presenta dos reglas importantes indicando que la deserción ocurre antes que al alumno finalice el segundo año de estudios, como se muestra en la Figura 5.24. En estos casos se refuerza la decisión cuando los alumnos tienen un promedio de notas entre un 3.0 y 4.0, esta decisión es más drástica en la medida que el alumno se encuentra en los inicios de su carrera.

Por otra parte en muchos análisis se ha identificado como una causa de la deserción la situación socioeconómica del alumno, Esto se valida con los resultados obtenidos de la clasificación realizada por el

algoritmo CART, donde se visualiza que los alumnos que presentan un bajo avance académico, con promedios de notas bajos y que han cursado al menos una asignatura por tercera vez, asociado a un bajo promedio bajo financiamiento con créditos menores a un 10%, el alumno renuncia a la carrera (ver Figura 5.25).

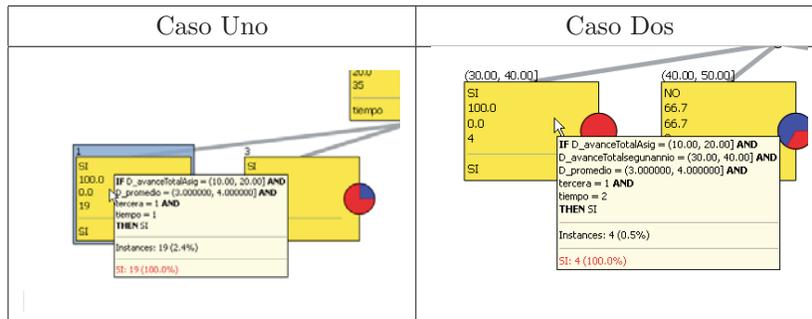


Figura 5.24: Casos en que se presenta baja tasa de avance total carrera en alumnos desertores.

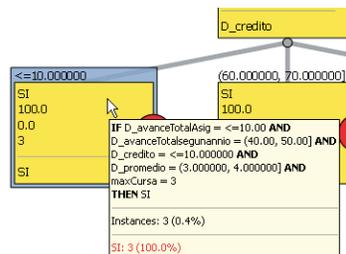


Figura 5.25: Clasificación con tasa de avance y crédito bajo el 10%.

Lo anterior se valida, también, debido a que alumnos con mejores asignaciones de crédito toman una decisión totalmente contraria al contar con un mejor financiamiento como se puede visualizar en la Figura 5.26.

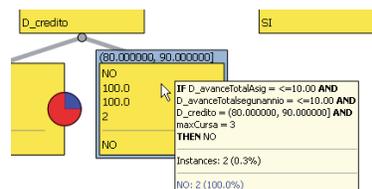


Figura 5.26: Clasificación con tasa de avance y crédito bajo el 10% caso dos.

Al evaluar la clasificación presentada por el algoritmo C4.5, se muestra una menor calidad en las reglas obtenidas debido a que presenta un modelo con menor cantidad de niveles y con menor conocimiento útil.

Sin embargo como conocimiento relevante de este modelo, se obtiene la validación del atributo crédito y la relevancia del financiamiento al momento de tomar la decisión de desertar el alumno, donde se presenta encuentra en sus reglas que alumnos con tasas de avance menor al 20 % y créditos menores al 10 % tienen una alta probabilidad de desertar.

Al ser un modelo más simple que CART el nivel de conocimiento obtenido no es mayor, que el que ya se tiene con respecto a la situación de los alumnos desertores. Debido a lo anterior, se realiza una comparación con los resultados obtenidos por el algoritmo ID3. Para realizar esta evaluación y debido a que la aplicación Orange no cuenta con la implementación de este algoritmo se utiliza la aplicación WEKA.

El algoritmo J48 permite ratificar los resultados obtenidos con los modelos CART y C4.5, dado que se encuentran reglas que identifican como desertores a los alumnos que tienen menos de una 10 % de avance y créditos asignados. (ver Figura 5.27)

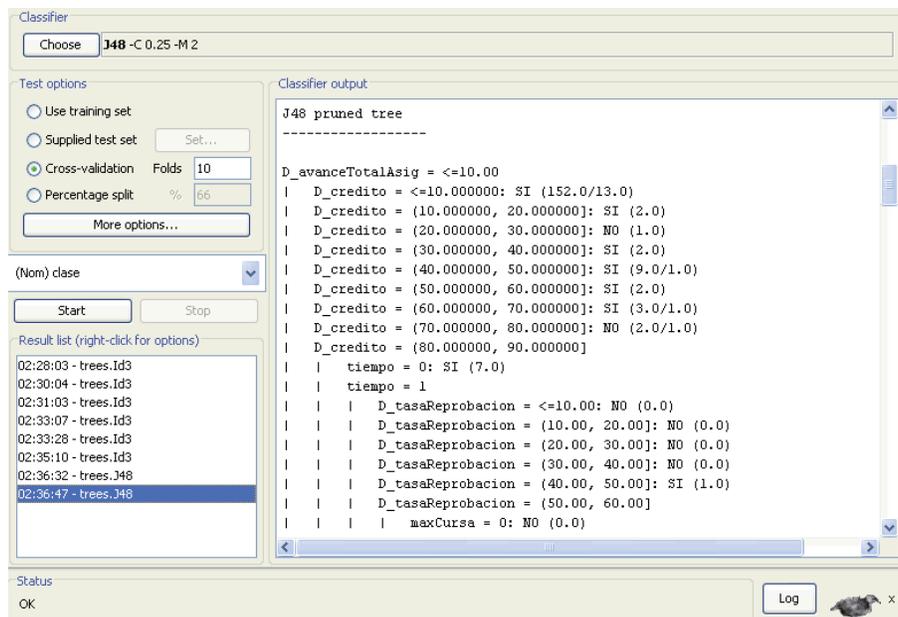


Figura 5.27: Clasificación árbol J48 con WEKA.

Al igual que en casos anteriores se puede ver que ante dos alumnos en la misma condición con respecto a su situación académica la decisión de la deserción del alumno se encuentra en el porcentaje de crédito asignado, para alumnos con mayor apoyo económico, es más probable que no deserte que aquellos que no lo tienen.

5.2.2.2. Evaluación de la clasificación

Aunque ninguno de los algoritmos permitió obtener un resultado bueno, el mejor de los casos obtenidos se encuentra en el algoritmo CART como se puede visualizar en las matrices de confusión presentadas en la Figura 5.28

| J48 | | | C4.5 | | | CART | | |
|-----|-----|-----|------|-------|-----|------|-------|-------|
| | NO | SI | | NO | SI | | NO | SI |
| NO | 85% | 18% | NO | 80,4% | 20% | NO | 83,2% | 17,4% |
| SI | 15% | 82% | SI | 19.6% | 80% | SI | 16,8% | 82,6% |

Figura 5.28: Comparativa de las Matrices de Confusión

Al realizar la evaluación de la carrera se identifica el impacto que existen la tasa de aprobación y de el avance académico para los alumnos desertores, en especial para los alumnos en sus primeros años; debido a esto se torna interesante evaluar las notas y las asignaturas que pueden ser causal de deserción de los alumnos en distintos niveles.

Junto a lo anterior aunque un poco más complejo, se podría identificar si existe un profesor que pueda referenciar mayormente a los desertores, es por esto que extrajo un nuevo conjunto de datos que se presenta en la Tabla 5.2.

| Atributo | Descripción |
|---------------------------|---|
| Nombre Asignatura | nombre de la asignatura que cursa |
| Código asignatura | código de la asignatura que cursa se requiere dado que hay veces en que el código se mantiene y el nombre cambia |
| Semestre de la asignatura | semestre en el cual debe ser cursada la asignatura |
| Cantidad de alumnos | cantidad de alumnos que tiene la asignatura, permite identificar si la deserción de los primeros años está fundamentada en la cantidad de alumnos en sala |
| Tasa de aprobación | tasa de aprobación de la asignatura compuesta por cantidad de aprobados cantidad de alumnos por 100 |
| Profesor | código de profesor de tal forma de identificar que hay casos en que el académico puede ser un impacto a la deserción de los estudiantes. |
| Nota final | nota final obtenida en la asignatura para validar si es un problema de aprendizaje. |
| Cod Evaluación | estado de la asignatura (no determinado, reprobada, aprobada) |
| Cursa | vez en la que cursa la asignatura. |

Tabla 5.2: Atributos para el set de datos de clasificación de la deserción con las evaluaciones de Ingeniería en construcción

Al realizar una evaluación visual, se identifica que existen asignaturas que los alumnos reprueban mayormente lo que implica que deben ser reforzadas: física, cálculo I y álgebra, representadas por las tres mayorías en la Figura 5.29

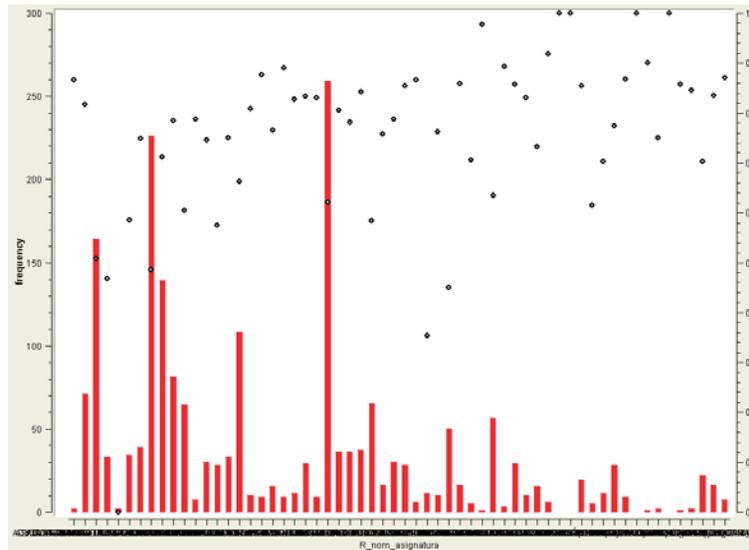


Figura 5.29: Cantidad de asignaturas cursadas alumnos desertores.

Posteriormente se pueden identificar a las asignaturas de Cálculo II y Estadísticas, en forma general, las asignaturas matemáticas son las que representan a los alumnos desertores. Al evaluar a los académicos de estas asignaturas, donde se identifica que sí existen profesores que pueden aportar a la deserción como se muestra en la Figura 5.30. En este gráfico se puede evidenciar, que al menos, dos académicos tienen altas tasas de reprobación y que, evidentemente, pueden influir en la deserción de las asignaturas de los alumnos; como se muestra en la Figura 5.31 la tasa de aprobación, el profesor y la asignatura que se encuentra cursando el alumno tienen una gran importancia en la decisión de desertar.

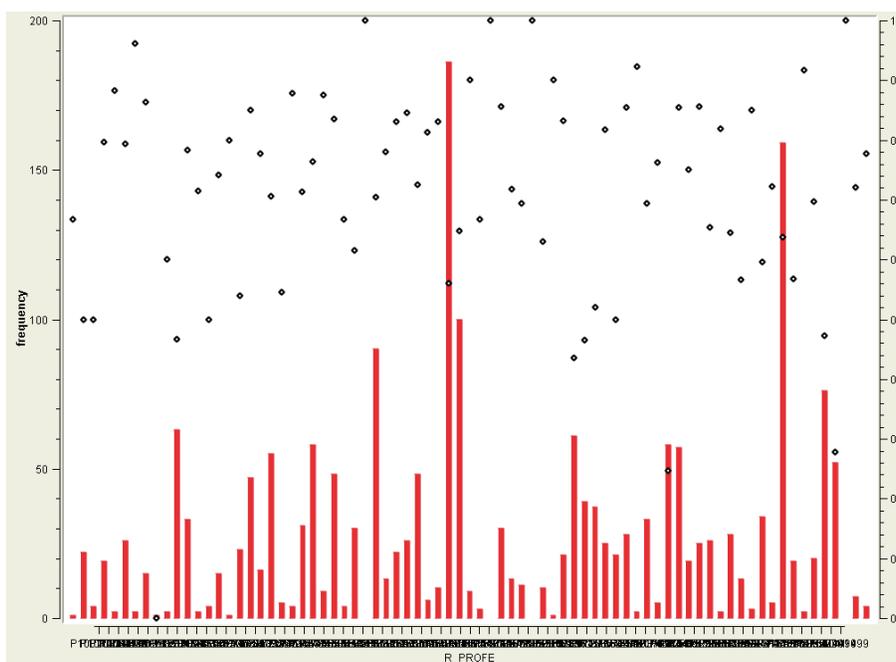


Figura 5.30: Cantidad de desertores por Profesores dictan asignaturas.

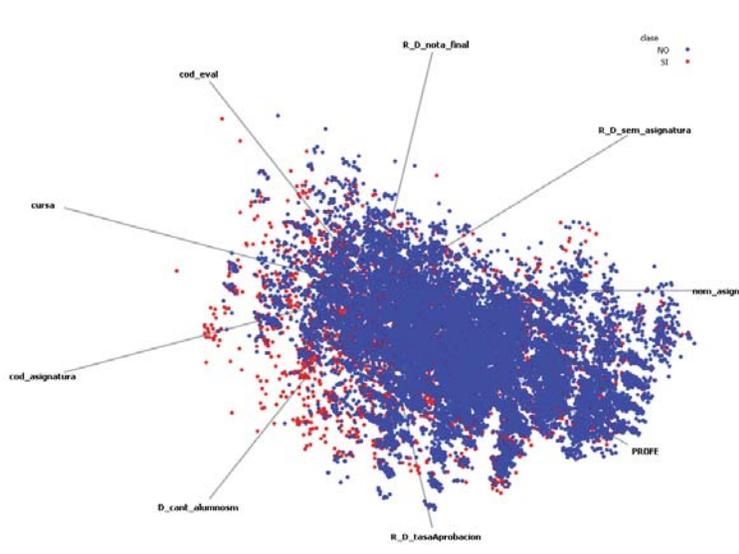


Figura 5.31: Distribución de registros según descripción de las evaluaciones.

Al evaluar la distribución sobre el total de las notas se identifican una serie de descripciones representativas que permiten identificar a los alumnos desertores, como la importancia de tasa de aprobación de los alumnos a la hora de aprobar, donde a mayor tasa de aprobación de la asignatura mayor es la probabilidad continuidad de los alumnos.

Pese a los bajos resultados de la clasificación, que en principio no permiten otorgar resultados definitivos,

se identifican reglas iniciales que podrían ser interesante evaluar en análisis futuros, dado que existen reglas que identifican a el resultado obtenido en ciertas asignaturas influyen en la decisión de desertar, dependiendo si aprueba o reprueba la asignatura(ver Figura 5.32).

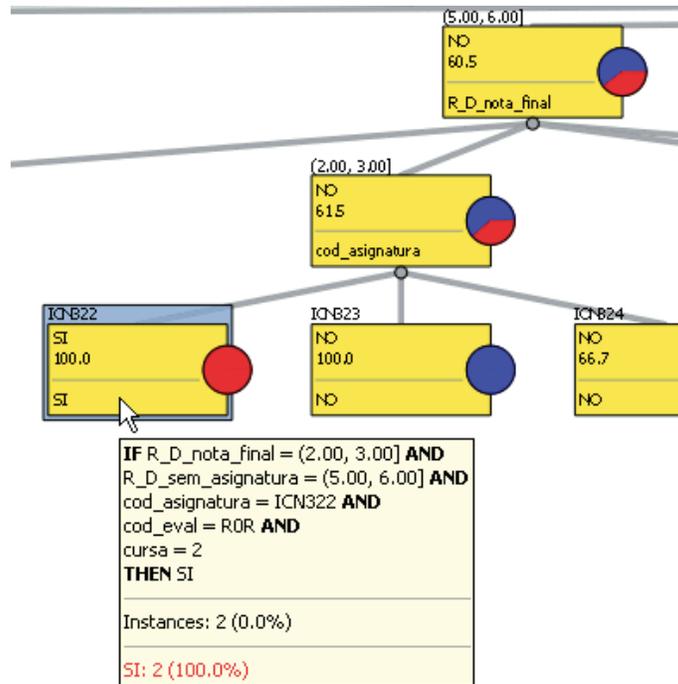


Figura 5.32: Clasificación de alumnos Ingeniería en Construcción y la reprobación de asignaturas

Un caso más complejo se encuentra en reglas determinadas por profesores que determinan la deserción frente a otros que determinan la continuidad de los alumnos como se visualiza en la Figura 5.33. En la figura se puede visualizar que alumnos con los mismos promedios finales, en las mismas asignaturas toman decisiones totalmente dispares dependiendo del académico que dicta la asignatura, aunque presentan un bajo soporte si tienen un 100 % de confianza lo cual debería ser evaluado en base a las debilidades que representan a cada uno de los profesores que dictan las asignaturas, quizás en estos casos sería interesante aplicar un tipo de encuesta que permita identificar el problema.

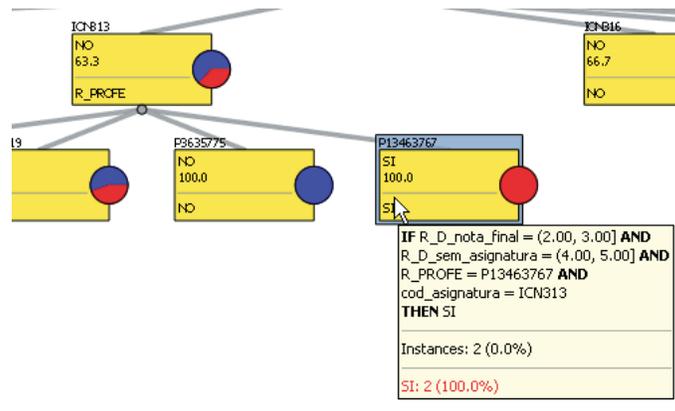


Figura 5.33: Clasificación de alumnos en la carrera A y la descripción de la deserción por el profesor

5.2.2.3. Impacto del Factor Tiempo.

Una de las dudas que nace con el análisis realizado al evaluar las tasas de avance es si el semestre en el cual deserta el alumno y la cantidad de semestres que ha cursado con respecto a los semestres que debería tener tienen importancia en la descripción del desertor; para evaluar esto se incorporan los atributos semestre en el cual deserta el alumno, cantidad de semestres cursados y la tasa de avance fundamentada en semestres.

Al visualizar la cantidad de desertores por semestre, se puede identificar que no es tal como se piensa dado que la cantidad de desertores es muy similar en cada semestre, como se puede visualizar en la Figura 5.34.

Además la tasa de avance según los semestres, se puede identificar que los semestres que ha cursado el alumno tienen una gran relevancia para la deserción debido a que la gran mayoría de los alumnos que desertan son aquellos que tienen una tasa de avance entre un 90% y 100% con respecto a su total, donde los alumnos desertan antes del quinto semestre por lo cual se hace relevante el seguimiento de los alumnos durante este periodo, lo cual se refleja en la Figura 5.35

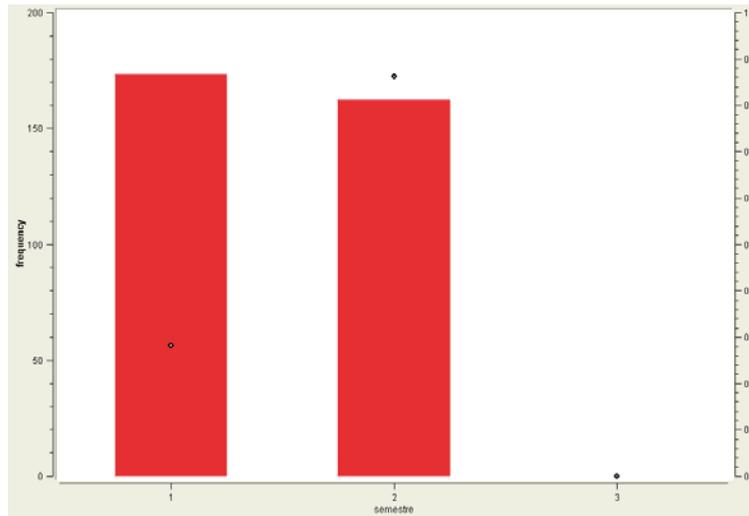


Figura 5.34: Cantidad de alumnos desertores por semestre en la carrera de Ingeniería en Construcción.

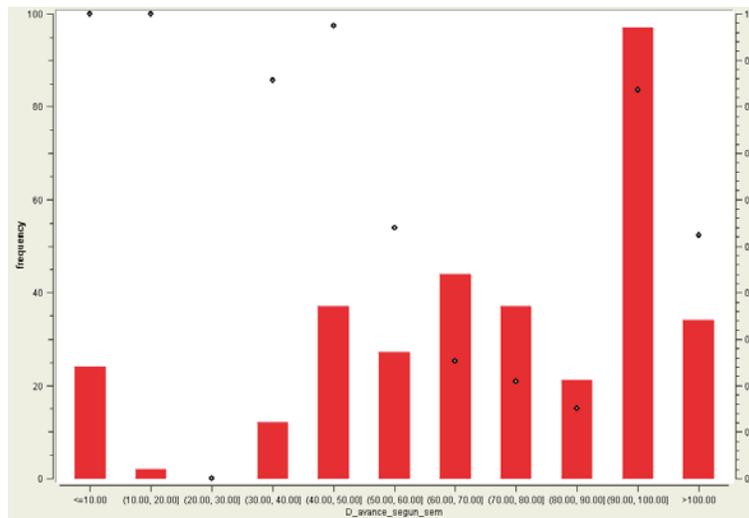


Figura 5.35: tasa de avance en semestre de los alumnos de la carrera Ingeniería en Construcción.

Además de la reevaluación del avance de los alumnos con respecto al avance en semestres también se redefine la tasa de aprobación por primera, segunda y tercera vez, dado que en las primeras evaluaciones se ha definido como una tasa sobre el total de asignatura de la carrera, lo que presentado un impacto fuerte en los alumnos con baja tasa de aprobación por primera vez, presentado nuevos resultados en la evaluación de la clasificación. Algo que se debe evaluar es que existe un número no menor de alumnos que tienen una tasa de aprobación por primera menor a un 10%.

Al realizar la reevaluación por medio de arboles de clasificación se identifica en primer lugar que el semestre si es representativo en la descripción de los alumnos desertores, junto a la tasa de avance de la carrera, el porcentaje de crédito asignado y la tasa de aprobación, según lo identificado al realizar el ranking

de los atributos. Pese a este nuevo análisis se puede identificar la relevancia que tiene la tasa de aprobación y el avance académico en los alumnos desertores, por lo cual se hace necesario evaluar la aprobación de los alumnos en sus asignaturas, que permita identificar las principales causas de los problemas en el avance académico de los alumnos desertores.

Por otra parte, la deserción principalmente ocurre en el primer semestre, lo anterior permite identificar una prioridad de evaluación en la carrera, donde se identifica que la deserción ocurre cuando los alumnos tienen un porcentaje de crédito menor al 10 %, con una baja tasa de avance y una tasa de reprobación normal entre un 20 % y un 30 %. Como se puede visualizar en la Figura 5.36.

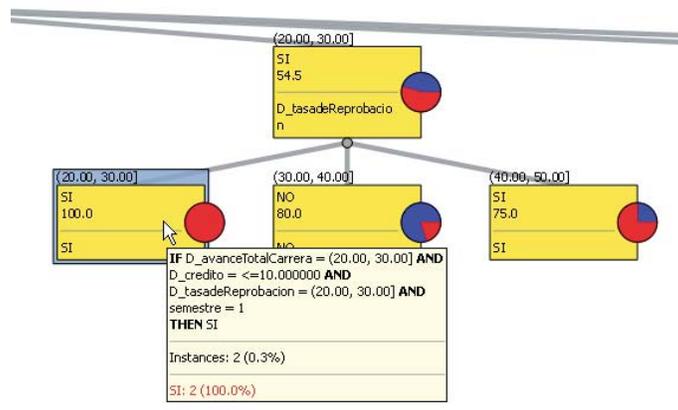


Figura 5.36: Clasificación con tasa de avance y crédito del alumno.

Dado el impacto que tiene la deserción en el primer semestre, se evalúa la clasificación en este semestre de los alumnos desertores, donde no se encuentran reglas importantes de los desertores y permite identificar atributos que permiten clasificar a los alumnos que no desiertan(ver Figura 5.37).

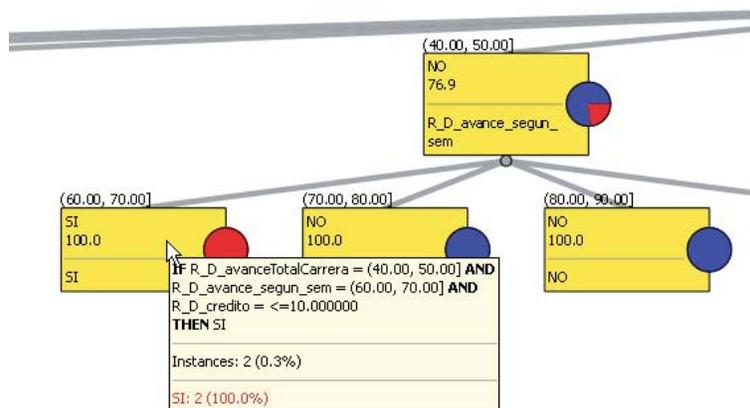


Figura 5.37: Clasificación con tasa de avance y crédito del alumno.

En este caso se muestra que el avance en semestres determina a un alumno desertor a mayor avance

del alumnos que no deserta. Junto a lo anterior, se identifica que la tasa de aprobación por primera vez no desertan cuando existe una mayor tasa de aprobación.

La primera evaluación no entrega mayores resultados al evaluar los atributos con valores únicos, una regla de interés se presenta as continuación donde se identifica que los alumnos con un avance según semestre menor al 50 % tienen una alta probabilidad de desertar a la carrera. Ver Figura 5.38

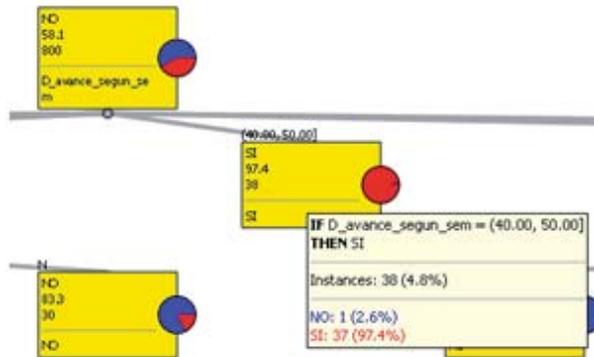


Figura 5.38: Poda clasificacion alumnos desertores y la tasa de avance segun semestre

En esta primera evaluación se identifica que con un soporte de un 4.8% los alumnos desertan en una tasa de avance según semestre menor al 50 %, pero no entrega mayores resultados esperados. Debido a lo anterior y para obtener se realizó una búsqueda exhaustiva, realizando mezclas de los valores de los atributos, identificando nuevas reglas con mayor soporte. La primera regla que se identifica se encuentra en la Figura 5.39, que representa al 7.9% de las instancias aumentado el soporte con 63 instancias de las cuales 58 ellas pertenecen a alumnos desertores que se encuentran identificados por una tasa total de avance de la carrera menor al 10% de la misma forma que la tasa de aprobación, donde esta regla tiene una tasa confianza de un 98.1% de los casos. Por lo anterior, se convierte en una regla muy interesante.

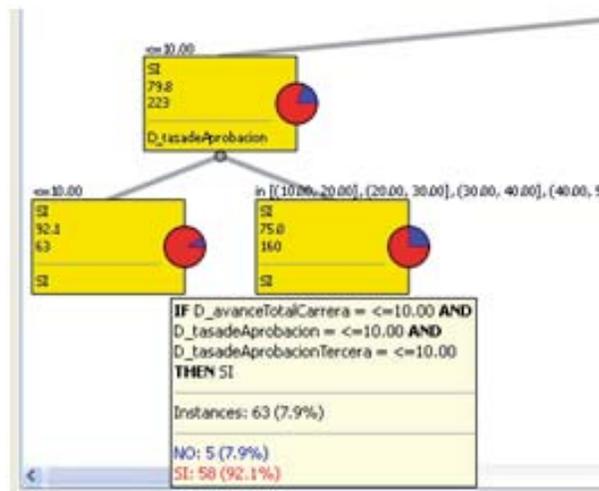


Figura 5.39: Poda clasificacion alumnos desertores y la tasa total de avance

Al evaluar las reglas obtenidas por medios de los arboles se identifica que los alumnos de la carrera de construcción desertan con un soporte de un 6.5% de los casos equivalente a 52 instancias, desertan cuando los alumnos tienen tasa de aprobación por primera y créditos menores al 10% con una confianza de un 96.2%. Ver Figura 5.40

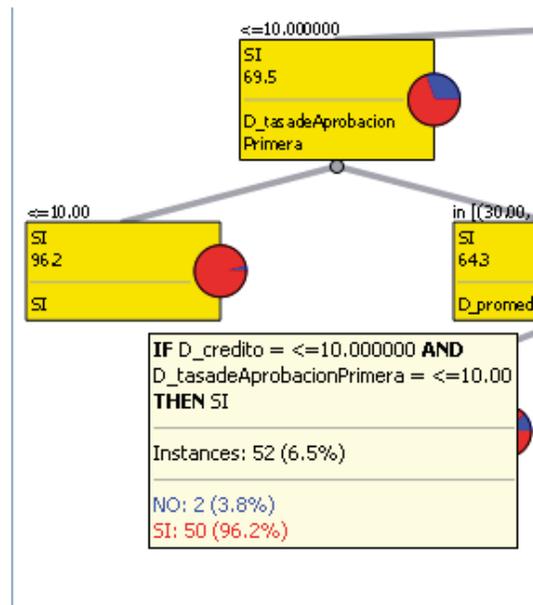


Figura 5.40: Poda clasificacion alumnos desertores y la tasa de avance

Otra regla importante se encuentra representada en la Figura 5.41, donde los alumnos desertores se encuentran representados por aquellos que tienen una tasa de avance según año menor al 10%, al igual que la tasa de aprobación. Junto a lo anterior, se ratifica que los alumnos desertores principalmente se encuentran representados por aquellos que tienen un bajo porcentaje de crédito, lo que representa un soporte de un 6,8% de la carrera con una confianza de un 96.3%.

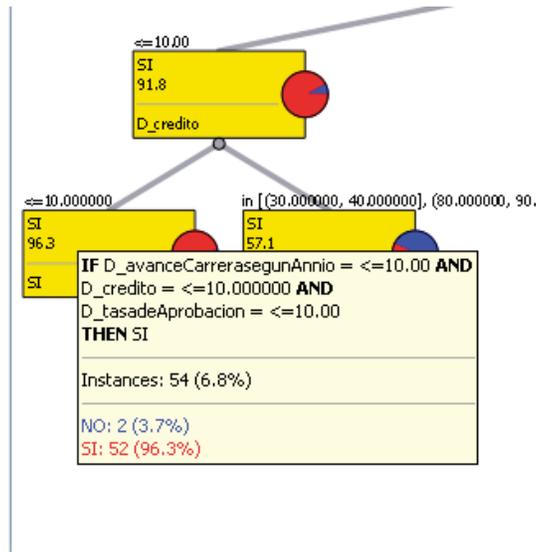


Figura 5.41: Poda clasificacion alumnos desertores y la tasa de avance

Por otra parte se visualiza también que aprobar las asignaturas juega un rol fundamental al momento de decidir desertar, los alumnos que tiene tasas de aprobación por primera menor al 10% desertan en al menos 59 instancias que representan el 7.4% de los casos con una confianza de un 91%. Ver Figura 5.42

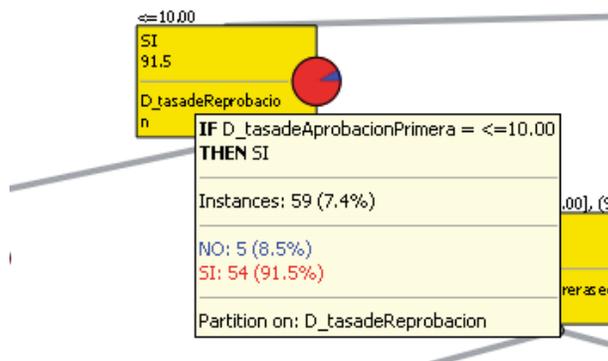


Figura 5.42: Poda clasificacion alumnos desertores y la tasa de aprobación

Junto a la baja tasa de avance ya sea por año o general en la carrera se identifica que los alumnos desertores de la carrera de Ingeniería en Construcción es muy importante el promedio de notas, dado que los alumnos con un promedio de notas menor a un 4.0 desentan de la carrera en un 6.9% de los casos con una confianza de un 87.3%, como se puede ver en la Figura 5.43

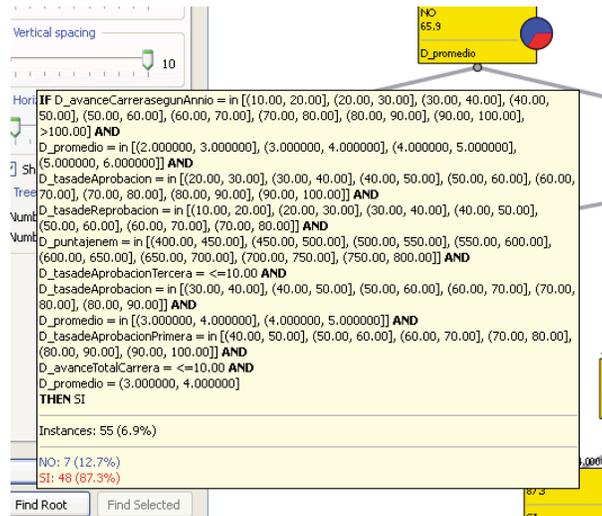


Figura 5.43: Poda clasificacion alumnos desertores y la tasa de avance

Estas nuevas reglas identifican la importancia que tienen para los alumnos de la carrera de construcción las variables de:

- Tasa de aprobación por primera menor al 10 %
- Tasa de aprobación menor al 10 %
- Tasa de avance total carrera menor al 50 %
- Crédito asignado menor al 50 %
- promedio de notas menor a un 4.0

5.3. Análisis de la Deserción en otras carreras

Al evaluar las reglas obtenidas en la primera evaluación, se visualizan varias reglas que tienen un bajo soporte, en muchos de los casos menor al 1 % del total de datos. En base a lo anterior, se ha realizado un nuevo análisis de los datos realizando una poda de los arboles obtenidos, para permitir reglas que contengan un mayor número de casos, sacrificando hasta un cierto nivel de confianza. Lo anterior, permitiendo identificar mejores soportes para la clasificación. Al realizar esta tarea se usaron reglas de asociación y optimización de los atributos y junto a ellos se han modificado las reglas de cantidad de instancias por nodo y la poda con m-estimado. Para la evaluación se ha utilizado orange canvas, debido a su simpleza y facilidad, lo que permite por medio de la interfaz presentada en la Figura 5.44 modificar y adaptar los datos para obtener los resultados esperados.



Figura 5.44: Ajuste atributos clasificación y poda

En el análisis por carrera se han obtenido resultados importantes dado que se ha evaluado la data completa nuevamente, dentro de los resultados más importantes se encuentran, lo presentados en las secciones a continuación.

5.3.1. Carrera de Arquitectura

Los resultados iniciales que se encuentra en el Anexo 1, presentan reglas de bajo soporte, que hacen necesario realizar la poda de los árboles obtenidos en la clasificación, con los nuevos resultados en la carrera se determinan algunas reglas que aportan conocimiento a la evaluación.

Donde el 1.4% de los alumnos tiene un promedio menor a un 5.0 y un avance total de la carrera pueden ser causal de deserción de la carrera. Ver Figura 5.45

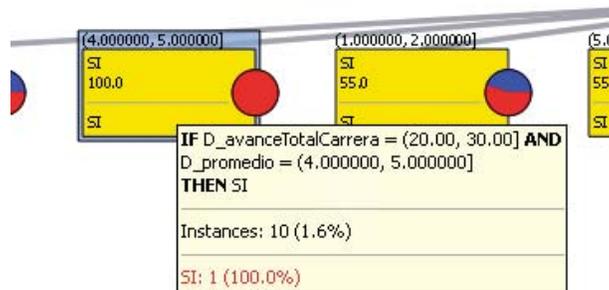


Figura 5.45: Poda clasificacion alumnos desertores y la tasa de avance total

El atributo de avance total carrera toma un gran relevancia en este caso dado que se encuentra en una regla con 39.7% de soporte y un 94% de confianza que identifica que los alumnos desertan antes del 10% de avance de su carrera. Ver Figura 5.46

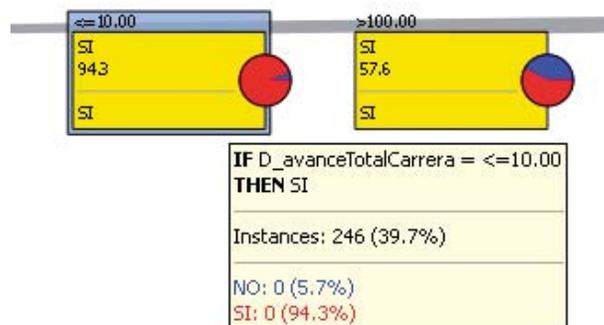


Figura 5.46: Poda clasificacion alumnos desertores y la tasa de avance total carrera

Junto a lo anterior se identifica con un soporte de 38 instancias que corresponden al 6.1% de los casos que si el avance según año es menor a un 10% pese a contar con avance según semestre mayores los alumnos desertan de la carrera con una confianza de un 97%, al tener una asignacion de créditos menor al 10%. Ver Figura 5.47

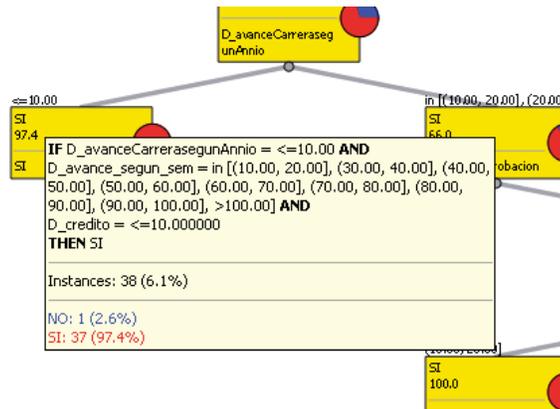


Figura 5.47: Poda clasificacion alumnos desertores y la tasa de avance segun año menor al 10%

Otra mezcla interesante de atributos corresponde a la regla presentada en la Figura 5.48 con un 4% del soporte y que tiene un 100% de confianza, en que los alumno con un crédito menor a un 10% y con avance según semestre entre un 40% y un 50% desertan sin importar la condición donde en que se encuentre, con respecto al avance según año.

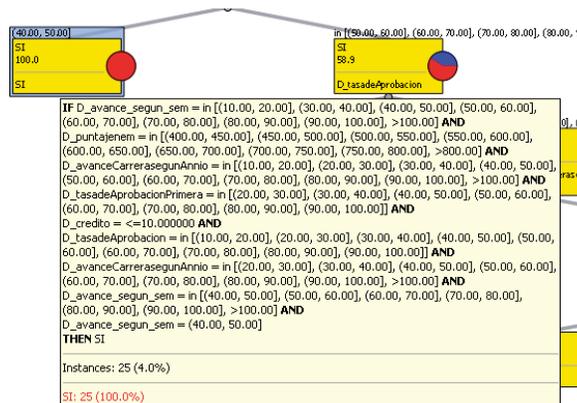


Figura 5.48: Poda clasificacion alumnos desertores y asignacion de crédito menor al 10%

Junto a lo anterior, se identifica que la tasa de avance del riesgo de deserción de los alumnos de arquitectura es mayor que en otras carreras dado que hay un alto porcentaje que se encuentra representado por aquellos alumnos con tasa de avance total menor a un 50%, lo anterior se representa con un soporte del 18% de los casos que equivalen a 116 instancias con una confianza de un 94.8%, que los alumnos desertan es esta condición. Ver Figura 5.49

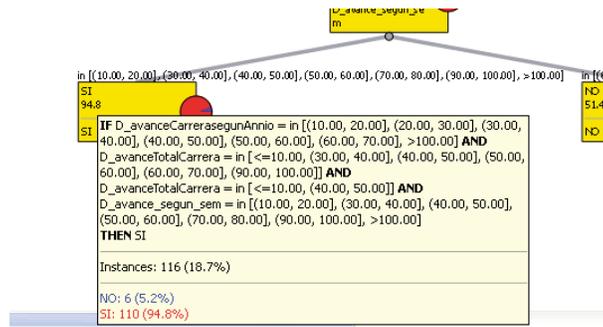


Figura 5.49: Poda clasificacion alumnos desertores y que tienen asignaturas por segunda

Un regla importante y que cuenta con un 14.9% del soporte y una confianza de un 85.9% fundamenta que el atributo de la tasa de avance según año no impacta directamente en la decisión de la deserción de los alumnos de la carrera al igual que el promedio dado que los alumnos sin importar que tengan una buena tasa de avance según año y buenos promedios de todas formas llegan a desertar en un gran número de casos. Ver Figura 5.50

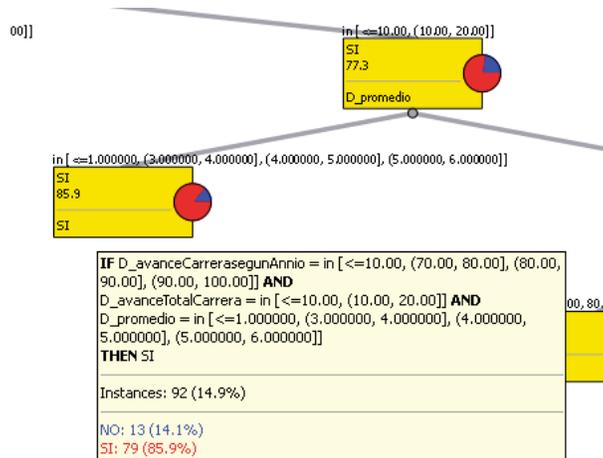


Figura 5.50: Poda clasificacion alumnos desertores y tasa de avance del alumno

Junto a lo anterior se identifica la relevancia que tiene la tasa de avance por primera dado que con un soporte de un 14.7% de los casos que equivalen a 91 instancias los alumnos que tienen una tasa menor al 10% de la tasa de aprobación por primera desertan, como se puede ver en la Figura 5.51

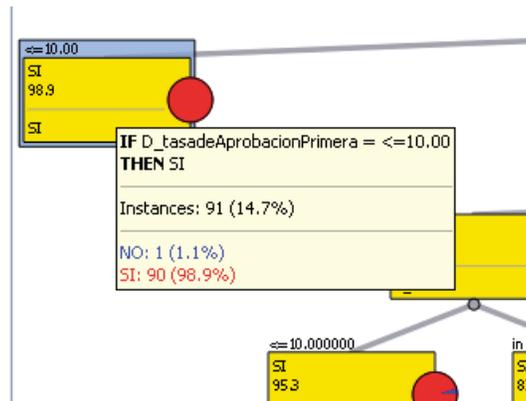


Figura 5.51: Poda clasificación alumnos desertores y tasa de aprobación por primera

En base a esta nueva evaluación se identifica que la carrera tiene como atributos más relevantes a los siguientes, donde el más relevante es la tasa de avance total de la carrera.

- Tasa de avance total de la carrera menos a un 50 %
- Tasa de aprobación por primera menor al 10 %
- Tasa de aprobación menor al al 10 %
- crédito menor al 50 %

5.3.2. Carrera de Auditoria (Diurno)

Al evaluar la clasificación en la carrera de auditoría se identifica que los atributos más relevantes se repiten como se puede ver en la Figura 5.52, todo alumno con créditos menor al 10 %, con un avance total de la carrera menor al 10 % con puntajes bajo 700 puntos, desertan de la carrera con un soporte de la regla de un 5.5% de los casos y una confianza de un 82.6%. Obteniendo con esta reevaluacion reglas de mejor soporte como las presentadas obtenidas en el análisis inicial presentado en el anexo 2.

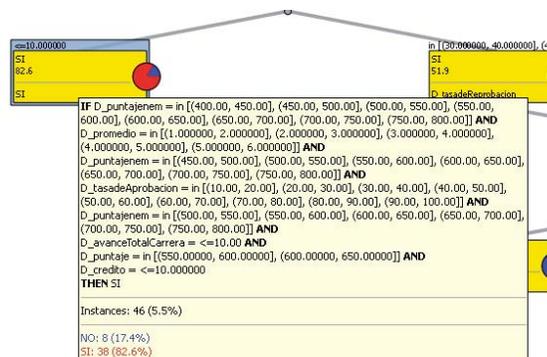


Figura 5.52: Alumnos desertores de carrera c y el porcentaje de asignación de créditos

Junto a lo anterior se identifica que los alumnos con una tasa de avance total menor al 10% desertan con un soporte de un 3% de los alumnos y una confianza de un 96%, ver Figura 5.53. La imagen explica y ratifica la importancia que tiene el avance general del alumno, al igual que en las carreras anteriores.

Por otra parte, se identifica que el puntaje si toma una importancia menor en la carrera dado que los alumnos con puntajes menores a 550 puntos tienen una mayor probabilidad de deserción.

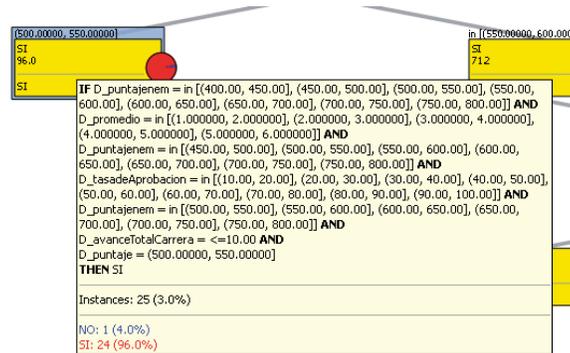


Figura 5.53: Alumnos desertores de carrera c y el porcentaje de asignación de créditos

Otra regla importante se identifica en la Figura 5.54, donde se puede ver que los alumnos que tienen tasas de aprobación menor al 10% desertan con un soporte de 1.9% y una confianza de un 100%, Demostrando que es necesario evaluar este atributo en los alumnos.

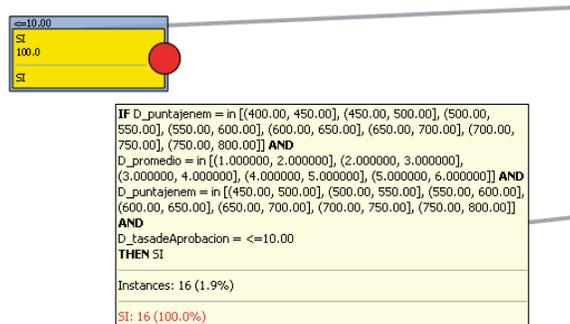


Figura 5.54: Poda clasificacion alumnos desertores y tasa de Aprobación del alumno

Al reevaluar la importancia de las tasas se identifica que la deserción de los alumnos principalmente se encuentra representada por la tasa de avance de los alumnos, donde alumnos con un avance menor al 10% desertan de forma general, como se puede ver en las Figuras 5.55, 5.56. En ambos casos los alumnos desertan en con un gran soporte y una alta confianza, con la diferencia que en alumnos con tasa de avance semestral menor al 10% desertan con una confianza de un 100%, al contrario de los otros porcentajes de avance según semestre donde también desertan pero con una confianza de un 82%.

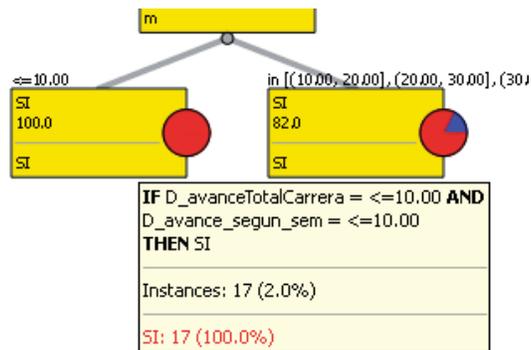


Figura 5.55: Poda clasificacion alumnos desertores y tasa de Avance del alumno

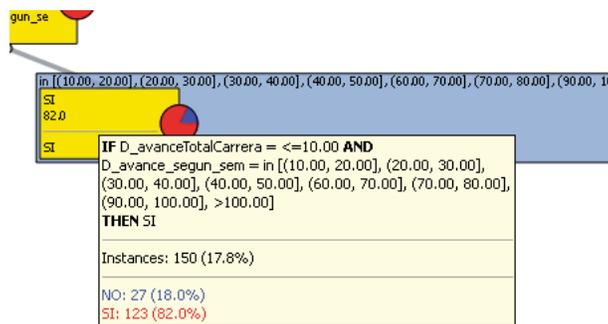


Figura 5.56: Poda clasificación alumnos desertores y tasa de Avance según semestre del alumno

Otra evidencia se identifica que los alumnos con una tasa de aprobación por primera menor al 10% desertan en un 3,3% de los casos con un 100% de confianza, como se puede ver en la Figura 5.57

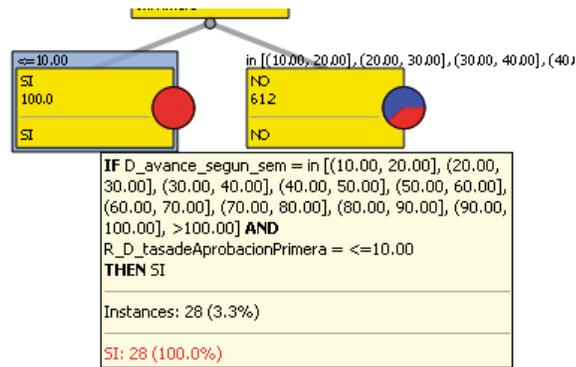


Figura 5.57: Poda clasificación alumnos desertores y tasa de Aprobación por primera del alumno

Al igual que en todas las carreras evaluadas toma importancia en el análisis en especial el atributo de avance total carrera donde el riesgo más alto se encuentra en tasas menores al 10 % en conjunto con los atributos.

- Tasa de avance total carrera menor al 40 %
- Tasa de aprobación menor al 50 %
- Crédito menor al 10 %
- Tasa de avance según semestre menor al 10 %
- Puntaje menor a 550 ptos.

Estos atributos también describen la deserción en la carrera de auditoría.

5.3.3. Carrera de Enfermería

Debido a la baja cantidad de eliminado en la carrera es complejo obtener reglas que identifiquen la deserción en la carrera, según los resultados obtenidos en la evaluación inicial que se presentan en el Anexo 3.

Lo anterior, dado que principalmente la carrera clasifica a sus alumnos con una alta probabilidad de finalizar sus estudios, pero dentro de estos casos se encuentran reglas que ratifican los atributos importantes ya descritos.

En la regla presentada Figura 5.58 se identifica que los alumnos con créditos menores a un 10 % y tasa de avance total en la carrera menor a un 20 % desertan a la carrera en un 2.9 % de los casos totales con una confianza de un 76.5 %.

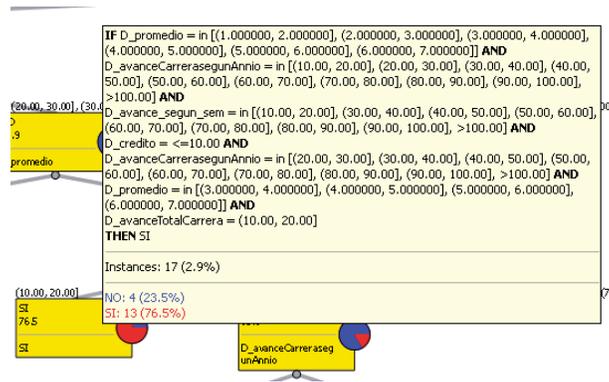


Figura 5.58: Reevaluación Enfermería con nuevas reglas en alumnos con créditos

Otra regla interesante es la que valida en la carrera la importancia que tiene el avance de los alumnos según su año de ingreso donde se presenta con una confianza de un 100 % y un soporte de un 3.5 % que los alumnos con tasas de avance según año menores al 10 % desertan. Ver Figura 5.59

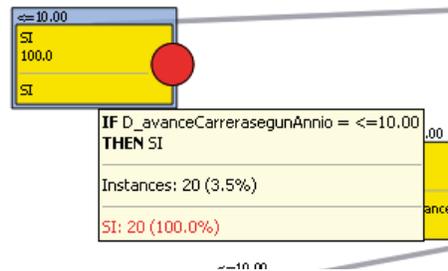


Figura 5.59: Reevaluación Enfermería con nuevas reglas en alumnos según año de ingreso

La regla presentada en la Figura 5.60, muestra la importancia que tiene el avance de general de los alumnos en la carrera al momento de desertar, siendo el atributo más importante al momento de tomar la decisión de desertar.

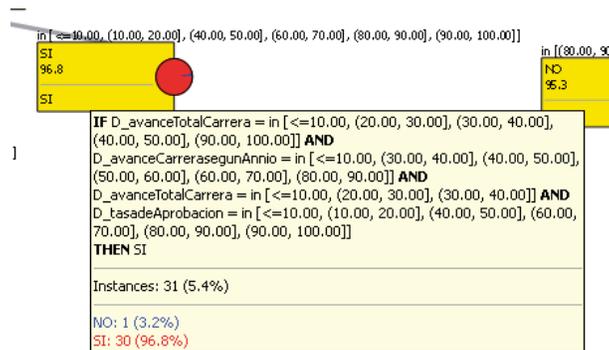


Figura 5.60: Reevaluación Enfermería con nuevas reglas en alumnos segun avance total.

Al evaluar las distintas reglas que se pueden obtener, se puede identificar una regla que en un menor grado de importancia que el avance total, pero no menor dado que puede llegar a describir por sí sola la deserción de los alumnos, se identifica que los alumnos con una tasa de aprobación menor al 10% desertan de la carrera con un soporte de un 2.6% y una confianza del un 100%. Lo anterior, implica que es necesario evaluar este atributo en los alumnos para identificar la deserción. . Ver Figura 5.61

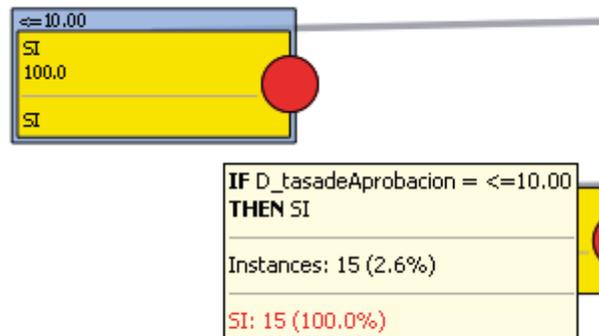


Figura 5.61: Reevaluación Enfermería con nuevas reglas en alumnos según Tasa de Aprobación.

Reglas mayores como la presentada en la Figura 5.62, ratifica los atributos más importantes que permiten describir la deserción de los alumnos se encuentran definidos por el avance total de la carrera, avance según año y el crédito asignado de los alumnos. Lo anterior, describe en la Figura con un soporte de un 1.9% de los casos y un 90.9% de confianza que los alumnos desertan al tener tasas de avance total menor al 20% y créditos menores al 10%.

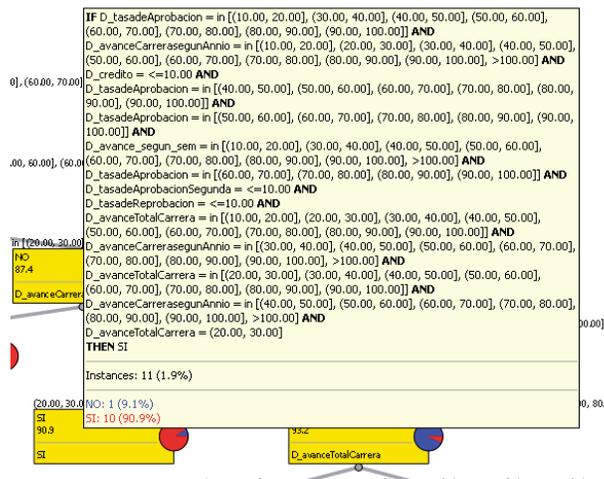


Figura 5.62: Reevaluación Enfermería con nuevas reglas en alumnos según Tasa de Aprobación.

Por último, en la regla presentada en la Figura 5.63, identifica la importancia que tiene la tasa de avance del alumno según año de ingreso, la que permite describir con un 100% de confianza y un soporte de un 3.5% de los casos totales, que los alumnos desiertan al tener avances menores al 10%.

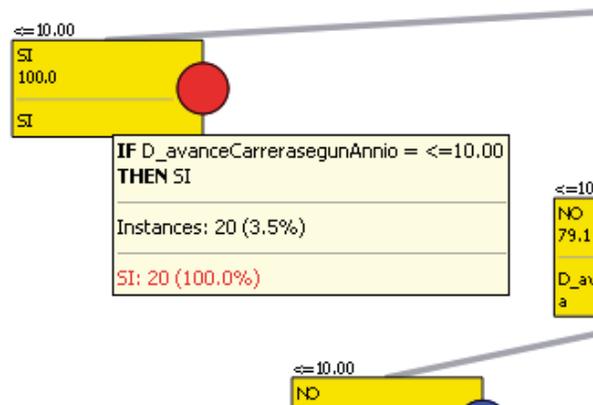


Figura 5.63: Reevaluación Enfermería con nuevas reglas en alumnos con Tasa avance según año.

Aunque el crédito y las becas juegan un rol en esta carrera no es tan relevante como en las carreras anteriormente evaluadas en si los alumnos desertores principalmente se encuentran descritos en la carrera por:

- Tasa de avance total carrera
- Tasa de aprobación por primera
- Crédito
- Tasa de avance según año

En forma general los atributos más relevantes que describen la deserción se encuentran definidos por:

- Tasa de avance total carrera menor a 10 %
- Tasa de avance según año menor a 20 %
- Tasa de aprobación por primera menor a un 10 %
- Créditos asignados menor al 10 %
- Tasa de reprobación mayor al 20 %
- Promedio de notas menor a un 4.0 %

5.4. Conclusiones

Según el análisis de datos realizado, por medio de técnicas de minería de datos, se puede inferir que no todos los atributos que describen a un alumno son importantes a la hora de analizar su probabilidad de deserción, junto a lo anterior también se extrae como conocimiento que las carreras describen su deserción en base a distintas condiciones de sus alumnos, donde no existe un único grupo de atributos que describa la deserción, pero si existen atributos que son transversales y que sería muy importante para las autoridades evaluar para evitar que los alumnos deserten de sus carreras, tal como se puede ver en la Figura 5.64.

Además se obtienen respuestas iniciales a las preguntas de las autoridades con respecto a la deserción de los estudiantes que se han descrito en la definición del problema, como:

- ¿es factible identificar el riesgo de la deserción de un estudiante previamente?

Si es factible, al contar con la identificación de los patrones que definen esta, se puede implementar un sistema de alerta que permita al los directores tomar medidas preventivas con estos alumnos.

- ¿existe una correlación entre los factores de ingreso y la continuidad de los alumnos en el programa?

Como se puede observar no existe implicancia de los factores de ingreso de los alumnos frente a la deserción. lo anterior dado que los factores que la definen estan centrados en los procesos de avance y de aprobación de sus asignaturas y no en su ingreso, donde en forma general se encuentran equilibrados en segmentos muy similares de puntaje, procedencia, lugar y preferencia.

- ¿existen asignaturas que destaquen la factibilidad de deserción de los alumnos en una carrera?

Aunque no se llego a algo definitivo en este estudio si se identifican asignaturas que pueden llegar incidir en la deserción de los estudiantes, pero para validar las causas es necesario contar con un histórico mayor de antecedentes las encuestas y las evaluaciones docentes.

- ¿realmente los alumnos con mal comportamiento de pago o aquellos que tienen muchos beneficios están más expuestos a la deserción en la educación superior?

El caso del comportamiento de pago no es un reflejo en la deserción dado que en general la mayoría de los alumnos tiene un mal comportamiento de pago en que la menos en mas de una oportunidad se

han retrasado en ellos, pero con respecto a los beneficios se encontró que el caso es inverso el problema se incrementa en alumnos que tienen asignados créditos muy bajos en porcentaje o simplemente no tienen beneficios económicos.

En general la deserción tiene causas particulares en cada carrera incidiendo directamente de forma transversal, el avance académico y el porcentaje de crédito asignado.

| | Ingeniería en Construcción | Arquitectura | Auditoría Diurna | Enfermería |
|-------------------------|-------------------------------|--------------|---------------------|------------|
| ingreso | - | - | - | - |
| per_acade | - | - | - | - |
| semestre | - | - | - | - |
| tiempo | - | - | - | - |
| via_ingreso | - | - | - | - |
| estado | - | - | - | - |
| tasadeAprobacion | 10 | 10 | 40 | 10 |
| tasadeReprobacion | - | - | - | - |
| tasadeAprobacionPrimera | 10 | 10 | - | - |
| tasadeAprobacionSegunda | - | - | - | - |
| tasadeAprobacionTercera | - | - | - | - |
| avanceTotalCarrera | 50 | 10 | 50 | 50 |
| avanceCarrerasegunAnio | - | - | - | 30 |
| promedio | 3 | - | - | - |
| cantidad_aprobadas | - | - | - | - |
| cantidad_reprobadas | - | - | - | - |
| primera | - | - | - | - |
| segunda | - | - | - | - |
| tercera | - | - | - | - |
| cuarta | - | - | - | - |
| mas_de_cuatro | - | - | - | - |
| minimo | - | - | - | - |
| maxCursa | - | - | - | - |
| segundasCursa | - | - | - | - |
| terceraCursa | - | - | - | - |
| beca | - | - | - | - |
| credito | 10 | 50 | 50 | 20 |
| atrasos | - | - | - | - |
| puntajenem | - | - | - | - |
| puntaje | - | - | - | - |
| lugar | - | - | - | - |
| preferencia | - | - | - | - |
| tipocolegio | - | - | - | - |
| grupo_dependencia | - | - | - | - |
| estadosolucion | - | - | - | - |
| ciudad | - | - | - | - |
| region | - | - | - | - |
| semestre | - | - | - | - |
| cant_semestre | - | - | - | - |
| avance_segun_sem | - | - | 40 | - |

Figura 5.64: Influencia de los atributos en la deserción.

Capítulo 6

Propuesta de Indicador de Riesgo de Deserción

Al realizar el análisis con minería de datos se han identificado una serie de atributos que permiten describir a los alumnos desertores, pero frente a este análisis nace una nueva pregunta ¿cuál es el impacto o la relevancia que tiene cada uno de los atributos?, y por otra parte ¿existe factibilidad que crear un indicador que permita describir a los alumnos desertores?, de tal forma de asignar un grado de criticidad ante la deserción. Para lo anterior dada su simpleza y la flexibilidad que otorga se ha decidido realizar un primer alcance del indicador de riesgo utilizando AHP (*Analytic Hierarchy Process*) o Proceso de Análisis Jerárquico.

6.1. Carrera de Ingeniería en Construcción

La evaluación de la carrera de Ingeniería en construcción se encuentra definida por:

1. Avance total carrera menor al 50 %
2. Tasa de aprobación por primera menor al 10 %
3. Tasa de aprobación menor al 10 %
4. Promedio menor o igual a un 3.0
5. Asignación de crédito menor al 50 %

En base a estos atributos se obtiene la matriz de pesos que permite obtener los pesos de los atributos.

| | avanceTotalCarrera <50 | tasadeAprobacionPrimera <10 | tasadeAprobacion <10 | promedio <3.0 | credito <50 | suma |
|-----------------------------|------------------------|-----------------------------|----------------------|---------------|-------------|-------|
| avanceTotalCarrera <50 | 1,00 | 0,14 | 0,33 | 0,11 | 0,33 | 1,92 |
| tasadeAprobacionPrimera <10 | 7,00 | 1,00 | 7,00 | 3,00 | 7,00 | 25,00 |
| tasadeAprobacion <10 | 3,00 | 0,14 | 1,00 | 7,00 | 0,33 | 11,48 |
| promedio <3.0 | 9,00 | 0,33 | 0,14 | 1,00 | 7,00 | 17,48 |
| credito <50 | 3,00 | 0,14 | 3,00 | 0,14 | 1,00 | 7,29 |

Figura 6.1: matriz de pesos atributos Construcción

Con esto pesos se obtiene valores de evaluación promedios, presentados en la tabla a continuación

| | avanceTotalCarrera <50 | tasadeAprobacionPrimera <10 | tasadeAprobacion <10 | promedio <3.0 | credito <50 | Evaluación |
|-----------------------------|------------------------|-----------------------------|----------------------|---------------|-------------|------------|
| avanceTotalCarrera <50 | 0,52 | 0,07 | 0,17 | 0,06 | 0,17 | 1,00 |
| tasadeAprobacionPrimera <10 | 0,28 | 0,04 | 0,28 | 0,12 | 0,28 | 1,00 |
| tasadeAprobacion <10 | 0,26 | 0,01 | 0,09 | 0,61 | 0,03 | 1,00 |
| promedio <3.0 | 0,51 | 0,02 | 0,01 | 0,06 | 0,40 | 1,00 |
| credito <50 | 0,41 | 0,02 | 0,41 | 0,02 | 0,14 | 1,00 |
| promedio | 0,40 | 0,03 | 0,19 | 0,17 | 0,20 | |

Figura 6.2: matriz de ajuste de pesos atributos Construcción y promedios obtenidos

Con estos resultados se aplica la estructura algorítmica descrita anteriormente que permite identificar a los alumnos desertores de la carrera de Ingeniería en construcción, donde se encuentra en que no todas las carreras tienen los mismos atributos relevantes y en el mismo rango de valores.

con los resultados se ha determinado que si la suma es mayor a 0,5 describe a un alumno desertor de la carrera, y que mientras mayor es el resultado de la misma, describe a un alumno desertor.

Frente a los resultados obtenidos, con un soporte de un 100% de los registros, que corresponden a 561 registros de los que 127 son alumnos desertores y 434 no, con estos valores aplicando el polinomio se obtienen los siguientes resultados.

1. La efectividad del polinomio se encuentra en un buen rango permitiendo una efectividad de un 90.1% de confianza determina clasifica correctamente a un alumno desertor o no.

2. El algoritmo identifica a un alumno desertor con una confianza de un 90.5%, con un soporte de un 100% de los alumnos desertores.

3. Con 89.8% de confianza el algoritmo clasifica correctamente a un alumno no desertor, lo que puede presupone un valor bajo en este caso sería preferible identificar a un alumno con un riesgo mayor como desertor que el caso contrario.

Básicamente el algoritmo permite obtener en un buen número de casos a los alumnos desertores.

6.2. Carrera de Arquitectura

En base a la evaluación de la minería de datos se han obtenido los atributos de:

1. Tasa de avance total menor al 10 %
2. Tasa de aprobación por primera menor al 10 %
3. Tasa de aprobación total menor al 10 %
4. Crédito asignado menor al 50 %

En base a esto se ha obtenido la matriz de preferencias que se presenta a continuación:

| | avance general menor al 10% | Tasa de aprobación por primera | Tasa de aprobación | crédito asignado | suma |
|--|-----------------------------|--------------------------------|--------------------|------------------|-------|
| avance general menor al 10% | 1.00 | 0.14 | 0.14 | 0.33 | 1.62 |
| Tasa de aprobación por primera menor 10% | 7.00 | 1.00 | 0.33 | 0.14 | 8.48 |
| Tasa de aprobación menor 10% | 3.00 | 3.00 | 1.00 | 0.20 | 7.20 |
| crédito asignado menor 50% | 3.00 | 7.00 | 5.00 | 1.00 | 16.00 |

Figura 6.3: matriz de pesos atributos Arquitectura

Una vez obtenido de los valores globales se define la matriz de pesos y el promedio de estos, con el cual se ha realizado la evaluación global de la deserción en la carrera.

| | avance general menor al 50% | Tasa de aprobación por primera | Tasa de aprobación | crédito asignado | suma |
|--|-----------------------------|--------------------------------|--------------------|------------------|-------|
| avance general menor al 10% | 0.62 | 0.09 | 0.09 | 0.21 | 11.33 |
| Tasa de aprobación por primera menor 10% | 0.83 | 0.12 | 0.04 | 0.02 | 17.14 |
| Tasa de aprobación menor 10% | 0.42 | 0.42 | 0.14 | 0.03 | 4.48 |
| crédito asignado menor 50% | 0.19 | 0.44 | 0.31 | 0.06 | 4.44 |
| avance general menor al 10% | 0.51 | 0.27 | 0.14 | 0.08 | 1 |

Figura 6.4: matriz de ajuste de pesos atributos Arquitectura y promedios obtenidos

En base a los pesos promedios se obtiene el siguiente algoritmo de evaluación

```

CríticidadDesercion(avanceTotal float , aprobacionPrimera float, aprobacionTotal float,
crédito float){
    Suma =0;
    avanceTotal < 50? Suma+=0.51:suma;
    aprobacionPrimera < 10? Suma+=0.27: suma;
    aprobacionTotal<10?Suma+=0.14:suma;
    crédito<50?Suma +=0.8:suma;
    return Suma;
}

```

Figura 6.5: Algoritmo aplicado a la carrera

Al aplicar el algoritmo se ha determinado que si suma es mayor a 0.5 el alumno deserta en una alta probabilidad, en caso contrario el alumno no deserta.

Frente a los resultados obtenidos se puede identificar que con un soporte total de 598 casos pertenecientes a la carrera de los cuales 338 desertan y 260 no desertan se obtiene se obtienen los siguientes resultados:

1.- 22 de los 598 alumnos se encuentran mal clasificados, por lo que el polinomio en forma general tiene una confianza de un 96.3 %.

2- Del total de desertores 8 de 338 que desertan fueron clasificados como no desertores, por lo que el porcentaje de desertores correctamente identificados es de 97.6 %

Estos valores en si son muy alentadores y se podría en una primera instancia indicar que es factible contar con instrumentos que permitan validar la probabilidad de deserción de los alumnos de esta carrera.

6.3. Carrera de Auditoria (Diurno)

La evaluación de la carrera de Ingeniería en Construcción se encuentra definida por:

1. Tasa de aprobación menor al 40
2. Avance total carrera menor al 50
3. Crédito asignado menor al 50
4. Avance del alumno según semestre menor al 40

En base a estos atributos se realiza una evaluación de los pesos de las variables que permiten identificar la siguiente matriz.

| | tasadeAprobacion <50 | avanceTotalCarrera <50 | credito < 50 | avancesegunssemetre <50 | suma |
|-------------------------|----------------------|------------------------|--------------|-------------------------|-------|
| tasadeAprobacion <50 | 1,00 | 3,00 | 5,00 | 0,20 | 9,20 |
| avanceTotalCarrera <50 | 0,33 | 1,00 | 0,20 | 0,33 | 1,87 |
| credito < 50 | 0,20 | 5,00 | 1,00 | 3,00 | 9,20 |
| avancesegunssemetre <50 | 5,00 | 0,33 | 5,00 | 1,00 | 11,33 |

Figura 6.6: matriz de pesos atributos Auditoría

Con los valores de los pesos se obtiene un promedio de ajustado presentado a Auditoría.

| | tasadeAprobacion <50 | avanceTotalCarrera <50 | credito < 50 | avancesegunssemetre <50 | suma |
|-------------------------|----------------------|------------------------|--------------|-------------------------|-------------|
| tasadeAprobacion <50 | 0,11 | 0,33 | 0,54 | 0,02 | 0,98 |
| avanceTotalCarrera <50 | 0,18 | 0,54 | 0,11 | 0,18 | 0,82 |
| credito < 50 | 0,02 | 0,54 | 0,11 | 0,33 | 0,67 |
| avancesegunssemetre <50 | 0,44 | 0,03 | 0,44 | 0,09 | 0,91 |
| Promedio | 0,19 | 0,36 | 0,30 | 0,15 | 1,00 |

Figura 6.7: matriz de ajuste de pesos atributos Auditoría y promedios obtenidos

Con estos valores se aplica el algoritmo que identifica la suma de los pesos en la totalidad de los alumnos de la carrera, evaluando con un soporte de un 100% de los datos, que corresponden a 819 registros que corresponden a 480 alumnos que no desertan y 339 que si desertan, donde se obtienen los siguientes resultados.

1. Se identifica que el polinomio tiene una menor efectividad que en los casos anteriores con una confianza de un 73.4%, el problema principal de la carrera, y que puede afectar esta asertividad radica en el gran número de alumnos que realizan cambio de jornada, por los que sería interesante evaluar a estos alumnos en conjunto con las estructuras curriculares paralelas que tienen.

2. Pese a todo el polinomio permite identificar a los desertores con una confianza de un 82.9% de confianza.

3. Además se obtiene un buen número de desertores para el caso de la evaluación de los que no desertan su resultado no es como se espera dado que clasifica a los no desertores con un 66.7% de confianza.

Para el caso de la carrera es un buen predictor de alumnos desertores pero se presenta la problemática de aquellos que no desertan a la carrera sino que cambian su jornada lo que sería interesante evaluar por separado.

6.4. Carrera de Enfermería

Para la carrera se identifican atributos similares a los casos anteriores, los cuales son:

1. Avance Total carrera menor a 50 %
2. Avance Carrera según año menor a 30 %
3. Crédito asignado menor o igual al 20 %
4. Tasa de aprobación menor o igual al 10 %

En base a estos atributos se realiza una evaluación de los pesos de las variables que permiten identificar la siguiente matriz.

| | avanceTotalCarrera | avanceCarrerasegunAnno | credito | tasadeAprobacion | suma |
|------------------------|--------------------|------------------------|---------|------------------|-------|
| avanceTotalCarrera | 1,00 | 0,33 | 0,33 | 0,20 | 1,87 |
| avanceCarrerasegunAnno | 3,00 | 1,00 | 0,14 | 0,20 | 4,34 |
| credito | 3,00 | 7,00 | 1,00 | 0,33 | 11,33 |
| tasadeAprobacion | 5,00 | 5,00 | 3,00 | 1,00 | 14,00 |

Figura 6.8: matriz de pesos atributos Enfermería

Con los valores de los pesos se obtiene un promedio de ajustado presentado a continuación.

| | avanceTotalCarrera | avanceCarrerasegunAnno | credito | tasadeAprobacion | suma |
|------------------------|--------------------|------------------------|-------------|------------------|-------------|
| avanceTotalCarrera | 0,54 | 0,18 | 0,18 | 0,11 | 1,00 |
| avanceCarrerasegunAnno | 0,69 | 0,23 | 0,03 | 0,05 | 1,00 |
| credito | 0,26 | 0,62 | 0,09 | 0,03 | 1,00 |
| tasadeAprobacion | 0,36 | 0,36 | 0,21 | 0,07 | 1,00 |
| promedio | 0,46 | 0,35 | 0,13 | 0,06 | 1,00 |

Figura 6.9: matriz de ajuste de pesos atributos Enfermería y promedios obtenidos

Con los valores obtenidos y aplicados a la carrera se puede identificar la deserción de los alumnos con un 86.7% de confianza en un soporte de un 100% de los alumnos, En realidad es un número importante de identificación debido a que la deserción de los alumnos de la carrera en total corresponde al 18.9%, con estos resultados generales se puede indicar.

1. El polinomio permite identificar correctamente con una confianza de un 86% a los alumnos desertores con un soporte de un 100% de los alumnos.

2. Por otra parte muestra una gran ventaja dado que se comporta de la misma forma al identificar a un alumno no desertor con una confianza de un 86.9% y un soporte de un 100%.

Capítulo 7

Conclusiones

Hoy en día la gran mayoría de las instituciones cuenta con una cantidad importante de información, que puede permitir obtener conocimiento de interés para el desarrollo de la empresa o institución, para esto existe distintas vías de iniciar el proceso, pero todos llegan a un mismo fin que es la búsqueda de conocimiento.

Para esto, en este documento se han presentado una serie de herramientas y alternativas como son los repositorios de datos y la inteligencia de negocios, pero solamente la minería de datos podría permitir encontrar conocimiento oculto en los grandes volúmenes de datos. Existen dos grandes tipo de conocimiento dentro de la minería de datos, que es descriptivo y predictivo, que permiten describir por métodos supervisados o no supervisados la información, lo que apoya la toma de decisiones.

Las universidades no están lejos de esto de esta realidad, la gran mayoría de ellas ya cuenta con grandes repositorios de información, que pueden ser utilizados para la búsqueda de conocimiento que permita el mejoramiento continuo y apoyar a los desafíos que imponen los procesos de acreditación, donde los indicadores juegan un rol primordial, como es la deserción de la educación superior.

En este caso, la deserción universitaria es un tema relevante en los análisis de la educación superior, debido a las consecuencias personales y económicas que ocurren a distintos niveles. Existen algunos estudios que hacen referencia a la minería de datos Y otros hacen recomendaciones basados en pseudo experimentos, pero ninguno ha analizado de forma transversal y utilizando técnicas de minería de datos el problema de la deserción.

El análisis exploratorio de los datos permitió obtener resultados previos ante las dudas presentadas por las autoridades, dentro de los resultados principales obtenidos se encuentra el hecho que los alumnos desertores postulas en su mayoría con buenos puntajes y en las primeras opciones a la universidad.

Junto a lo anterior, también se logro determinar que los antecedentes económicos y de notas juegan un rol principal ante la descripción de los desertores a la universidad.

Hay que destacar que el análisis exploratorio es tan sólo un análisis general y requiere mayor analisis en profundidad, permite en gran medida obtener la vista minable y los atributos principales que permitirán aplicar los algoritmos descriptivos y los posibles algoritmos predictivos que permitan obtener las claves para determinar la probabilidad que un alumno sea eliminado académicamente o un desertor.

Al realizar las primeras evaluaciones, se identifica que el conocimiento obtenido con los datos extraídos no es un gran aporte, se encuentran relaciones un poco obvias entregado un conocimiento menor, por lo que

es necesario realizar relaciones que puedan otorgar algún tipo de conocimiento, para esto se crearon distintas tasas, entre otras, la tasa de avance y tasa de aprobación, que permiten obtener conocimiento nuevo con los datos de los alumnos.

Dentro de las preguntas a responder se encuentra, si existen atributos que identifiquen la deserción, en este análisis ya se identifican atributos que permiten identificar a los alumnos desertores, destacando el crédito asignado, la tasa total de avance de los alumnos y la tasa de aprobación, por otra parte pierde relevancia atributos de ingreso como puntaje, preferencia y otros que se podrían pensar que si impactan directamente en la permanencia de los alumnos.

Uno de los problemas de los resultados iniciales, pese a la calidad de las reglas obtenidas en base a su confianza se encuentra en el bajo soporte obtenido, por parte de los resultados. Debido a lo anterior, ha sido necesario realizar nuevas mezclas que permitan ampliar el soporte de los resultados aplicando recursividad y poda en los arboles.

Con estos nuevos resultados obtenidos, se han identificado reglas con buen soporte que han permitido identificar cada uno de los atributos que impactan directamente en la deserción de los alumnos, donde se identifica que esta ocurre principalmente en alumnos con tasas de avance menor al 30 %, donde su tasa de aprobación es baja y lo más importante es el bajo porcentaje de crédito asignado.

Por otra parte, se identifica que no existe un set de atributos que describan a todos los alumnos desertores de forma única, cada una de las carreras tiene sus propios atributos característicos que permiten describir a los alumnos desertores, es así como en la cuatro carreras que se evaluaron se obtienen distintos resultados, donde sólo dos atributos se hacen transversales a todas las carreras que corresponden a la tasa de avance total de la carrera y el crédito asignado.

Con las últimas evaluaciones se logró obtener un set de atributos por cada carrera que permiten describir la deserción de los alumnos, como uno de las intenciones de este trabajo se centraba en ver la factibilidad de obtener un polinomio que pudiese describir numéricamente la deserción.

Utilizando metodología AHP, se ha identificado el peso de cada uno de los atributos, con lo que se han creado un algoritmo de evaluación, pese a todo pronostico se identificó con un promedio de un 90 % de confianza la totalidad de los desertores de cada una de las carreras, tomando un soporte del 100 % de los alumnos que desertan y no desertan.

Este valor es muy alentador, y que podría ajustarse un poco más dada la variabilidad de los valores discretos utilizados, lo que podría permitir aumentar la confianza de la predicción de la deserción en las carreras. Lo anterior, con la intención de llegar a resultados como los obtenidos en la carrera de Arquitectura que permite identificar un 96 % de los casos de deserción, sin afectar los correctamente clasificados en los no desertores.

Una de las preguntas complejas que se realizan en este análisis, se centra en si existen académicos que puedan llegar a tener mayor injerencia que otros en la deserción, al evaluar se identifica en todas las carreras en mayor o menor escala existen académicos que concentran una mayor cantidad de desertores, lo que hace pensar que debería haber algún tipo de estrategia de evaluación dentro de las cátedras que permita evaluar el proceso.

Es un tema complejo en realidad, que debe ser abordado por separado en este análisis dado que se requieren de otros instrumentos que pudiesen permitir evaluar a los académicos con mejor detalle.

Lo que si se debe destacar, es que no todos los académicos se encuentran capacitados o tienen las

competencias para enfrentar cualquier asignaturas, dado que existen casos particulares como el presentado en la Figura A.19, donde un mismo académico al realizar distintas cátedras permite que el alumno sea clasificado como desertor o no, esto explica que se hace importante clarificar las competencias de los académicos y el perfeccionamientos en los temas de su experticia.

En cierta forma, este análisis demuestra que con la aplicación de técnicas de minería de datos es factible obtener algunas causas de la deserción de los estudiantes, que permitirían otorgar medidas preventivas como por ejemplo implementar el polinomio y evitar que alumnos que se encuentren en alto riesgo de deserción pudiesen modificar su toma de asignaturas aumentando su carga académica.

Otra alternativa sería que todo alumno que ingresa a primer año, si reprueba un cierto porcentaje de asignaturas que lo deje en una tasa de avance menor al promedio de las carreras, enviarlo a un proceso de nivelación, permitiendo comenzar desde cero al siguiente año, evitando así mayores retrasos o perdida de estudios.

Son muchos los desafíos que nos deja hacia el futuro este análisis, recién hace poco más de un año se inicio el registro de las encuestas de apreciación docente, por lo que se hace interesante esperar unos periodos que permitan validar los problemas de los académicos con los resultados de las encuestas, para ver posibles planes de mejoras que permitan aumentar la tasa de retención de los alumnos.

Al conversar con la prorectoría académica, se hace de mucho interés evaluar las carreras que quedaron ausentes en especial aquellas que han iniciado sus procesos de innovación curricular de tal forma de contar con una mirada previa antes del cambio que permita tener un valor de partida.

Por otra parte, se hace muy interesante evaluar más en detalle la interacción que tienen las notas en el avance académico y la docencia para lo que sería interesante mezclar de alguna forma las tasas obtenidas por los alumnos en conjunto con sus evaluaciones y la asignación de crédito más los resultados de su postulación.

En base a las necesidades presentadas por la División académica crear indicadores de riesgo para los directores en base al polinomio que debe ser refinado, que permita a los alumnos aumentar su tasa de avance y con ello aumentar la tasa de retención.

Bibliografía

- [1] Pilar Armanet. ¿porque asegurar la calidad? en la educación superior. *Calidad en la educación, Concejo Superior de Educación*, 1(2):15, Agosto 2004.
- [2] Thomas M. Connolly; Carolyn E. Begg. Database systems : A practical approach to design, implementation, and management / T.M. connolly, C.E. begg., 2002.
- [3] Toscano Hurtado;Gérard Bruno. Proceso de análisis jerárquico (ahp). http://sisbib.unmsm.edu.pe/bibvirtualdata/tesis/basic/toskano_hg/cap3.pdf, 2006.
- [4] Fernando Arrau C. Indicadores de calidad de la educación superior en chile. *Biblioteca del Congreso Nacional de Chile Departamento de Estudios extencion y publicaciones*, 1(1):14, Enero 2005.
- [5] Cristina García Cambronero. Algoritmos de aprendizaje: Knn & kmeans. <http://www.it.uc3m.es/jvillena/irc/practicas/08-09/06.pdf>, 2009.
- [6] CNA Chile. Normas y procedimientos acreditación institucional. *Comisión Nacional de Acreditación*, 1(1):12, Enero 2010.
- [7] Fernando de la Jara Goyeneche. Necesidades de información desde la perspectiva desde la perpectiva del concejo superior de educación (cse). *Seminario Internacional, Concejo Nacional de Educación*, 0(5):8, Marzo 2001.
- [8] Francisco José Cortijo Bon; Universidad de la Republica Uruguay. Técnicas supervisadas ii. http://iie.fing.edu.uy/ense/asign/recpat/material/tema3_00-01/tema3_00-01_www.html, 2001.
- [9] Centro de Microdatos Universidad de Chile. Estudio sobre causas de la deserción universitaria. Informe interno, Agosto 2008.
- [10] Universidad de Waikato. Página principal del producto. <http://www.cs.waikato.ac.nz/ml/weka/>, 2011.
- [11] Emilio Rodríguez Ponce; Nicolás Fleet ; Milagros Delgado. La acreditación en la generación de información sobre la calidad de la educación superior. *Calidad en la educación, Concejo Superior de Educación*, 2(31):19, Julio 2009.
- [12] Luis R. Rivera Fernández. Data warehousing and business intelligence. *Universidad Interamericana de Puerto Rico, Inteligencia de negocio y bodegas de datos*, 1(1):20, 2009.
- [13] Cristian Aedo; Luis Eduardo González. La educación superior en chile. *Calidad en la educación, Concejo Superior de Educación*, 1(1):12, Enero 2010.
- [14] Rafael Jesús Montero González. Algoritmo de inducción de árboles de decisión (id3). [www.uco.es/ i92mogor/trabajos/id3/id3.doc](http://www.uco.es/i92mogor/trabajos/id3/id3.doc), 2003.
- [15] Cindi Howson. *Successful Business Intelligence Secrets to Making BI a Killer App*, volume 1. Mc Graw Hill, 1 edition, 2008.

- [16] IIIA-CSIC. Minería de datos o data mining. <http://www.iiia.csic.es/udt/files/DataMining.pdf>, 2009.
- [17] kdnuggets.com. Data mining/analytic tools used. <http://www.kdnuggets.com/polls/2011/tools-analytics-data-mining.html>, 2011.
- [18] Jing Luan PhD. Data mining applications in higher education. *SPSS Inc*, 1(1):8, 2006.
- [19] José Hernández Orallo ; María José Ramírez Quintana; César Ferri Ramírez. *Introducción a la minería de datos*, volume 1. Pearson, 1 edition, 2004. ISBN: 8420540919.
- [20] Iberoamerica S.A. Business intelligence. <http://managersmagazine.com/wp-content/uploads/2009/10/download.pdf>, Octubre 2009.
- [21] Eduardo Nicolás Campazzo; Virginia Inés Santos. Business intelligence: Negocios inteligentes para empresas inteligentes. *Universidad Nacional de La Rioja*, 1(1):7, 2008.
- [22] Óscar Arias Sebastián Donoso; Gonzalo Donoso. Iniciativas de retención de estudiantes de educación superior. *Revista Calidad en la Educación, Concejo Nacional de Educación*, 2(33):8, Julio 2010.
- [23] Corp Solutions. La importancia business intelligence. <http://www.corp-solutions.com.ar/briefs/Business.Intelligence.pdf>, 2008.
- [24] Pete Chapman (NCR); Julian Clinton (SPSS); Randy Kerber (NCR); Thomas Khabaza (SPSS); Thomas Reinartz (DaimlerChrysler); Colin Shearer (SPSS) and Rüdiger Wirth (DaimlerChrysler). Crisp-dm process model 1.0. <http://www.crisp-dm.org/CRISPWP-0800.pdf>, 2000.
- [25] Xindong Wu · Vipin Kumar ; J. Ross Quinlan ; Joydeep Ghosh ; Qiang Yang ; Hiroshi Motoda ; Geoffrey J. McLachlan ; Angus Ng · Bing Liu ; Philip S. Yu ; Zhi-Hua Zhou ; Michael Steinbach ; David J. Hand ; Dan Steinberg. *Top 10 algorithms in data mining*, volume 1. Springer, 1 edition, 2007.
- [26] IESALC UNESCO. Estudio sobre la repitencia y deserción en la educación superior chilena. *Digital Observatory for higher education in latin america and the caribbean*, 1:55, Abril 2002.
- [27] Luis Eduardo González; Daniel Uribe. Estimaciones sobre la repitencia y deserción en la educación superior chilena. consideraciones sobre sus implicaciones. *Calidad en la educación, Consejo Superior de Educación*, 2(17):19, Agosto 2002.
- [28] Nadja Alejandra Starocelsky Villavicencio. Aplicar minería de datos a la predicción de la deserción retención de estudiantes. Master's thesis, Universidad Técnica Federico Santamaría, 2010.
- [29] R. Alcover; J. Benlloch; P. Blesa; M. A. Calduch; M. Celma; C. Ferri; J. Hernández-Orallo; L. Iniesta; J. Más; M. J. Ramírez-Quintana; A. Robles; J. M. Valiente; M. J. Vicent; L. R. Zúnica. Análisis del rendimiento académico en los estudios de informática de la universidad politécnica de valencia aplicando técnicas de minería de datos. *XIII Jornadas de Enseñanza Universitaria de la Informática*, 1(1):8, Julio 2007.

Apéndice A

Anexo 1

A.1. Análisis Carrera de Arquitectura

La carrera de Arquitectura es una de las carreras más antigua en la universidad, cuyo plan de estudios se encuentra compuesto por 53 asignaturas en total. Para esta evaluación se utilizaron los datos de los alumnos que ingresaron a esta posterior al año 2004, debido que desde esta fecha se rinde la PSU, siendo la cantidad de registros utilizados igual a 619.

A.1.1. Análisis Exploratorio

Posterior al trabajo de discretizar los valores, se realizó una análisis exploratorio de los datos para identificar algunas reglas genéricas de la información. El primer primer aspecto detectado es que existe una mayor cantidad de alumnos desertores que de aquellos que continúan en la carrera: de un total de 619 alumnos, mas de 300 de ellos han desertado de la carrera lo que corresponde al 57,5%. Esta información es muy relevante dado que la permanencia de los alumnos es muy baja con respecto a las otras carreras, lo que podría justificar evaluar la estructura de la carrera. Por otra parte al evaluar el impacto de los atributos en los alumnos se puede identificar que la tasa de aprobación y el puntaje ponderado con el cual ingresa el alumno juegan un rol fundamental en la decisión del alumno al momento de desertar a la carrera, como se muestra en la Figura A.1. Frente a la evaluación de los atributos de la carrera se identifica que los alumnos desertores principalmente lo realizan antes de cumplir un 10% de avance de la carrera, lo que equivale a no más de 5 asignaturas aprobadas, siendo más extremista que lo ocurrido con la carrera de Ingeniería en Construcción, por ende, hay un problema grave en el primer año de estudios de los alumnos.

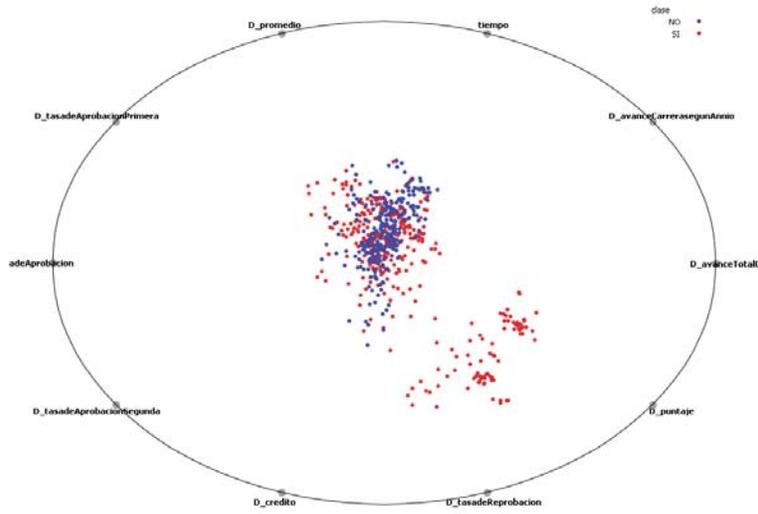


Figura A.1: Distribucion de datos según atributos

Un caso importante que debe ser evaluado es que al considerar la tasa de avance por año de ingreso, es posible identificar dos grupos de desertores: los que tienen menos de un 10% y los que alcanzan un 100% de la tasa de avance; situación que se puede corroborar en la Figura A.2.

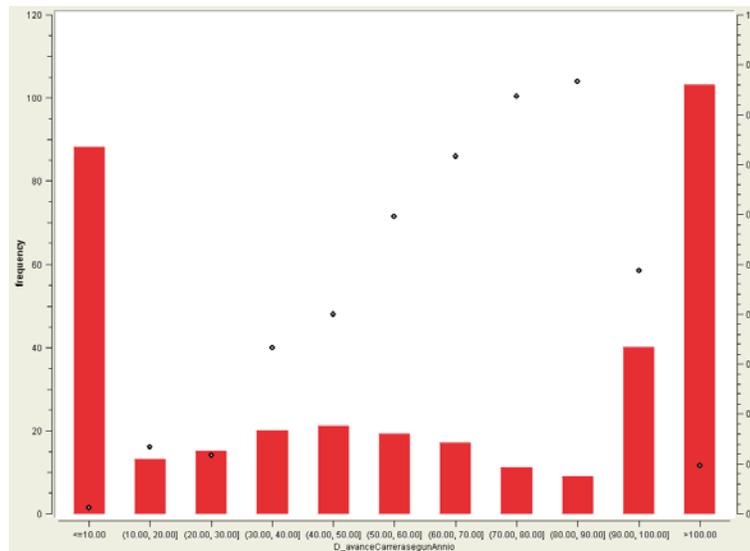


Figura A.2: Alumnos desertores según tasa de avance

Debido a que la tasa de aprobación es muy baja por parte de los alumnos, el hecho que los alumnos tengan una tasa de avance tan alta con respecto a su año de ingreso se puede deber a que ellos pudiesen haber convalidado asignaturas, lo que sería muy interesante de evaluar al futuro.

Junto con lo anterior se valida que existen atributos que inciden en varias carreras, donde junto con la baja tasa de aprobación de los alumnos, se repite que los alumnos desertores prácticamente no tienen créditos o becas como se puede observar en la Figura A.3.

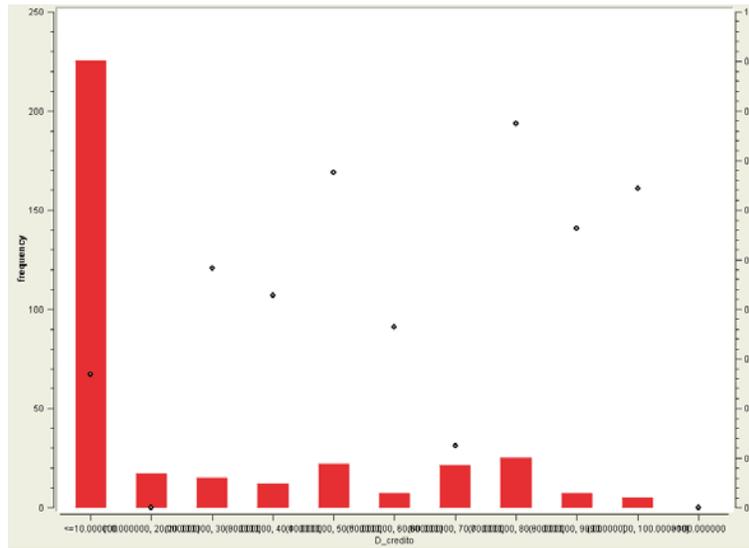


Figura A.3: Crédito asignados a los alumnos desertores

A.1.2. Clasificación

Para realizar la evaluación se ha utilizado el mismo conjunto de datos y el mismo ranking de atributos utilizando el algoritmo de ganancia de información, del cual se ha obtenido como atributos más relevantes 14 de ellos. A diferencia de la carrera de Ingeniería en Construcción, en este caso sí toman relevancia atributos que pertenecen a la condición de ingreso del estudiante como son el puntaje y el lugar de selección. Para realizar esta parte del análisis se han aplicado los árboles ID3, CART y C4.5 aplicados con anterioridad, comenzando con CART; este entrega muchas reglas en un modelo con muchos niveles, donde algunas de ellas consideran además del crédito del alumno, como atributo que condiciona la renuncia, el promedio ponderado como se puede visualizar en la regla a continuación. Figura A.4. En esta regla se puede ver que a pesar que las reglas de clasificación son iguales y el promedio es quien determina la clasificación de estas, donde alumnos con un promedio entre un 3.0 y 4.0 desertan al contrario de quienes tienen promedios entre un 2.0 y un 3.0, cuando la tasa de avance es muy baja y tienen una importante tasa de aprobación en segunda oportunidad.

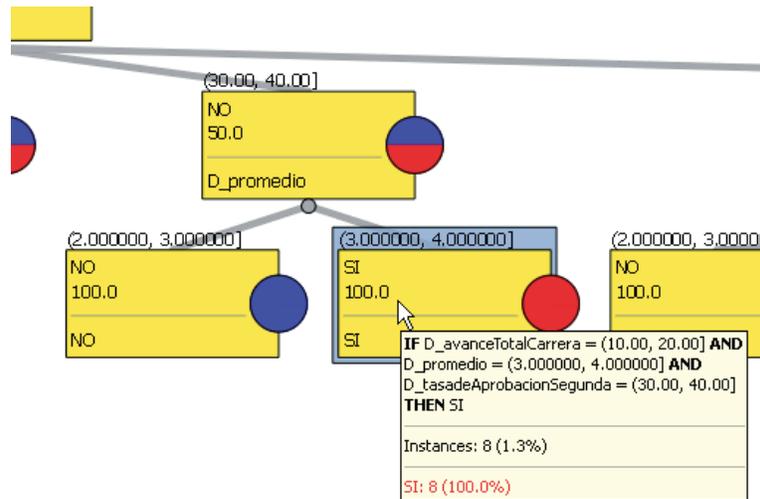


Figura A.4: Clasificación alumnos desertores debido a diferencias en el promedio de notas.

Un caso particular se produce con los alumnos que en su segundo año desertan principalmente por una baja tasa de avance en la carrera donde aquellos que tienen menos de un 20% de avance general en la carrera y ya han cursado su segundo año desertan a la carrera, pese a contar con una alta tasa de aprobación de asignaturas cursadas por primera, donde estos casos tienen muy pocas asignaturas cursadas por segunda vez que los han retrasado fuertemente y que provocan que renuncien a la carrera. Figura A.5

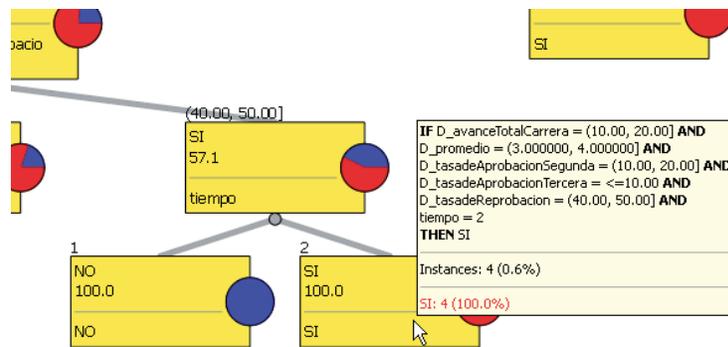


Figura A.5: Clasificación alumnos desertores baja tasa de avance.

Las otras reglas encontradas en este modelo ratifican en gran medida lo ya expuesto en el caso estudiado con anterioridad. Junto con esta evaluación se debe destacar lo que se encontró al usar el algoritmo ID3 implementado en la aplicación WEKA, a través del cual se han obtenido los resultados como que el crédito asignado a un alumno sí es importante a la hora de determinar si renuncia o no, como se muestra en la Figura A.6.

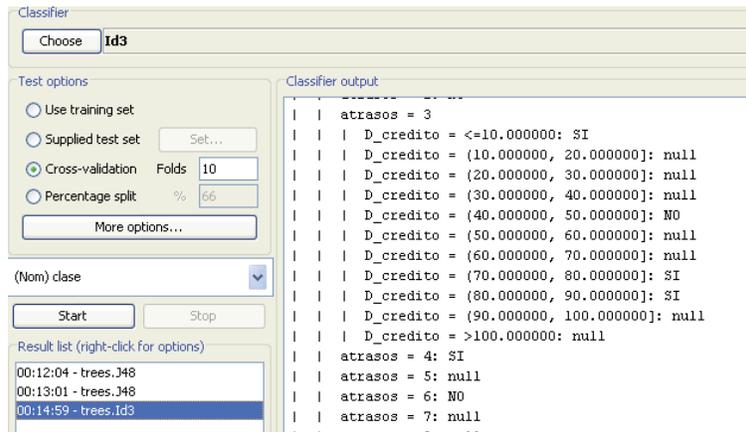


Figura A.6: Clasificación alumnos desertores usando ID3 con WEKA .

Se ratifica la importancia que tienen para los alumnos de Arquitectura el promedio de notas y el avance según su año ingreso, ante una misma condición de promedio y avance general, el avance individual del alumno es determinante para la decisión de desertar o no.

Además de lo anterior se observa que la tasa de avance académico menor 50 % aumenta la probabilidad de desercion, donde los alumnos con un 20 % y un 30 % de avance académico desertan, por el contrario alumnos tiene una tasa de avance mayor al 50 % no deserta pese a tener un bajo promedio y una baja tasa de avance total en la carrera.

Con respecto al árbol C4.5 en realidad el conocimiento otorgado es muy bajo, aunque se realizaron muchos cambios sus parametros de ejecución no se logró obtener alguna regla que pudiese ser relevante o interesante.

A.1.2.1. Evaluación de la Clasificación

Pese a que las reglas obtenidas en la evaluación de CART en Orange son más definidas, de todas formas se ha evaluado cada una de las matrices de confusión obtenidas con los distintos algoritmos, como se muestra en la Figura A.7.

| | | J48 | | | | C4.5 | | | | CART | |
|----|--|-----|-----|--|--|-------|-----|--|--|-------|-------|
| | | NO | SI | | | NO | SI | | | NO | SI |
| NO | | 85% | 18% | | | 80,4% | 20% | | | 83,2% | 17,4% |
| SI | | 15% | 82% | | | 19.6% | 80% | | | 16,8% | 82,6% |

Figura A.7: Comparación matrices de confusión.

Tal como se puede observar fue el algoritmos ID3 de WEKA el que obtuvo mejores resultados de la clasificación; sin embargo no se puede desechar que reglas muy importantes se obtuvieron con CART de Orange, pues las reglas obtenidas ratifican el aprendizaje obtenido. Se debe agregar que el peor caso se refiere a C4.5 de Orange donde la clasificación tiene muchos errores de clasificación, pese a que los resultados en CART Orange no son excelentes si presentan buenos resultados como se muestra en su curva ROC (ver Figura A.8).

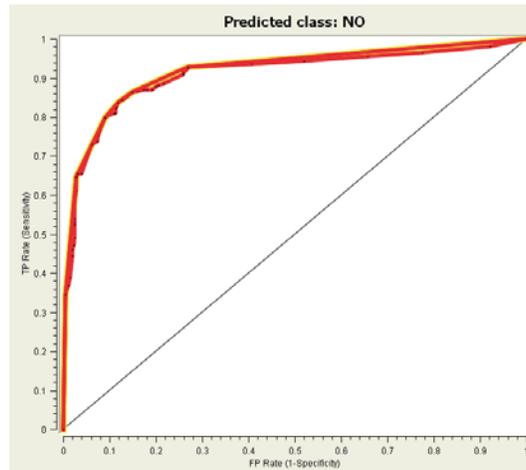


Figura A.8: Curva ROC árbol CART

A.1.3. Evaluación de la Carrera con la aplicación de Tasa en Semestres

Al utilizar el avance en semestre de los alumnos se puede determinar que la deserción ocurre principalmente en el segundo semestre de la carrera. Dado esto se puede indicar que el semestre sí es importante al momento de realizar un seguimiento de los alumnos.

Además, se repite que principalmente la deserción ocurre en los primeros semestres de la carrera, de preferencia en el segundo semestre y en general antes del cuarto semestre, tal como se muestra en la Figura A.9

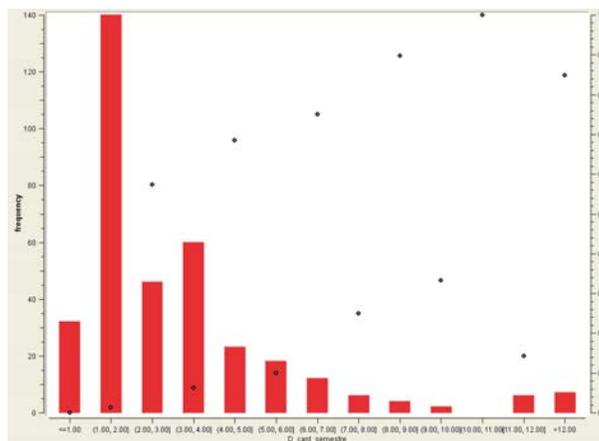


Figura A.9: Cantidad de semestres que cursa un alumnos desertor en la carrera de Arquitectura.

Por otra parte durante el análisis se ha identificado que existe un gran número de alumnos con una tasa de avance menor al 50% y que en algunos casos es menor al 10%, validando la causa de eliminación por motivos académicos de los alumnos, donde la cantidad de alumnos eliminados académicamente de la carrera es muy representativa como se visualiza en la Figura A.10, por lo que la evaluación de las notas es muy importante para identificar mayores detalles de la deserción de los estudiante.

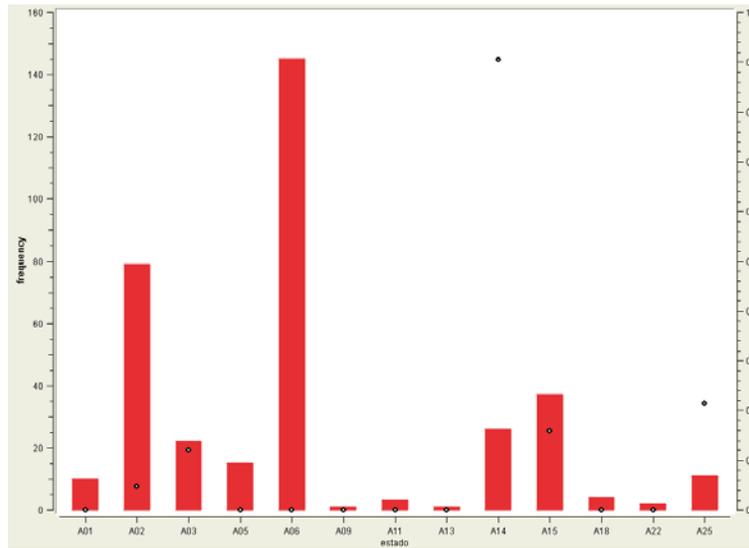


Figura A.10: Cantidad de alumnos por estados en la carrera de Arquitectura.

Al realizar un ranking usando como medida la ganancia de información, es posible determinar que los datos de ingreso (puntaje y lugar de selección) tienen un peso significativo para describir la deserción, como se muestra en la Figura A.11.

| | Attribute | # | RelieFF | Inf. gain | Gain ratio | Gini | log OR (for 'SI') |
|----|--------------------|----|-----------|-----------|------------|----------|-------------------|
| 1 | D_cant_semestre | 13 | 0.460341 | 0.471098 | 0.146111 | 0.129283 | NA |
| 2 | semestre | 3 | 0.357099 | 0.395454 | 0.408760 | 0.108570 | NA |
| 3 | D_avanceTotalCa... | 11 | 0.409342 | 0.387709 | 0.145769 | 0.109289 | NA |
| 4 | D_avanceCarrer... | 11 | 0.436704 | 0.344532 | 0.106009 | 0.099822 | NA |
| 5 | D_avance_segun... | 11 | 0.453531 | 0.336923 | 0.115370 | 0.098769 | NA |
| 6 | D_tasadeAproba... | 11 | 0.353008 | 0.248915 | 0.079988 | 0.072009 | NA |
| 7 | D_tasadeAproba... | 11 | 0.207549 | 0.204570 | 0.080400 | 0.058525 | NA |
| 8 | D_tasadeReprob... | 11 | 0.308099 | 0.202003 | 0.063889 | 0.060085 | NA |
| 9 | D_credito | 11 | 0.092444 | 0.162624 | 0.069034 | 0.051156 | NA |
| 10 | D_tasadeAproba... | 11 | 0.113502 | 0.127075 | 0.071738 | 0.041488 | NA |
| 11 | D_promedio | 8 | 0.255390 | 0.086015 | 0.040510 | 0.023600 | NA |
| 12 | D_beca | 11 | 0.060248 | 0.074019 | 0.062553 | 0.020447 | NA |
| 13 | D_puntaje | 10 | 0.068177 | 0.024244 | 0.013933 | 0.008086 | NA |
| 14 | D_lugar | 11 | 0.008310 | 0.021555 | 0.006329 | 0.007264 | NA |
| 15 | D_region | 17 | -0.028281 | 0.019988 | 0.009496 | 0.005997 | NA |
| 16 | tipocolegio | 10 | -0.027999 | 0.014785 | 0.013393 | 0.003550 | NA |
| 17 | D_puntajenem | 10 | 0.100668 | 0.014730 | 0.005292 | 0.004253 | NA |
| 18 | D_tasadeAproba... | 11 | 0.062757 | 0.009119 | 0.015502 | 0.002074 | NA |
| 19 | preferencia | 9 | -0.057915 | 0.005180 | 0.002242 | 0.001428 | NA |
| 20 | grupo_dependen... | 4 | 0.054545 | 0.005044 | 0.002861 | 0.001699 | NA |
| 21 | estadosseleccion | 3 | 0.006117 | 0.001517 | 0.002200 | 0.000518 | NA |

Figura A.11: Ranking atributos carrera B.

Al tomar en cuenta los semestres se ve que la cantidad de semestres cursados comparado con la tasa de avance y el hecho de no tener crédito asignado puede ser causa de deserción. Figura A.12. De la misma figura se puede observar que los alumnos con un bajo avance en la carrera y que se encuentran en el 4 año de su

carrera con una importante tasa de aprobación por segunda vez, son clasificados como desertores, lo que se revierte con los alumnos que tienen mejores tasas de avance en la carrera.

Lo anterior se refuerza con la importancia que presenta para los alumnos la tasa de avance total de la carrera en reglas de clasificación, en que la decisión de deserción de la carrera se encuentra depende del porcentaje de avance.

De forma general, el avance académico de los alumnos en la carrera de Arquitectura en conjunto con el porcentaje de crédito asignado, juegan un rol fundamental al momento de tomar la decisión de desertar, por lo que se hace muy importante evaluar los antecedentes académicos y las notas de la carrera.

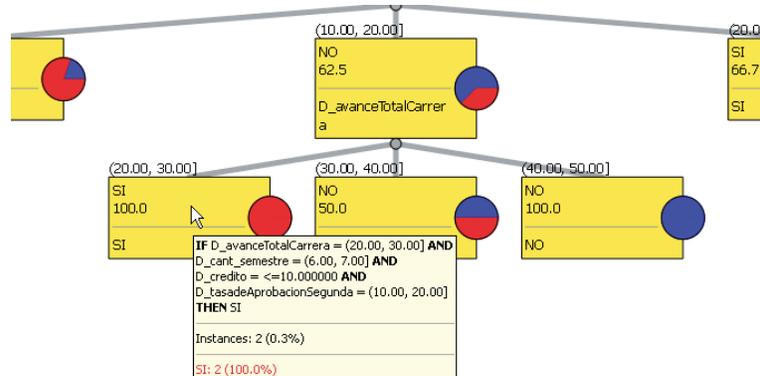


Figura A.12: Clasificación importancia avance total carrera carrera de Arquitectura.

A.1.4. Minería de datos en las evaluaciones de los alumnos de Arquitectura.

Para evaluar la antecedentes de las notas de los alumnos se ha utilizado el mismo conjunto de datos anteriormente descrito, en este caso el grupo de datos está compuesto por 11063 registros que corresponden a todas las notas de todos los alumnos de la carrera de Arquitectura con la que se ha analizado la deserción. A primera vista de los datos se puede identificar que los atributos que describan fuertemente la deserción en base a las notas se encuentra descrita principalmente por la asignatura, el profesor y la cantidad de alumnos que tiene la asignatura, por lo que es interesante identificar las asignaturas que pueden ser mas representativas de la deserción, como lo muestra la Figura A.13

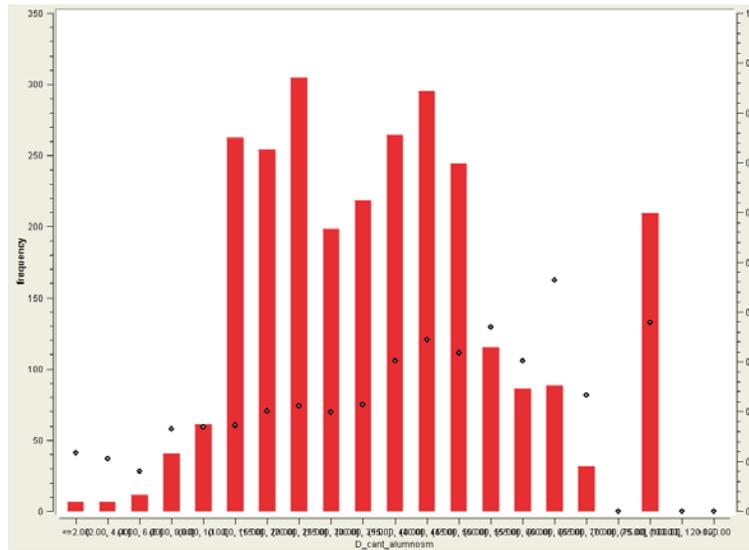


Figura A.15: Cantidad de desertores por asignatura en la carrera de Arquitectura.

Al igual que el caso anterior y con la complejidad que implica existen al menos dos casos relevantes de académicos que podrían representar la deserción de los alumnos de arquitectura, como se puede visualizar en la Figura A.16. En este caso se ha evaluado sobre el total de las notas incluyendo aprobadas y reprobadas siendo un resultado genérico.

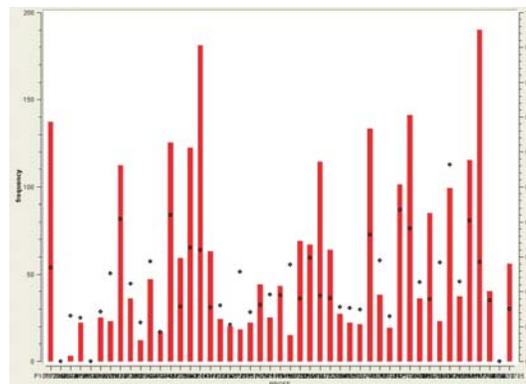


Figura A.16: Cantidad de desertores por profesor en la carrera de Arquitectura.

Al evaluar el caso de los alumnos reprobados se identifica que efectivamente hay al menos tres académicos que podrían llegar a describir la deserción en los alumnos que reprueban asignaturas, lo que complica el análisis dado que es necesario otro tipo de medidas por parte de las autoridades de la carrera y requiere que se evalúe el problema por medio de instrumentos adicionales.

La clasificación otorga reglas interesantes para la evaluación de la carrera la primera de ellas revela la importancia que tiene en algunos casos la cantidad de alumnos donde para el mismo profesor en las mismas condiciones identifica a los alumnos como desertores dependiendo directamente de la cantidad de alumnos que tiene la asignatura. Lo extraño y que se debe analizar las causas por que cuando se tiene menor cantidad de alumnos provoca deserción. Figura A.17

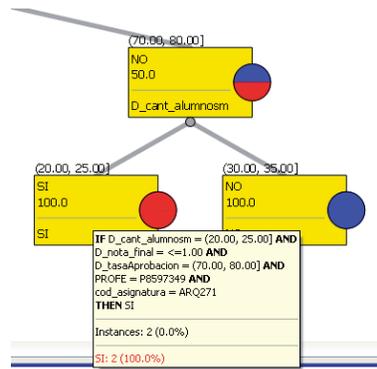


Figura A.17: Cantidad de alumnos en cátedra carrera de Arquitectura.

Un punto muy relevante en la carrera de Arquitectura es que el árbol de clasificación describe en su raíz al profesor de la asignatura, entregando una gran importancia a este atributo, donde depende de las condiciones y el curso que imparta la condición que puede generar ante la decisión de desertar o continuar; existen profesores que son muy relevante a la continuidad de los alumnos, como los casos contrarios que complican el análisis dado que apunta a personas y su trabajo.

Otra regla interesante se encuentra en la tasa de aprobación que logran los académicos que debe ser un punto de partida por parte de ellos debido que a mayor tasa de aprobación en algunos casos se prolonga la continuidad del alumno que en casos contrarios que deserta como se muestra en la Figura A.18.

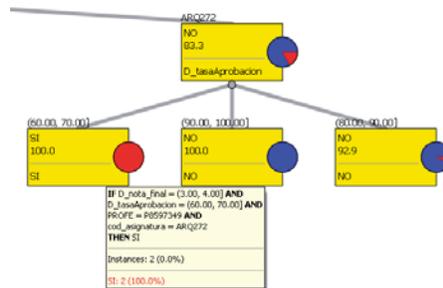


Figura A.18: Tasa de aprobación por asignatura carrera de Arquitectura.

Por último, en este caso se identifica que debe existir algún tipo de parámetro de evaluación de los docentes que imparten algunas asignaturas dado que en algunos casos pueden apoyar la continuidad de los alumnos y en otros pueden llegar a decidir la deserción como se muestra en la Figura A.19.

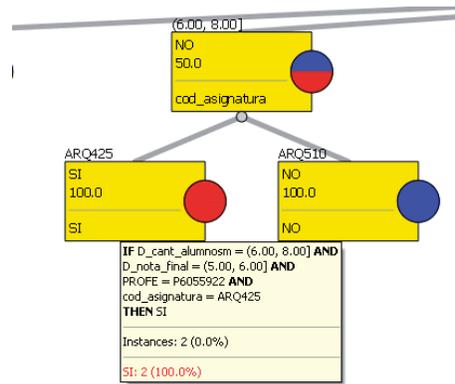


Figura A.19: Caso Clasificación deserción dependiendo de la asignatura en Arquitectura.

Apéndice B

Anexo 2

B.1. Análisis Carrera de Auditoría (Diurno)

La carrera de Auditoría (Diurno) es una carrera del área económica y se encuentra compuesta por 48 asignaturas en total, para esta evaluación se utilizaron los datos de los alumnos que ingresaron a ésta posterior al año 2004, donde la cantidad de registros obtenidos corresponde a 844 items.

Los atributos utilizados para esta evaluación son los mismos que se han utilizado en las dos carreras anteriores, por lo que este conjunto de datos ya incluye las tasas por asignaturas y por semestres.

B.1.1. Análisis Exploratorio

El análisis al igual que todas las carreras se realizó utilizando la herramienta Orange canvas; posterior a la discretización de los datos se puede identificar que a pesar que la cantidad de desertores es menor que el total de alumnos que no desertan, el número no deja de ser menor, acercándose al 50 % del total.

Un antecedente interesante en la carrera se encuentra en que el 39 % de los alumnos proviene de colegios técnicos, lo que implica, que estos alumnos tienen una menor preparación para enfrentar las asignaturas de ciencias básicas, lo que puede impactar directamente en los resultados de éstas, en comparación al 49 % de los alumnos que provienen de colegios o liceos científicos humanistas.

Al explorar los datos basados se puede identificar que los alumnos desertores se encuentran descritos en aquellos cuya tasa de avance total de la carrera es menor a un 20 %, lo que infiere que en su mayoría son alumnos que se encuentran en sus primeros años, lo que puede ser observado en la Figura B.1.

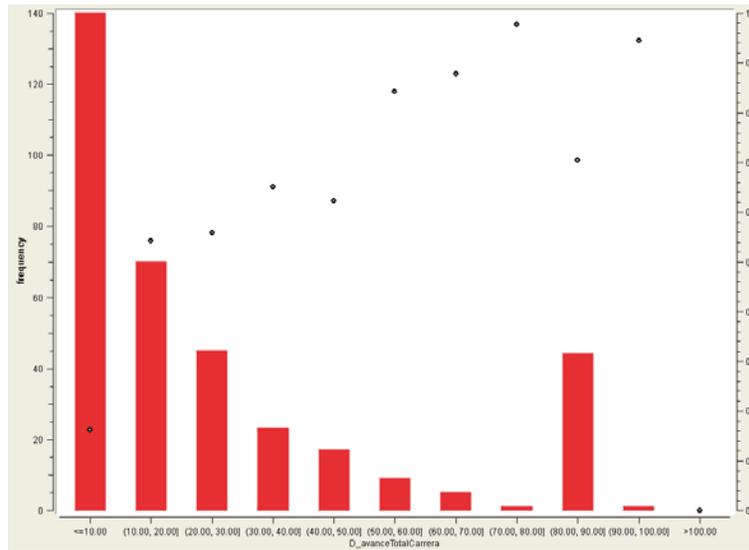


Figura B.1: Cantidad de alumnos desertores según tasa total de avance en la carrera

Por otra parte se debe destacar que existe un problema de avance académico por parte de los alumnos; un grupo menor de alumnos tiene un avance del 100% con respecto al que debería llevar y por otra parte se encuentran demasiados alumnos con problemas graves de avance los que tienen menos de un 10% con respecto al tiempo que tienen.

El problema académico que se identifica (la tasa de avance según año de ingreso) queda ratificado al visualizar los promedios de notas de los alumnos desertores los que principalmente presentan promedios menor a un 4.0, lo que implica directamente que se requiere una herramienta inicial que permita a los alumnos una mayor continuidad. Al igual que en las carreras ya evaluadas los alumnos desertores se encuentran descritos por aquellos alumnos que no tienen apoyo económico de créditos y becas, lo que podría inferir en la necesidad de otorgar mejores apoyos económicos para los alumnos. ver Figura B.2

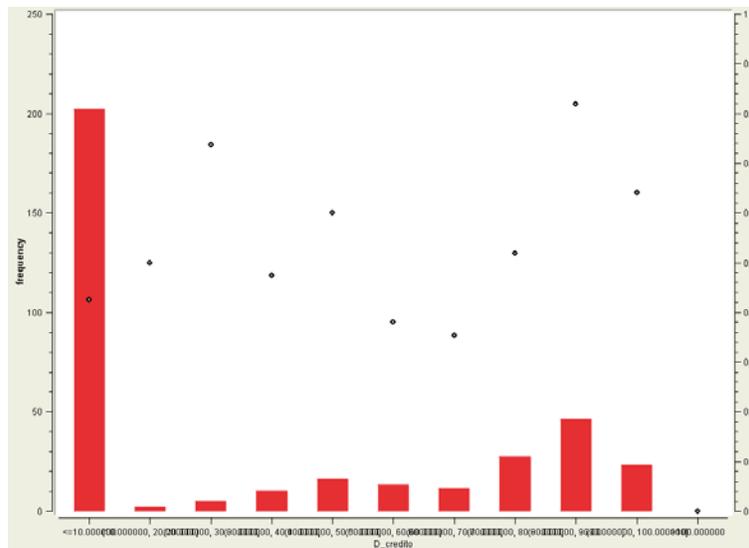


Figura B.2: Cantidad de alumnos desertores según según porcentaje de crédito asignado

Un punto importante es que los alumnos desertores son principalmente por alumnos que ingresan de colegios particulares subvencionados (código 2) y municipalizados (código 3) por lo que sería interesante planificar algún tipo de refuerzo al momento de ingresar estos alumnos. ver Figura B.3

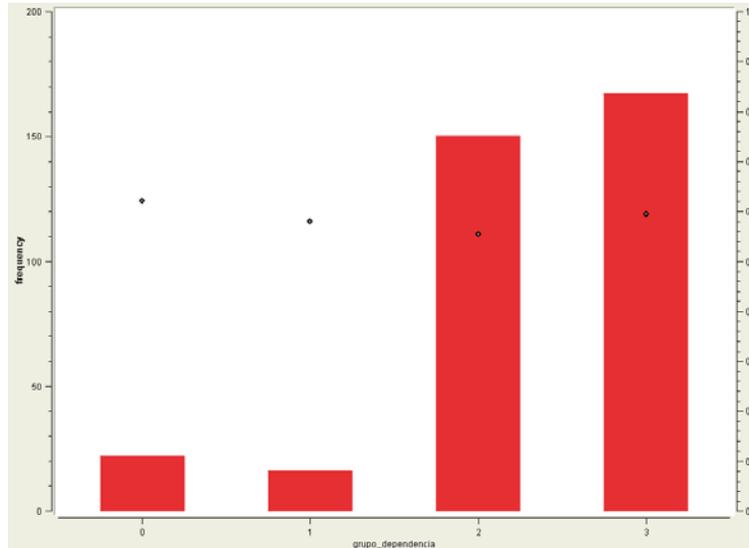


Figura B.3: Cantidad de alumnos desertores según el tipo de dependencia de los colegios.

Junto a lo anterior, se identifica que los alumnos desertores son en su mayoría aquellos que quedan convocados en un lugar sobre el puesto 100 y que pertenecen principalmente a la lista de seleccionados. ver Figura B.4

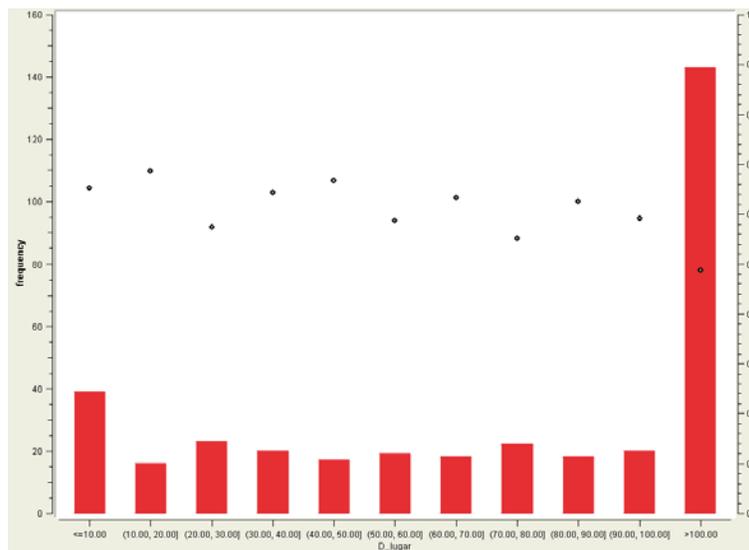


Figura B.4: Cantidad de alumnos desertores por tramo de selección.

Al igual que los casos anteriores la deserción por parte de los alumnos de la carrera ocurre principalmente en los primeros años de la carrera, lo que hace muy importante el avance total de alumno, al momento de evaluar el riesgo de deserción.

B.1.2. Evaluación de la Carrera con Algoritmos

B.1.2.1. Clasificación

Para realizar la evaluación de la clasificación de los datos generales se ha realizado una evaluación y ranking de éstos utilizando ganancia de información, obteniéndose 11 atributos representativos los cuales se muestran en la Figura B.5:

| | Attribute | # | ReliefF | Inf. gain | Gain ratio | Gini | log OR (for 'SI') |
|----|---------------------------|----|-----------|-----------|------------|----------|-------------------|
| 1 | D_cant_semestre | 15 | 0.219068 | 0.252583 | 0.072808 | 0.075837 | NA |
| 2 | D_avance_segun_sem | 11 | 0.197348 | 0.214100 | 0.077954 | 0.066295 | NA |
| 3 | D_avanceTotalCarrera | 11 | 0.176920 | 0.193653 | 0.063410 | 0.058607 | NA |
| 4 | semestre | 3 | 0.192811 | 0.190169 | 0.272414 | 0.058941 | NA |
| 5 | D_avanceCarreraSegunAño | 11 | 0.181942 | 0.187807 | 0.059010 | 0.056709 | NA |
| 6 | D_tasadeAprobacion | 11 | 0.104865 | 0.166495 | 0.054295 | 0.049849 | NA |
| 7 | D_tasadeReprobacion | 11 | 0.121688 | 0.145472 | 0.048769 | 0.045290 | NA |
| 8 | D_tasadeAprobacionPrimera | 11 | 0.068016 | 0.117516 | 0.047135 | 0.034069 | NA |
| 9 | D_credito | 11 | 0.154293 | 0.095285 | 0.040950 | 0.030478 | NA |
| 10 | D_promedio | 8 | -0.016433 | 0.077501 | 0.042504 | 0.022790 | NA |
| 11 | D_beca | 11 | 0.010375 | 0.061300 | 0.079257 | 0.017049 | NA |
| 12 | D_tasadeAprobacionSegunda | 11 | 0.009630 | 0.027790 | 0.015007 | 0.008853 | NA |
| 13 | D_puntaje | 10 | 0.066946 | 0.027115 | 0.019176 | 0.009139 | NA |
| 14 | D_tasadeAprobacionTercera | 11 | 0.025744 | 0.021885 | 0.039820 | 0.006557 | NA |
| 15 | tipocollegio | 10 | -0.062876 | 0.019450 | 0.012030 | 0.006241 | NA |
| 16 | D_lugar | 11 | -0.115524 | 0.015594 | 0.005036 | 0.005251 | NA |
| 17 | D_region | 17 | 0.014233 | 0.007888 | 0.005448 | 0.002445 | NA |
| 18 | D_puntajenem | 10 | 0.018426 | 0.007392 | 0.002740 | 0.002523 | NA |
| 19 | preferencia | 9 | -0.095074 | 0.003909 | 0.001574 | 0.001317 | NA |
| 20 | grupo_dependencia | 4 | -0.074053 | 0.001351 | 0.000900 | 0.000456 | NA |
| 21 | estadoseleccion | 3 | -0.029597 | 0.000616 | 0.000981 | 0.000206 | NA |

Figura B.5: Ranking de los atributos utilizados en la carrera de Auditoria (Diurno).

Con este nuevo conjunto de dato se obtienen algunas reglas interesantes, siendo el atributo más importante la cantidad de semestres. Luego, al pasar a la etapa de aplicación de algoritmos, se visualiza la importancia que tiene la cantidad de semestres junto a la tasa de avance según año del alumno, mientras menor es la tasa de avance mayor es la probabilidad de deserción; alumnos con una tasa de avance menor al 20 % son calificados como alumnos desertores, pese a tener avances según año de un 90 %. Figura B.6.

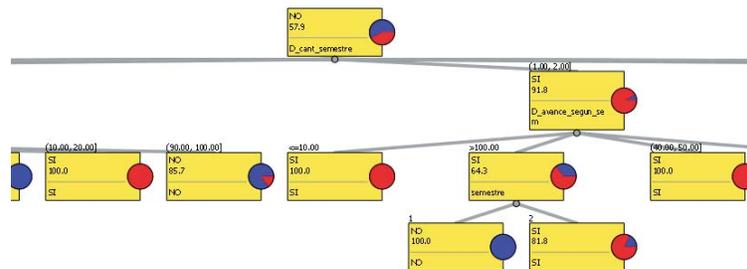


Figura B.6: Clasificación importancia de los semestres cursados.

Además alumnos con una tasa de aprobación menor a un 60 %, identifica deserción cuando la tasa de avance según semestre es mayor al 100 %; esto ocurre con los alumnos que han homologado sus asignaturas, provocando un gran avance de sus asignaturas generando una muy baja tasa de avance total de la carrera.

Por este motivo se hace importante realizar un seguimiento de los alumnos que ingresan por traslados, transferencia y que homologan sus asignaturas. Además las reglas identifican la importancia que tienen los

antecedentes académicos en la evaluación de la deserción de la carrera, para alumnos con una baja tasa de avance por semestres y bajos promedios serán desertores. ver Figura B.7

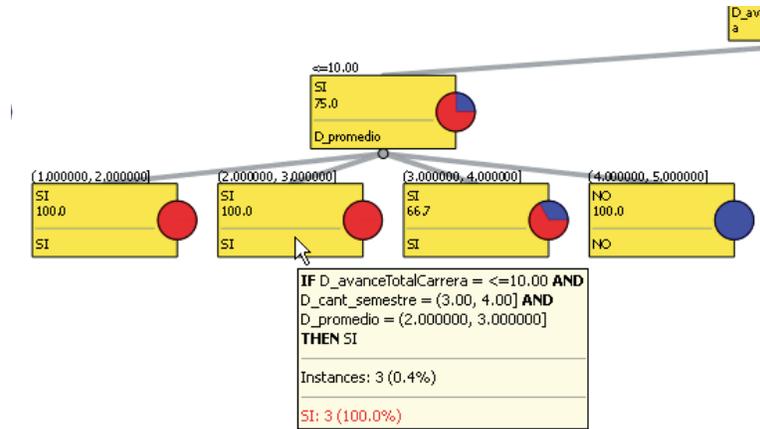


Figura B.7: Clasificación importancia tasa de avance según año.

Al aplicar el algoritmo C4.5, la cantidad de reglas son menores que las presentadas por el algoritmo CART anterior; lo que sí identifica es que repite la importancia en el avance académico por parte de los alumnos como (ver Figura B.8).

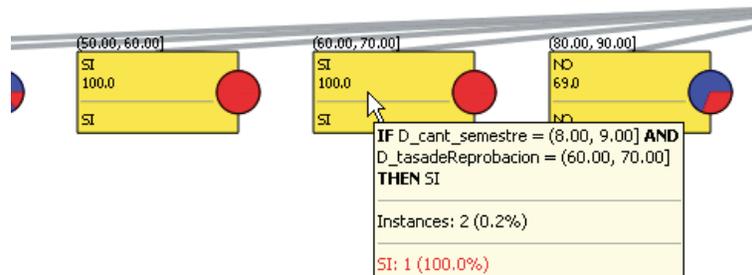


Figura B.8: Clasificación aplicacndo algoritmo C4.5.

El caso particular presentado revela que es importante el seguimiento de los alumnos pues quienes tienen una alta tasa de aprobación encontrándose en los últimos semestres desertan en comparación a los alumnos con baja tasa de aprobación; esto se puede deber por que se trata de alumnos en los semestres finales y prefieren salir a trabajar que terminar su ciclo de estudios. Es muy común que en estos niveles los alumnos tengan ofertas tentadoras de trabajo en especial aquellos que tienen buena tasa de aprobación.

Con el algoritmo C4.5 se puede identificar que la deserción de los alumnos en la carrera de Auditoría depende directamente del avance en semestres que se tenga; además puede indicar que a menor avance mayor es el riesgo de la carrera ante la posible deserción, donde alumnos con baja cantidad de semestres son clasificados como desertores.

A medida que el alumno avanza en la carrera se identifican nuevos atributos que toman relevancia al momento de desertar, como es el porcentaje de crédito que tienen con dos a tres semestres cursados, en el cual depende directamente del porcentaje de crédito asignado en que se presentan alumnos con créditos asignados menores al 10 % clasificándolos como desertores.

En cursos superiores para alumnos que tienen ocho o nueve semestres se identifica el atributo tasa de reprobación como relevante al momento de decidir la deserciónlo anterior se puede ver en la Figura B.9

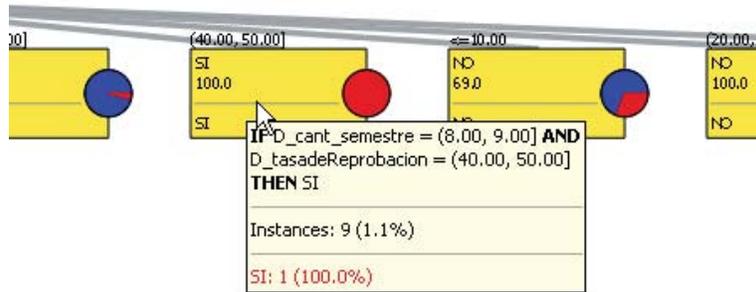


Figura B.9: Clasificación Efecto de la tasa de reprobación

Utilizando el algoritmo ID3 de WEKA, se identifican reglas interesantes; en primer lugar se ratifica la importancia de la cantidad de semestres en el conjunto de datos, a medida que el avance es mayor, menor es la probabilidad de deserción. Al evaluar esta deserción se identifica la importancia que tiene el avance académico en los primeros cursos de los alumnos desertores, donde todo alumno con tasa aprobación menor a un 50 % en los primeros semestres de su carrera puede ser un alumno desertor, como se puede ver en la Figura B.10.

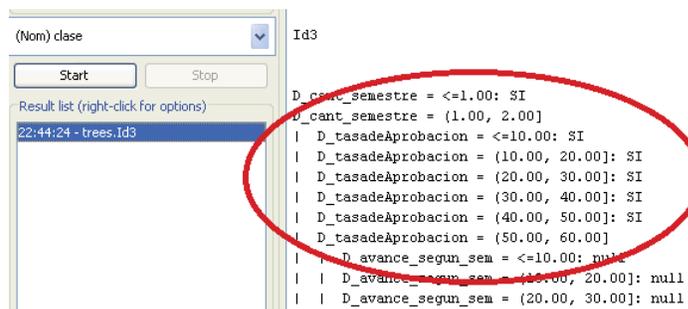


Figura B.10: Clasificación Efecto de la tasa de reprobación

Para otros casos, el avance total de la carrera juega un rol fundamental para determinar la continuidad de un alumno. Además, la clasificación permite identificar un parámetro de decisión de los alumnos con bajo crédito que corresponde a la tasa de reprobación que tienen en los primeros años: alumnos con una importante tasa de reprobación tienen una alta probabilidad de deserción al no contar con apoyo financiero. En forma general los antecedentes académicos de los alumnos son muy importantes al momento de tomar la decisión de desertar es por esto que es relevante identificar si se trata de una asignatura o una condición de las evaluaciones.

B.1.2.2. Evaluación de la clasificación

Aunque ninguno de los algoritmos utilizados permite una clasificación sin errores, se identifica que la mejor clasificación se ha obtenido directamente con el algoritmo ID3 en la aplicación Weka, la cual entrega además una gran cantidad de reglas que permiten identificar la deserción con un menor cantidad de mal clasificados (ver Figura B.11)

| ID3 | | | | C4.5 | | | | CART | | |
|-----|-----|-----|--|------|-----|-----|--|------|-----|-----|
| | NO | SI | | | NO | SI | | NO | SI | |
| NO | 427 | 47 | | NO | 436 | 53 | | NO | 427 | 62 |
| SI | 70 | 244 | | SI | 105 | 250 | | SI | 106 | 249 |

Figura B.11: Matrices de confusión de la clasificación de Auditoría (Diurno)

Aunque c4.5 entrega una mejor clasificación que CART, este último entrega mejores resultados que C4.5, comparando la calidad con el mejor de los casos que corresponde al algoritmo ID3, dado que la profundidad y la cantidad de atributos involucrado en las reglas encontradas, permite visualizar conocimiento interesante para la carrera.

B.1.3. Minería de datos en las evaluaciones de los alumnos de Auditoría.

Para evaluar la carrera de auditoría se ha extraído un conjunto de datos equivalente al total de las notas de asignaturas que corresponde a 22.588 registros de todos los alumnos que han sido considerados en la evaluación de la deserción. La evaluación realizada se ha hecho en primer lugar con el total de las asignaturas y debido a la importancia de la reprobación de asignaturas, de forma independiente se ha evaluado las asignaturas reprobadas solamente. Al igual que en los casos anteriores se identifican claramente que existen asignaturas más representativas para los desertores, las cuales son en este caso particular las pertenecientes a los dos primeros semestres. ver Figura B.12.

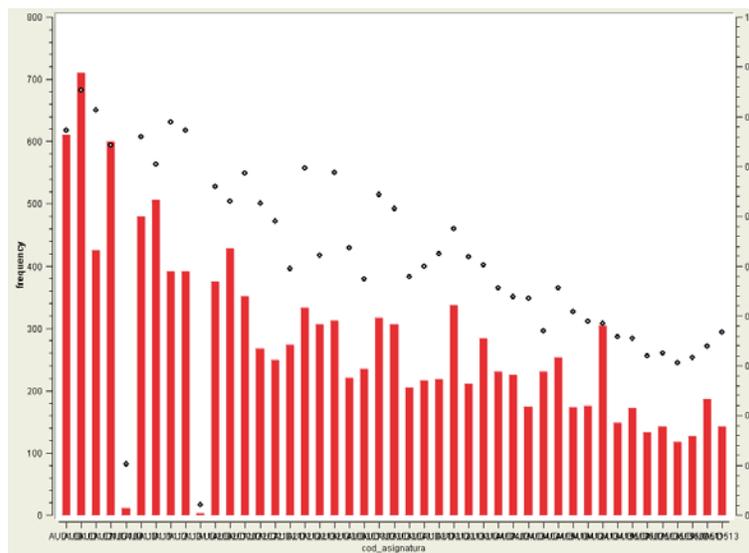


Figura B.12: Cantidad de desertores por asignaturas

Por otra parte también se distinguen académicos que tienen una mayor cantidad de desertores que otros; como se puede ver en la Figura B.13, al menos dos profesores influyen fuertemente la deserción de los alumnos.

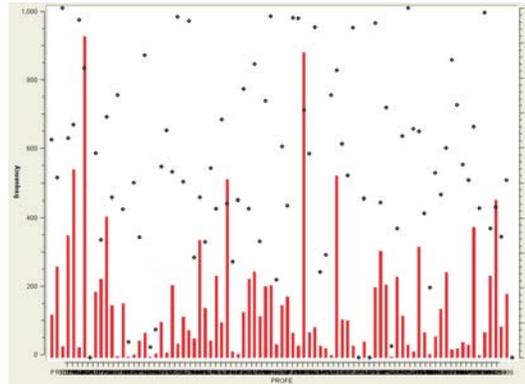


Figura B.13: Cantidad de desertores por profesor

La tasa de aprobación juega un rol importante en las asignaturas de los primeros años donde a menor tasa de aprobación de la asignatura dictada la probabilidad de deserción sube, como se muestra en la Figura B.14. Junto a lo anterior las asignaturas de alumnos de últimos años presentan una posibilidad de deserción con las asignaturas que tienen un gran cantidad de alumnos, aunque es una regla débil es interesante evaluar a los cursos grandes para identificar las causas independientes de esta deserción.

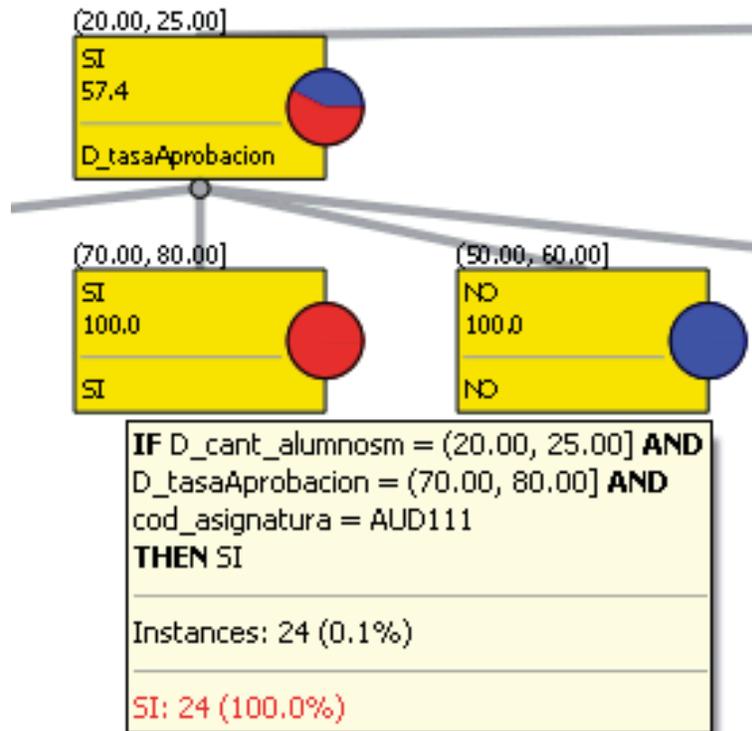


Figura B.14: Clasificación alumnos desertores y la tasa de aprobación.

Ademas,es importante la nota final obtenida en las asignaturas del primer semestre de la carrera, debido

a que los alumnos con muy baja evaluación en las asignaturas desertan en esta etapa inicial de estudios.

Al evaluar los alumnos desertores por medio del algoritmo C4.5 se identifica que el atributo más importante para describir la deserción de los estudiantes de la carrera de Auditoría es el profesor; la condición general que se identifica se encuentra en la cantidad de alumnos del curso, donde para distintos cursos dependiendo de la cantidad de alumnos determina el abandono, lo que hace muy importante regular los cupos de cada asignatura o evaluar la factibilidad de levantar una mayor cantidad de paralelos, ver Figura B.15.

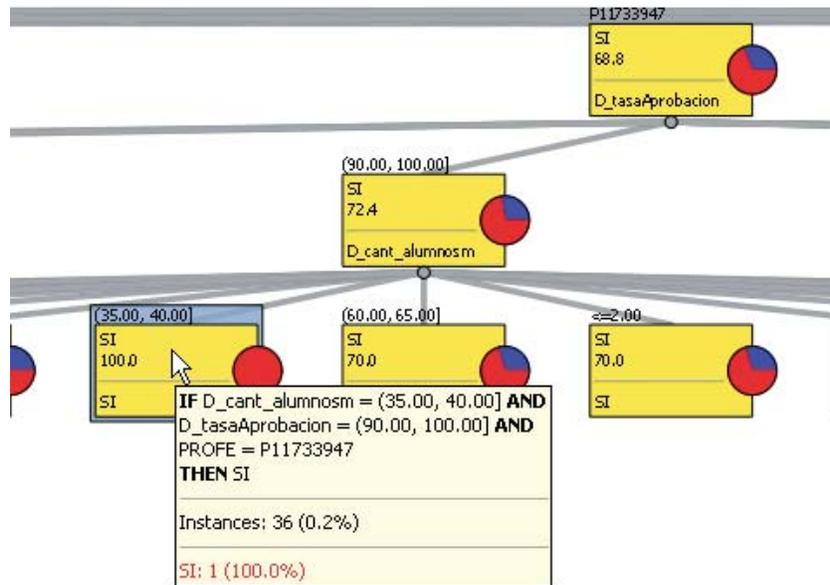


Figura B.15: Clasificación notas alumnos desertores con algoritmo C4.5.

En forma general la tasa de aprobación y la cantidad de alumnos juega un rol fundamental al momento de evaluar la deserción de los alumnos de la carrera de Auditoría Diurna en que implica que cursos muy grades pueden causar la deserción.

Apéndice C

Anexo 3

C.1. Análisis Carrera de Enfermería

La carrera de Enfermería es una de las carreras del área de la salud, que cuenta con 40 asignaturas y una estructura de carácter mixto semestral-anual. Por otra parte, las carreras del área de la salud en su mayoría se caracterizan por no tener una gran cantidad de desertores, en el caso de la carrera de enfermería es una de las que más desertores tienen en su área, lo cual hace interesante su evaluación. El conjunto de datos evaluado cuenta con 579 registros, correspondiente a los alumnos ingresado entre el 2004 y el 2010.

Al igual que todas las evaluaciones anteriores se ha utilizado el conjunto de datos final que incluye las tasa de avance por semestres y de avance por asignaturas que permita identificar la causa de deserción de los alumnos y la similaridad que pueda existir entre las carreras.

C.1.1. Análisis Exploratorio

El primer acercamiento a los datos de la carrera, posterior a la discretización, permite identificar algunas reglas generales, como es que la deserción principalmente ocurre en el primer semestre, aunque esto es muy relativo dado que se trata de una carrera que en muchos alumnos no tiene segundo semestre, al tratarse de una asignatura anual donde sólo hay una apertura de semestre único en el año.

Por otra parte se identifica que los alumnos desertores principalmente son aquellos que abandonan la carrera siendo su estado final " Espera de cancelacion de derecho básico de matrícula (A15)", factor diferenciador de otras carreras previamente evaluadas en que la deserción principalmente se fundamenta en la eliminación académica; la cantidad de alumnos que abandonan la carrera es mucho mayor en otros casos como se puede ver en la Figura C.1.

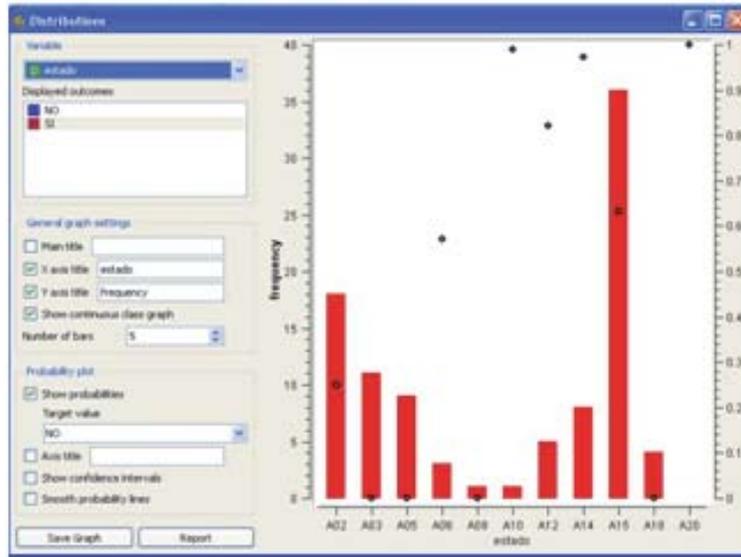


Figura C.1: Cantidad de alumnos desertores por estados

Un punto importante se muestra con la tasa de avance total de la carrera donde los alumnos desertores tienen una tasa menor al 30% del avance total de la carrera, lo que implica que se encuentran en sus primeros años de carrera (ver Figura C.2).

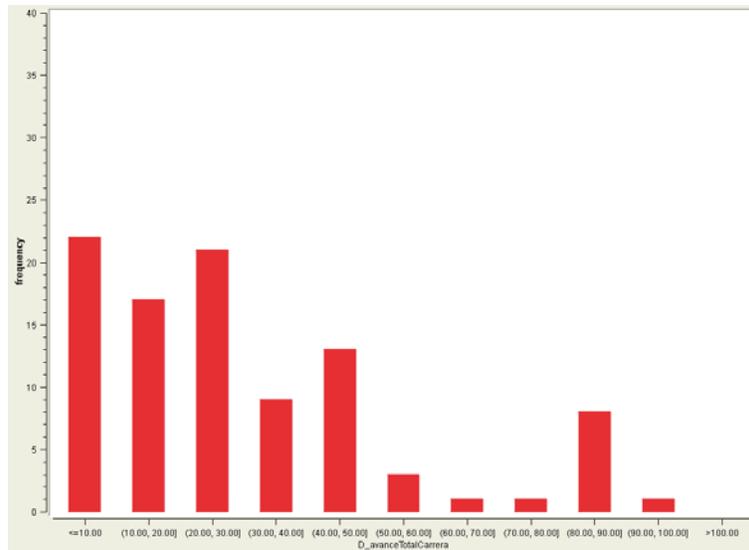


Figura C.2: Cantidad de alumnos desertores por avance total carrera

La misma condición se presenta al evaluar la cantidad de semestres cursados por los alumnos, donde la gran mayoría han cursado menos de cinco semestres. Al visualizar la distribución de los datos aprecia que la deserción de los alumnos se encuentra definida por la tasa de aprobación y el avance del alumno en la carrera y muy levemente en el porcentaje de crédito asignado, (ver Figura C.3)

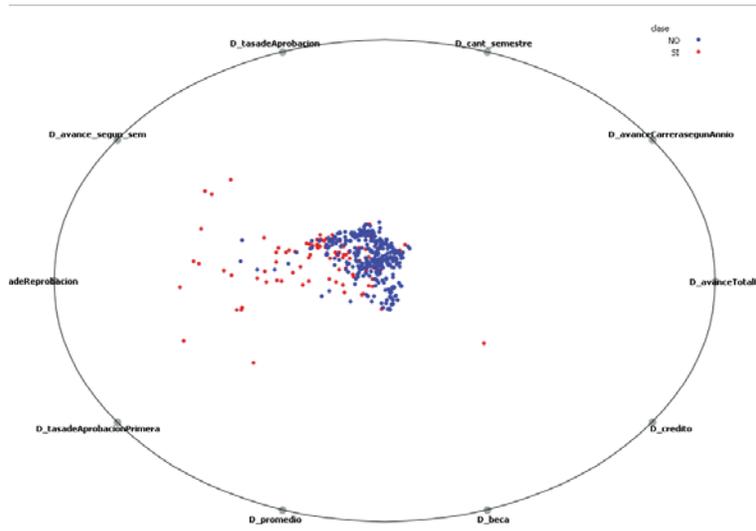


Figura C.3: Distribución de datos por atributos relevantes

C.1.2. Evaluación de la Carrera con Algoritmos

C.1.2.1. Clasificación

Al realizar la clasificación de los datos, en primer lugar se ha realizado una ranking de los atributos más importantes, utilizando el algoritmo de ganancia de información, lo que ha permitido obtener 14 atributos más representativos, ver en la Figura C.4

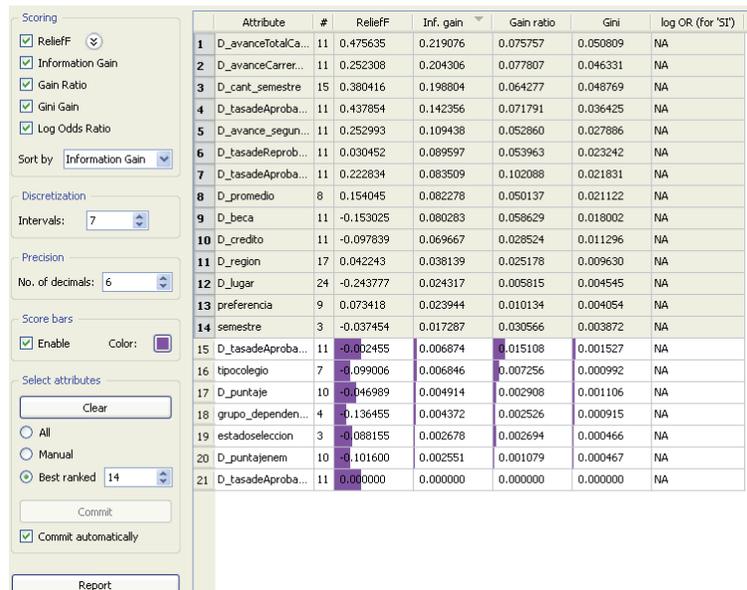


Figura C.4: Ranking atributos relevantes para la clasificación

Con estos atributos se han aplicado algoritmos de clasificación con árboles de decisión utilizando los algoritmos ID3, CART y C4.5; con estos algoritmos se identifican reglas interesantes en la carrera, pues

pese a que la deserción ocurre principalmente por causas académicas (como se muestra en la Figura C.5), el avance total de la carrera es muy importante a la hora de evaluar la deserción, siendo el atributo principal de la clasificación.

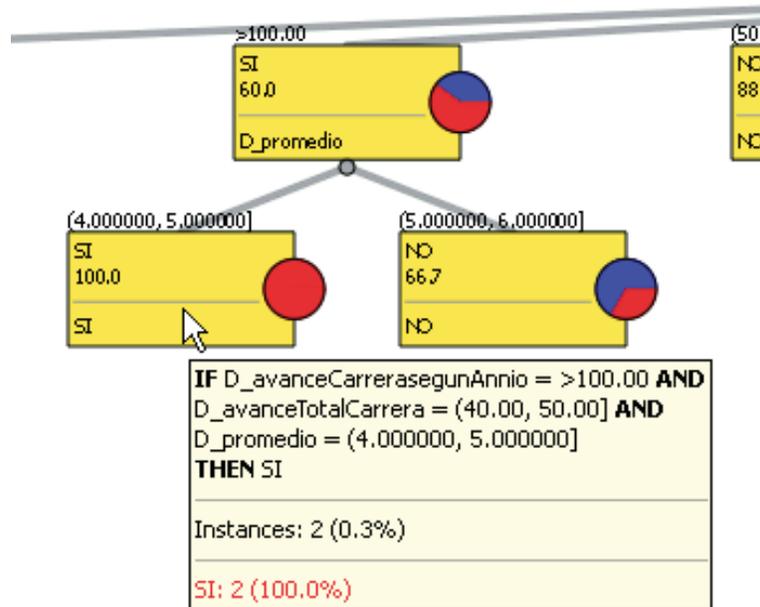


Figura C.5: Clasificación alumnos desertores y el avance total Carrera

En los primeros años la tasa de avance en la carrera juega un rol importante dado que pese a contar con una buena tasa de aprobación los alumnos jóvenes prefieren desertar antes de continuar, como se muestra en la Figura C.6, donde se identifica que a menor cantidad de semestres mayor es la probabilidad de deserción por parte de los alumnos, en una tasa de avance de un 40 % a 50 %.

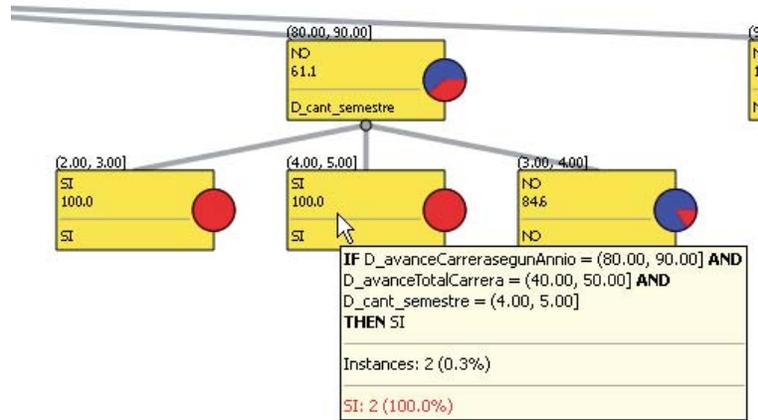


Figura C.6: Clasificación alumnos desertores y la tasa de aprobación

Por otra parte la tasa de aprobación particular de los alumnos debe ser una preocupación de la carrera, debido que en iguales condiciones los alumnos con menor tasa de aprobación deciden desertar, como se muestra en la Figura C.7. Pese a que son pocos los alumnos desertores en la carrera en la mayoría de los casos juega un rol importante los antecedentes académicos, por lo que se hace interesante la evaluación de las asignaturas.

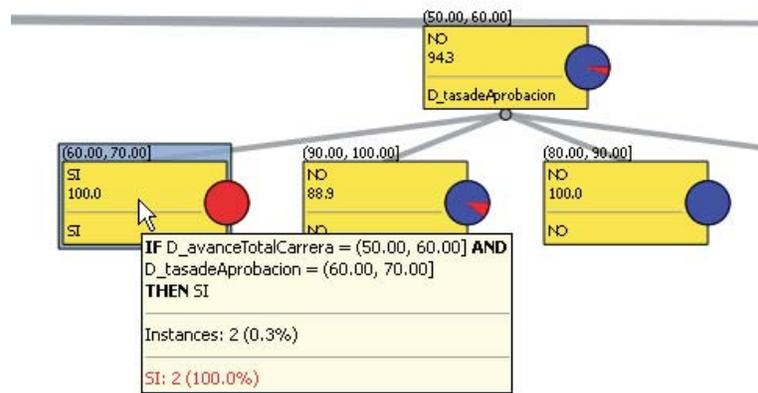


Figura C.7: Clasificación Alumnos desertores y la tasa de aprobación

Al evaluar la deserción utilizando el algoritmo C4.5, sólo se ratifica la importancia que tiene la tasa de avance total en la carrera para los alumnos de Enfermería.

El algoritmo ID3 muestra nuevas reglas basadas en los antecedentes de ingreso donde identifica en que los alumnos que postulan en tercer o cuarto lugar de preferencia pueden ser potenciales desertores de la carrera, pese a que tienen una tasa de avance según semestre de un 100%. Dado que se encuentran en los primeros años de la carrera, se hace importante el seguimiento de ellos, debido a que puede tratarse alumnos con dudas vocacionales o de entendimiento de la carrera, como se muestra en la Figura C.8 que los alumnos

desertan al tener tasa de avance total de la carrera menores a un 20% y tasa de avance según año de ingreso menores al 50%.

```

| D_avanceCarrerasegunAnio = (70.00, 80.00]: null
| D_avanceCarrerasegunAnio = (80.00, 90.00]: null
| D_avanceCarrerasegunAnio = (90.00, 100.00]: null
| D_avanceCarrerasegunAnio = >100.00: SI
D_avanceTotalCarrera = (10.00, 20.00]
| D_avance_segun_sem = <=10.00: SI
| D_avance_segun_sem = (10.00, 20.00]: SI
| D_avance_segun_sem = (20.00, 30.00]: null
| D_avance_segun_sem = (30.00, 40.00]
| | D_avanceCarrerasegunAnio = <=10.00: null
| | D_avanceCarrerasegunAnio = (10.00, 20.00]: null
| | D_avanceCarrerasegunAnio = (20.00, 30.00]: SI
| | D_avanceCarrerasegunAnio = (30.00, 40.00]: SI
| | D_avanceCarrerasegunAnio = (40.00, 50.00]: NO
| | D_avanceCarrerasegunAnio = (50.00, 60.00]: null
| | D_avanceCarrerasegunAnio = (60.00, 70.00]: null
| | D_avanceCarrerasegunAnio = (70.00, 80.00]: null
| | D_avanceCarrerasegunAnio = (80.00, 90.00]: null
| | D_avanceCarrerasegunAnio = (90.00, 100.00]: null
| | D_avanceCarrerasegunAnio = >100.00: null
| D_avance_segun_sem = (40.00, 50.00]: SI
| D_avance_segun_sem = (50.00, 60.00]: SI
| D_avance_segun_sem = (60.00, 70.00]: NO
| D_avance_segun_sem = (70.00, 80.00]: NO
| D_avance_segun_sem = (80.00, 90.00]: null
| D_avance_segun_sem = (90.00, 100.00]: SI
| D_avance_segun_sem = >100.00
| | preferencia = 0: NO
    
```

Figura C.8: Clasificación Alumnos desertores utilizando algoritmo ID3 en WEKA

C.1.2.2. Evaluación de la clasificación.

Al evaluar la bondad de la clasificación y la calidad de las reglas de los tres algoritmos aplicados se desprende que el mejor de ellos es el algoritmos ID3 que identifica una gran cantidad de reglas con una menor tasa de errores, ver Figuras C.9. Pese a que el algoritmo CART es el peor calificado no está muy distante, pues aunque cuenta con una gran cantidad de errores de clasificación, no obstante tiene la gran ventaja de la cantidad de reglas interesantes que otorga.

| ID3 | | | C4.5 | | | CART | | |
|-----|-----|----|------|-----|----|------|-----|----|
| | NO | SI | | NO | SI | | NO | SI |
| NO | 429 | 29 | NO | 466 | 17 | NO | 498 | 15 |
| SI | 35 | 44 | SI | 40 | 56 | SI | 44 | 52 |

Figura C.9: Matrices de clasificación de la carrera de Enfermería

C.1.2.3. Minería de datos en las evaluaciones de los alumnos de Enfermería.

Para evaluar las calificaciones de la carrera de enfermería se ha utilizado el mismo grupo de datos que corresponde a los utilizados con anterioridad; en este caso se han identificado 12007 ítem de notas donde se encuentran mayoritariamente alumnos con continuidad de estudios, ver en la Figura C.10.

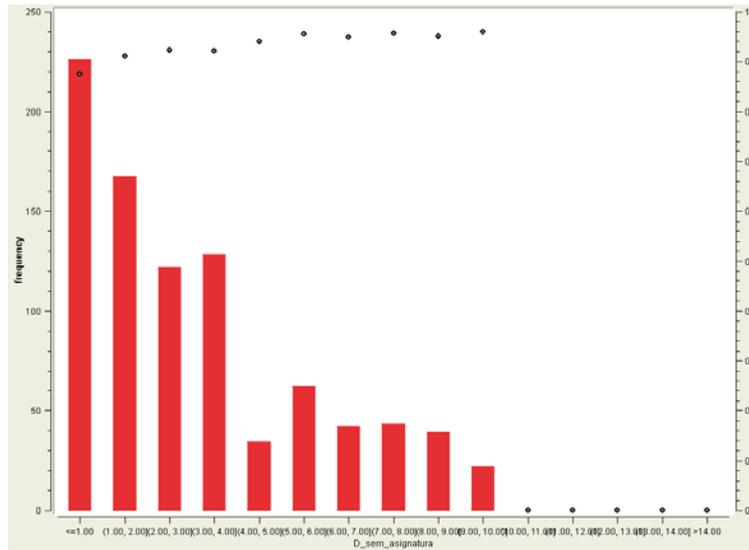


Figura C.10: Cantidad de alumnos desertores por semestre asignatura

Principalmente la deserción se encuentra afectada por la cantidad de alumnos que tienen las asignaturas, como se puede ver en la siguiente Figura C.11, las asignaturas cursadas por los alumnos desertores tienen más de 75 alumnos

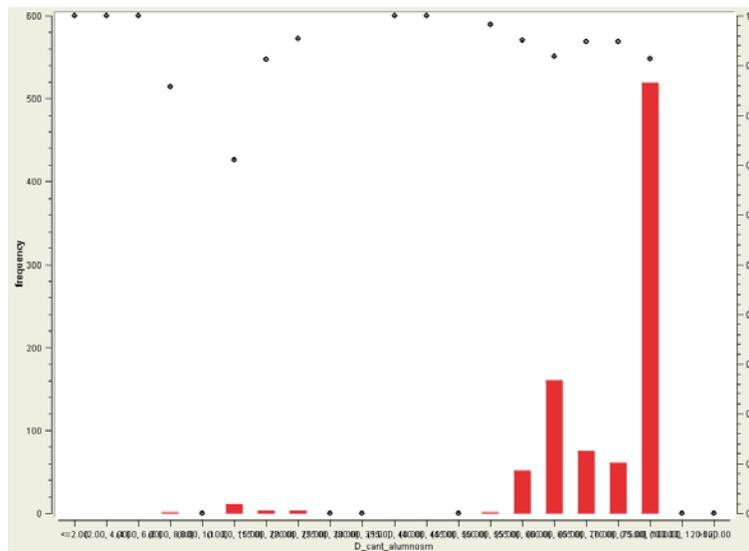


Figura C.11: Cantidad de alumnos desertores por cantidad de alumnos en una asignatura

Por otro lado, a diferencia de los casos anteriores, donde la deserción evaluada por medio de las notas esta fuertemente representada en algunos casos por los profesores, en el caso de la carrera de Enfermería no hay un impacto directo de los académicos y menor las asignaturas.

Además, debido a la baja cantidad de desertores que hay se obtuvo un subgrupo para identificar los desertores en base a las asignaturas reprobadas, donde se visualiza que la deserción está principalmente determinada por la tasa de aprobación y la cantidad de alumnos que se encuentran cursando la asignatura,

lo cual hace muy interesante validar la factibilidad de reducir los cursos o la forma en la cual se dictan las asignaturas.

Una regla interesante que se aprecia en algunos casos determina la importancias que tienen el promedio de notas versus la tasa de aprobación y cantidad de alumnos de un curso, como por ejemplo la Figura C.12, donde los alumnos desertores están descritos por aquellos que tienen cursos pequeños una nota final menor a un 3.0 y una tasa de aprobación de la asignatura mayor a un 80 %, lo anterior hace muy interesante identificar la situación particular de cada alumnos previo a la toma de asignaturas, de tal forma de darle una mayor probabilidad de éxito en la aprobación del curso.

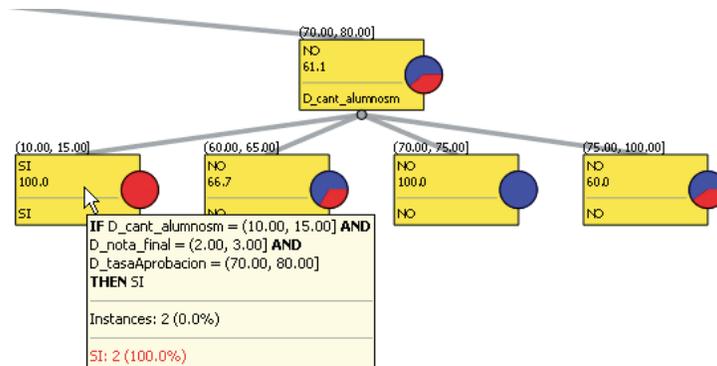


Figura C.12: Clasificación con CART y la cantidad de alumnos en catedra

En resumen es de gran relevancia la relación que existe entre la tasa de aprobación de la asignatura y la nota obtenida por el alumno en especial en los primeros años; más que esperar a reprobar es la decisión de desertar, lo que puede ocurrir debido a que los alumnos de estas carreras, suelen pertenecer a segmentos académicos altos y más que una necesidad se puede deber a una frustración personal que lo lleva a desertar.