



UNIVERSIDAD DE CHILE

FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS

DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**DISEÑO DE SISTEMA DE GESTIÓN PARA LA RETENCIÓN
DE ALUMNOS PARA ASIGNACIÓN OPORTUNA DE
BENEFICIOS ESTUDIANTILES**

PROYECTO DE GRADO PARA OPTAR AL GRADO DE MAGÍSTER EN
INGENIERÍA DE NEGOCIOS CON TECNOLOGÍAS DE LA INFORMACIÓN

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

MANUEL ALFREDO FONSECA CRUZ

PROFESOR GUÍA:
EZEQUIEL MUÑOZ KRSULOVIC

PROFESOR CO-GUÍA:
CECILIA SAINT-PIERRE CORTÉS

MIEMBRO DE LA COMISIÓN:
EDUARDO OLGUÍN MACAYA

SANTIAGO DE CHILE
2019

Resumen Ejecutivo

La Universidad de Chile se ha caracterizado por ser históricamente una universidad al servicio de la sociedad, lo cual a través del tiempo se ha ido materializando en políticas internas para acompañar a sus alumnos en un ambiente de equidad e inclusión. Fiel reflejo es la promulgación en 2014 de la política interna de Equidad e Inclusión, la cual garantiza el acompañamiento de estudiantes en el acceso, permanencia y egreso oportuno de sus alumnos. Parte importante de esta política corresponde a que la Universidad de Chile debe realizar esfuerzos por retener a sus alumnos, respondiendo a un cuerpo estudiantil cada vez más diverso en términos económicos, sociales y culturales. Pese a los esfuerzos, aún continua existiendo una marcada heterogeneidad en la retención de alumnos de primer año en sus distintas facultades, siendo la más alta en la Facultad de Ciencias Físicas y Matemáticas, con un 95 % de retención, y la más baja en Bachillerato, con sólo un 53 % de retención.

Debido a estos antecedentes, es que se propone en este proyecto de tesis generar una nueva capacidad en la Universidad de Chile que permita detectar a alumnos en riesgo de deserción para así ofrecer beneficios estudiantiles de forma proactiva.

La metodología del proyecto corresponde a la metodología de Ingeniería de Negocios y Diseño de Servicios, mientras que la metodología de minería de datos utilizada corresponde a CRISP-DM.

Los modelos para predecir la deserción utilizados fueron de Árboles de Decisión, Random Forest y Regresión Logística. Para entrenar los modelos se utilizaron datos de alumnos nuevos que ingresaron a la Universidad de Chile en 2017 y que rindieron la PSU. El tipo de deserción estudiada correspondió a la deserción institucional, es decir, alumnos que abandonan de forma voluntaria o involuntaria la Universidad de Chile. El modelo que presentó los mejores resultados fue el de Regresión Logística, el cual de haber sido aplicado para alumnos de primer año de 2017, hubiese detectado la deserción de 311 alumnos, o de forma equivalente, el 41 % de los alumnos desertores. Además, se validó que el modelo fuera generalizable, es decir, se probó que el modelo aplica al universo de alumnos de primer año que ingresa a la Universidad de Chile, lo cual fue realizado a través de un análisis retrospectivo con datos de alumnos nuevos del cohorte del año anterior (2016).

La Arquitectura Tecnológica propuesta se compone de un software de minería de datos, una herramienta de visualización y un Data Mart de Alumnos. Este último se propone para contar con datos de alumnos de diferentes fuentes de información bajo un único estándar de calidad, para posteriormente ser utilizados en los modelos de minería de datos.

Por último, la evaluación económica del proyecto, indicó que el proyecto debe ser realizado, obteniendo un Δ VAN de \$ 1.520 millones con respecto a la situación actual optimizada, considerando una duración de tres años de proyecto.

*A mi familia, Virginia, Manuel, Tamara, Valeria y Martín
que todos estos años me han acompañado
y me han brindado su inmenso amor.*

Agradecimientos

A mis padres por haberme acompañado durante toda mi vida tanto en lo académico como personal, entregándome su infinito amor y consejo en todo momento. Por fin se termina un ciclo! Los amo con todo mi corazón!

A mis hermanas por la paciencia y el tiempo dedicado a escucharme.

A mis abuelos, quienes estuvieron atentos al desarrollo de mi proyecto y por haberme entregado el gran regalo de conocer a Dios, quien me guía en los diferentes aspectos de mi vida.

A mi polola por haberme acompañado en todo este proceso, por todo su cariño y soporte, brindándome ternura y amor en todo momento. Eres muy especial para mí!!

Al profesor Ezequiel y Cecilia por responder a mis dudas, dedicar su tiempo, paciencia y orientarme en este arduo trabajo. Y a Laura y Ana María, quienes me han apoyado durante todo el proceso de mi paso por el MBE. Gracias!

Al equipo de la Dirección de Bienestar Estudiantil, Rodrigo Carmona, Rafael Díaz y Pamela Díaz-Romero, quienes sin ser de la misma área de trabajo, creyeron en mí, brindándome todo su apoyo y guía en temáticas de bienestar estudiantil.

A Paulina Tapia por darme todas facilidades y brindarme el tiempo necesario para lograr sacar adelante mi tesis. Y a mis compañeros de trabajo, Jorge y Ricardo, que me apoyaron y ayudaron en este proceso. Gracias!!

A mis amigos, Negro, Kevin, Lillo, Robert, Pipe y Hevia quienes siempre estuvieron a mi lado pese a que tuve que faltar a muchas juntas y carretes. Aún así me entregaron su comprensión durante todo este tiempo y me apoyaron con sus consejos. Ahora se vienen todos los panoramas!

A todos aquellos que creyeron en mí, brindaron su apoyo y me escucharon largas conversaciones acerca de variados temas relacionados con mi tesis.

Por sobre todo agradecer a Dios! Por darme el valor, coraje y fuerza para mantenerme en este desafiante proceso.

Tabla de Contenido

Resumen Ejecutivo	I
Dedicatoria	II
Agradecimientos	III
Tabla de Contenido	IV
Índice de Figuras	VIII
Índice de Tablas	X
1. Introducción y Contexto de la Universidad de Chile	1
1.1. Descripción general Universidad de Chile	1
1.1.1. Estructura Organizacional	2
1.1.2. Vicerrectoría de Asuntos Estudiantiles y Comunitarios	4
1.1.2.1. Dirección de Bienestar Estudiantil (DBE)	4
1.1.2.2. Estructura Organizacional DBE	4
1.2. Antecedentes del problema en la educación superior	6
2. Proyecto	9
2.1. Antecedentes y justificación	9
2.2. Objetivo General	10
2.3. Objetivos Específicos	10
2.4. Resultados Esperados	11
2.5. Alcance	11
2.6. Criterios de Éxito	12
2.7. Riesgos Potenciales	13
3. Marco Teórico	14
3.1. Metodología de Ingeniería de Negocios y Diseño de Servicios	14
3.1.1. Patrón de Negocio	15
3.1.2. Patrones de Arquitectura y Procesos de Negocio	18
3.2. Metodología de Minería de Datos CRISP-DM	21
3.3. Método de selección de variables <i>Forward Selection and Backward Elimination</i>	23
3.4. Modelos de Minería de Datos	24

Tabla de Contenido

3.4.1.	Modelos Supervisados	24
3.4.1.1.	Regresión Logística	25
3.4.1.2.	Árboles de decisión	26
3.4.2.	Modelos No Supervisados	30
3.4.2.1.	Análisis de Cluster	30
3.4.2.2.	Random Forest	33
3.4.3.	Validación Cruzada	34
3.4.4.	Matriz de Confusión	35
3.4.5.	ROC Curve y AUC	37
3.5.	Concepto de Deserción	38
3.5.1.	Estudios Relacionados	39
3.5.1.1.	Modelos de Deserción Estudiantil	39
3.5.1.2.	Un modelo analítico para la predicción del rendimiento académico de estudiantes de ingeniería	41
3.5.1.3.	El impacto del tipo de financiamientos sobre la probabilidad de retención de primer año en la educación superior - El caso de la Universidad de Chile.	42
3.5.1.4.	Impacto de los distintos tipos de ayuda financiera gubernamental e institucional en la perseverancia de los estudiantes de la Universidad de Chile.	43
3.5.1.5.	Determinantes de deserción en la educación superior chilena, con énfasis en efecto de becas y créditos.	44
3.5.1.6.	Revisión estudio “ <i>Competing on Analytics</i> ”	45
4.	Planteamiento Estratégico	48
4.1.	Balanced Scorecard	48
4.2.	Posicionamiento Estratégico Universidad de Chile	50
4.3.	Política de Equidad e Inclusión Estudiantil	51
4.3.1.	Modelo de Desarrollo Integral del Estudiante	51
4.3.2.	Ámbitos de acción del Modelo de Desarrollo Integral del Estudiante	53
4.4.	Visión, Misión y Objetivos Estratégicos DBE	54
4.4.1.	Visión	54
4.4.2.	Misión	54
4.4.3.	Objetivos Estratégicos	55
4.5.	Análisis PEST DBE	55
4.5.1.	Político y Económico	55
4.5.1.1.	Gratuidad	56
4.5.1.2.	Ley de Universidades Estatales	57
4.5.1.3.	Sistema de Financiamiento Solidario	58
4.6.	Modelo de Negocios DBE	59
4.6.1.	Propuesta de Valor	59
4.6.2.	Recursos Clave	59
4.6.3.	Procesos Clave	60
4.6.4.	Fórmula de utilidades	60

Tabla de Contenido

5. Situación Actual	61
5.1. Patrón de Negocio a utilizar	61
5.2. Arquitectura de negocio Universidad de Chile	63
5.3. Diagnóstico Arquitectura Multinivel	71
5.4. Diagnóstico Situación Actual	73
5.5. Cuantificación del Problema u Oportunidad	75
6. Propuesta de Diseño de Procesos	77
6.1. Direcciones de Cambio y Alcance	77
6.2. Arquitectura de Procesos <i>to be</i>	82
6.3. Diseño de Proceso “Analizar Comportamiento y Segmentación de Estudiantes”	86
6.4. Diseño de Lógica de Negocios	87
6.4.1. Comprensión y preparación de datos	89
6.4.1.1. Fuentes de Datos	89
6.4.1.2. Descripción de Variables	91
6.4.1.3. Tratamiento de datos faltantes	93
6.4.1.4. Construcción de nuevas variables	96
6.4.1.5. Distribución de variables	98
6.4.1.6. Relación entre variables	99
6.4.1.7. Selección de variables	103
6.4.2. Análisis de Cluster	104
6.4.2.1. Resultados clusters con variables Reales	105
6.4.2.2. Resultados clusters con variables Categóricas	106
6.4.3. Procedimiento Análisis Predictivos	107
6.4.4. Resultados y Análisis de Modelos Predictivos	110
6.4.5. Modelo seleccionado	113
7. Propuesta de apoyo tecnológico	114
7.1. Especificación de requerimientos	115
7.1.1. Requerimientos funcionales	115
7.1.2. Requerimientos no funcionales	116
7.2. Arquitectura Tecnológica	116
7.3. Diseño de aplicación	117
7.3.1. Casos de Uso	117
7.3.2. Diagrama de arquitectura del sistema	118
7.3.3. Diagrama de despliegue	120
8. Gestión del Cambio	121
8.1. Modelo Integral de Liderazgo y Gestión del Cambio	121
8.2. Caracterización del cambio	126
8.3. Factores críticos de éxito	126
8.4. Plan de gestión del cambio	127
9. Evaluación del Proyecto	128
9.1. Análisis Retrospectivo	128

Tabla de Contenido

9.1.1. Presentación de Datos	128
9.1.2. Resultados obtenidos	129
9.1.3. Validación Análisis Retrospectivo	130
9.2. Definición de Beneficios y Costos	132
9.3. Flujo de Caja	134
9.4. Análisis de Sensibilidad	138
10. Conclusiones	139
10.1. Lecciones Aprendidas	141
10.2. Trabajos Futuros	142
Bibliografía	145
A. Marco Teórico	146
A.1. Parámetros árboles de decisión	146
A.2. Parámetros Random Forest	147
B. Lógica de Negocio	148
B.1. Detalle atributos Guía Matrícula e Índice de Vulnerabilidad Escolar (IVE)	148
B.2. Créditos Universitarios	153
B.3. Becas de Arancel Internas y Externas	154
B.4. Becas de Mantenimiento Internas y Externas	155
B.5. Resultados proceso PSU	156
B.6. Creación de Nuevas Variables	156
B.7. Distribución de Variables	166
B.8. Proceso clusterización con variables reales	169
B.8.1. X Means	169
B.8.2. DBSCAN	171
B.9. Selección epsilon y min points DBSCAN	175
B.10. Proceso clusterización con variables categóricas	175
B.10.1. X Means	175
B.10.2. DBSCAN	176
B.11. Puntaje Ranking	178
B.12. Índice de Vulnerabilidad Escolar (IVE)	179
B.13. Coeficientes Regresión Logística	181
C. Evaluación Económica	183
C.1. Análisis Retrospectivo - Gratuidad	183
C.2. Análisis Retrospectivo - Beca Vocación Profesor	183
C.3. Análisis Retrospectivo - Tramo Socioeconómico	184
C.4. Análisis Retrospectivo - Carrera	185
C.5. Análisis de Sensibilidad	186

Índice de figuras

1.1. Organigrama Universidad de Chile	3
1.2. Organigrama DBE	5
1.3. Evolución tasas de retención Instituciones de Educación Superior. Fuente Elaboración Propia, a partir de Informe de Retención de primer año de pregrado SIES Cohortes 2012- 2016.	6
1.4. Retención 1er año Cohorte 2016. Fuente Elaboración Propia, a partir de datos de retención de primer año de pregrado SIES 2016	7
1.5. Tasa de retención al primer año por unidad académica. Fuente: Anuario 2017 Universidad de Chile	8
3.1. Metodología de diseño de Servicios. Fuente Bussiness Engineering and Service Design.	15
3.2. Estructura básica de organización de servicios. Fuente: Business Engineering and Service Design.	16
3.3. Patrón de Negocio 1. Fuente: Business Engineering and Service Design.	17
3.4. Macroprocesos. Fuente: Business Engineering and Service Design.	19
3.5. Arquitectura Multinivel Fuente: Business Engineering and Service Design.	20
3.6. Metodología CRISP-DM. Fuente: Smart Vision Europe.	22
3.7. Puntos Core, Border y Ruido DBSCAN. Fuente: Elaboración propia.	32
3.8. Validación Cruzada (k=4). Fuente: Elaboración propia.	35
3.9. ROC Curve. Fuente: Elaboración propia.	37
4.1. Modelo de Desarrollo Integral del Estudiante. Fuente: Política de Equidad e Inclusión Estudiantil 2014.	52
5.1. Patrón de Negocio 1 - Universidad de Chile. Fuente: Elaboración propia.	62
5.2. Arquitectura Multinivel, Universidad de Chile. Fuente: Elaboración propia.	65
5.3. Gestión Centralizada, Universidad de Chile. Fuente: Elaboración propia.	66
5.4. Administración de la Relación con el Estudiante, Universidad de Chile. Fuente: Elaboración propia.	68
5.5. Atención Basada en el Conocimiento del Estudiante, Universidad de Chile. Fuente: Elaboración propia.	70
5.6. Analizar comportamiento y segmentación de estudiante, Universidad de Chile. Fuente: Elaboración propia.	71
6.1. Rediseño Administración de la Relación con el Estudiante, Universidad de Chile. Fuente: Elaboración propia.	84

6.2. Rediseño Atención Basada en el Conocimiento del Estudiante, Universidad de Chile. Fuente: Elaboración propia.	86
6.3. Rediseño Proceso Analizar Comportamiento y Segmentación de Estudiantes Fuente: Elaboración propia.	87
6.4. CAE y Deserción. Fuente: Elaboración propia.	101
6.5. Carrera Anterior No Terminada y Deserción. Fuente: Elaboración propia.	101
6.6. CAE-FSCU y Deserción. Fuente: Elaboración propia.	102
6.7. DBSCAN. Fuente: Elaboración propia.	105
6.8. DBSCAN - Desertores. Fuente: Elaboración propia.	106
6.9. ROC y AUC ID3. Fuente: Elaboración propia.	111
6.10. ROC y AUC CART. Fuente: Elaboración propia.	111
6.11. ROC y AUC CHAID. Fuente: Elaboración propia.	112
6.12. ROC y AUC Random Forest. Fuente: Elaboración propia.	112
6.13. ROC y AUC Regresión Logística. Fuente: Elaboración propia.	112
7.1. Casos de Uso. Fuente: Elaboración propia.	118
7.2. Arquitectura Tecnológica. Fuente: Elaboración propia.	119
7.3. Diagrama de despliegue. Fuente: Elaboración propia.	120
9.1. Desvío absoluto predicción vs realidad. Fuente: Elaboración propia.	132
B.1. Histograma PSU Matemáticas. Fuente: Elaboración propia.	166
B.2. Histograma PSU Lenguaje. Fuente: Elaboración propia.	166
B.3. Histograma PSU Ranking. Fuente: Elaboración propia.	167
B.4. Histograma IVE (1). Fuente: Elaboración propia.	167
B.5. Histograma raíz cúbica IVE (1). Fuente: Elaboración propia.	167
B.6. Histograma Porcentaje Cobertura Arancel. Fuente: Elaboración propia.	168
B.7. Histograma Porcentaje Cobertura Crédito. Fuente: Elaboración propia.	168
B.8. Cluster X Mean	170
B.9. Cluster X Mean - Desertores. Fuente: Elaboración propia.	170
B.10. PSU Matemática, Lenguaje Final e IVE. Fuente: Elaboración propia.	171
B.11. DBSCAN. Fuente: Elaboración propia.	173
B.12. DBSCAN - Desertores. Fuente: Elaboración propia.	173
B.13. DBSCAN - PSU Matemática, PSU Lenguaje e IVE. Fuente: Elaboración propia.	174

Índice de tablas

2.1. Riesgos Potenciales	13
3.1. Matriz de Confusión	36
3.2. Tipos de Deserción	38
4.1. Ámbitos de acción MDIE	53
6.1. Dirección de cambio: Estructura de empresa y mercado	77
6.2. Dirección de cambio: Anticipación	79
6.3. Dirección de cambio: Coordinación	79
6.4. Dirección de cambio: Prácticas de Trabajo	80
6.5. Dirección de cambio: Integración de Procesos Conexos	81
6.6. Dirección de cambio: Mantención de Estado	82
6.7. Descripción de Instituciones	89
6.8. Extracción de Datos	90
6.9. Atributos Guía Matrícula e Índice de Vulnerabilidad Escolar (IVE)	91
6.10. Nuevas variables PSU	94
6.11. Nuevas variables “IVE (1) e IVE (2)”	95
6.12. Correlación con variable deserción	100
6.13. Correlación entre variables predictoras	102
6.14. Variables seleccionadas por método <i>Forward and Backward Elimination</i>	104
6.15. Resultados modelos seleccionados	110
7.1. Requerimientos no funcionales	116
9.1. Resultados análisis Retrospectivo	129
9.2. Análisis Retrospectivo - CAE. Fuente: Elaboración propia.	130
9.3. Análisis Retrospectivo - Índice de Vulnerabilidad Escolar (IVE). Fuente: Elaboración propia.	131
9.4. Flujo de Caja - Escenario sin Proyecto	136
9.5. Flujo de Caja - Escenario con Proyecto	137
A.1. Parámetros árboles de decisión	146
A.2. Parámetros Random Forest	147
B.1. Atributos Guía Matrícula e Índice de Vulnerabilidad Escolar (IVE)	148
B.2. Créditos Universitarios	153

B.3. Becas de Arancel Internas y Externas	154
B.4. Becas de Mantenimiento Internas y Externas	155
B.5. Resultados proceso PSU	156
B.6. Nueva variable “Enfermedad Permanente o Crónica”	157
B.7. Nueva variable “Hacinamiento”	158
B.8. Nueva variable “Nivel de Educación Familiar”	158
B.9. Nueva variable “Región”	159
B.10. Nueva variable “Carrera anterior no terminada”	159
B.11. Nueva variable “CAE - FSCU”	160
B.12. Análisis Becas de Arancel	161
B.13. Nuevas variables “Becas de Arancel tipo I y II”	163
B.14. Selección epsilon y min points DBSCAN	175
B.15. Cluster X- Mean: Proporción de true values de variables binarias	176
B.16. DBSCAN: Proporción de true values de una selección de variables categóricas	177
B.17. Coeficientes Regresión Logística	181
B.18. Coeficientes Regresión Logística (continuación)	182
C.1. Análisis Retrospectivo - Gratuidad. Fuente: Elaboración propia.	183
C.2. Análisis Retrospectivo - Beca Vocación Profesor. Fuente: Elaboración propia.	183
C.3. Análisis Retrospectivo - Tramo Socioeconómico. Fuente: Elaboración propia.	184
C.4. Análisis Retrospectivo - Carrera. Fuente: Elaboración propia.	185
C.5. Análisis de sensibilidad	186

Capítulo 1

Introducción y Contexto de la Universidad de Chile

En este capítulo se describe la Universidad de Chile y la unidad donde se centra el trabajo de tesis la cual corresponde a la Dirección de Bienestar Estudiantil (DBE). Además se entregan antecedentes de la problemática de deserción estudiantil a nivel de educación superior, universidades chilenas y facultades de la Universidad de Chile.

1.1. Descripción general Universidad de Chile

La Universidad de Chile es una institución de Educación Superior del Estado de carácter nacional y público, con personalidad jurídica, patrimonio propio y plena autonomía académica, económica y administrativa, dedicada a la enseñanza de educación superior, investigación, creación y extensión de las ciencias, las humanidades, las artes y las técnicas, al servicio del país en el contexto universal de cultura ¹.

Esta cuenta con más de 175 años al servicio al país, 236 postgrados y especializaciones, 183 premios nacionales, 2 premios nobel, 20 exalumnos que han sido presidentes de Chile , 43.802 m2 de salas

¹Estatuto de la Universidad de Chile

de clases, 40.494 estudiantes de pregrado y postgrado, 3.825 académicos con un 60,7% de ellos con grado de doctorado, 1.728 académicos jornada completa (más de 33 horas) y 9.234 funcionarios ².

1.1.1. Estructura Organizacional

La estructura de la Universidad de Chile se encuentra definida en sus Estatutos, el cual fue aprobado en marzo de 2006 en reemplazo al decreto con fuerza de ley de 1981. En estos se establece que los Organismos Superiores de Administración de la universidad corresponden al Rector, el Consejo Universitario y el Senado Universitario, quienes se encargan de dirigir, gestionar, proyectar y normar a la universidad en base al cumplimiento de la misión institucional.

Bajo dependencia del Rector, se encuentran, como unidades ejecutivas centrales, la Prorectoría, las Vicerrektorías y demás unidades académicas requeridas para la administración superior. Estas se encargan de orientar, coordinar, apoyar y desarrollar la labor universitaria.

En tanto que la estructura académica se encuentra conformada por Facultades, Departamentos, Institutos, Centros y Escuelas, las cuales facilitan el ejercicio de la enseñanza de educación superior, investigación y extensión del conocimiento, gozando de autonomía dentro del marco de sus funciones universitarias, siendo el Decano la máxima autoridad de la Facultad.

En la figura 1.1 se visualiza la estructura organizacional de la universidad, siendo esta compuesta por los Organismos Superiores de Administración, 6 Vicerrektorías, 14 Facultades, 5 Institutos, 1 Hospital, 1 Centro de Extensión, entre otros.

²Una comunidad, una misión, un país. Universidad de Chile 2014-2018

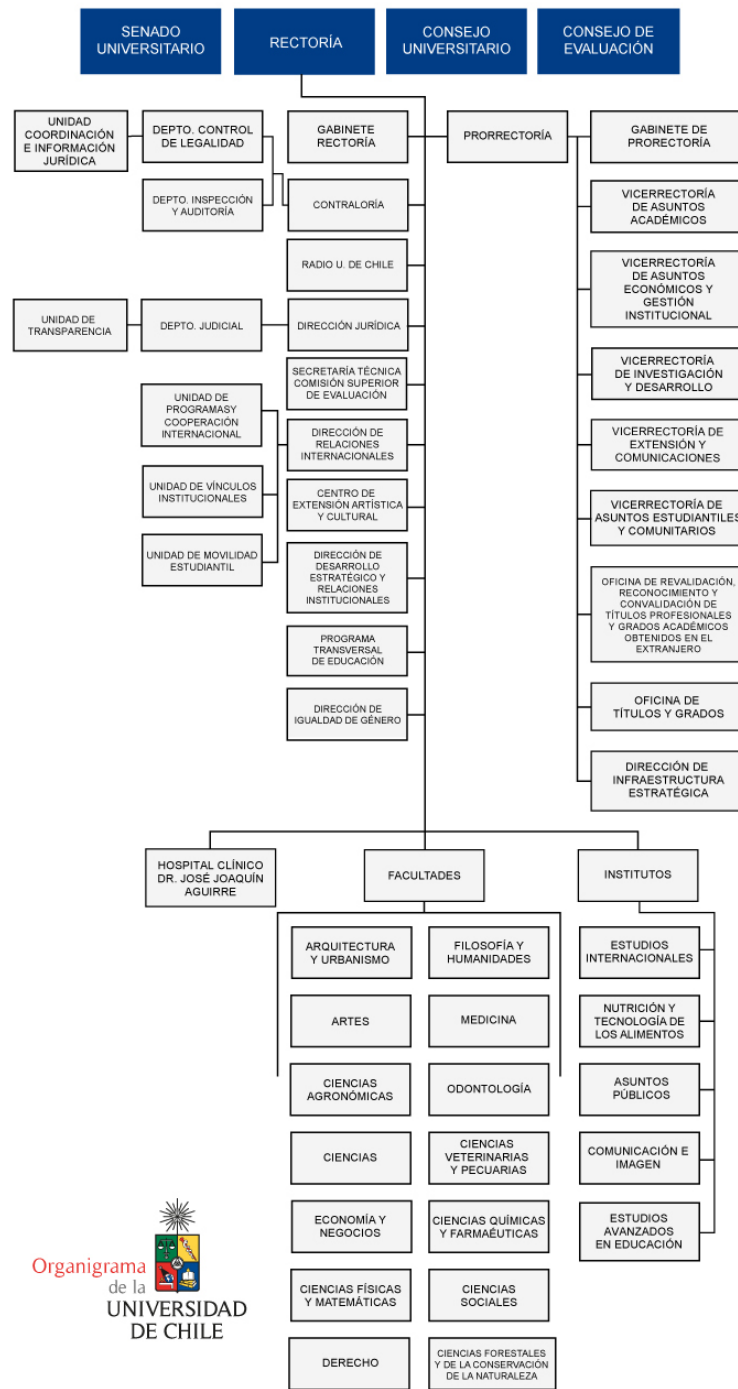


FIGURA 1.1: Organigrama Universidad de Chile

1.1.2. Vicerrectoría de Asuntos Estudiantiles y Comunitarios

El presente trabajo se centra en la Vicerrectoría de Asuntos Estudiantiles y Comunitarios (VAEC), la cual se encarga de contribuir a la formación y desarrollo integral de los estudiantes de la Universidad de Chile, bajo un contexto de Equidad e Inclusión. Particularmente, el foco se encuentra en una de sus direcciones: la Dirección de Bienestar Estudiantil.

1.1.2.1. Dirección de Bienestar Estudiantil (DBE)

Esta dirección se encuentra a cargo de generar políticas, lineamientos y programas para contribuir al acceso, retención y avance académico de estudiantes de contextos vulnerables, generando de esta manera condiciones de mayor equidad en el acceso, desarrollo y egreso de los alumnos.

Dentro de las actividades de la DBE se encuentran administrar las becas de arancel y mantención de la universidad; administrar programas de ingreso prioritario (SIPPE) para alumnos con necesidades socioeconómicas; administrar programas de residencia universitaria para alumnos de regiones; mejorar y actualizar los mecanismos de evaluación y asignación de becas; articular beneficios estudiantiles entre nivel central y facultades; y crear nuevos programas que ayuden al acceso, desarrollo y egreso del estudiante.

La DBE presta servicio a las carreras de pregrado de 14 facultades y 5 institutos que posee la universidad, las cuales ofrecen un total de 70 carreras de pregrado, un programa de bachillerato, un Plan Común de Ingeniería y 4 Programas de Etapa Básica de Artes. En 2017 estas carreras estuvieron conformadas por un total de 6.526 estudiantes nuevos y un total de 32.422 estudiantes de pregrado (78 % del total de la matrícula)³.

1.1.2.2. Estructura Organizacional DBE

Esta dirección posee 5 áreas de trabajo:

Unidad de Gestión y Proyectos: Encargada de la planificación, diseño de procesos, gestión administrativa y presupuestaria de la Dirección de Bienestar Estudiantil. De igual forma, lidera las tareas de vinculación con el medio.

³Anuario Universidad de Chile 2017

Área de Beneficios Estudiantiles: Encargada de dirigir los procesos de financiamiento estudiantil asociados a las becas de mantención y de aranceles, tanto externas como internas. De igual forma, establece acciones de apoyo a partir de la caracterización socioeconómica de los estudiantes.

Programa de Residencias Universitarias: Encargada de coordinar y gestionar beneficios de apoyo a la residencia, asignando cupos en los 3 Hogares Universitarios bajo su administración, y becas de residencia para estudiantes provenientes de regiones y de situación socioeconómica deficitaria.

Área de Calidad de Vida: Orientada al desarrollo de programas de autocuidado y desarrollo, e implementación de proyectos estudiantiles y de desarrollo artístico cultural.

Área de Comunicaciones: Establece y gestiona canales informativos, tanto para los estudiantes como para contrapartes institucionales, vinculados a los procesos que lidera la Dirección. Adicionalmente, tiene a cargo la Mesa de Ayuda y las comunicaciones transversales de la VAEC.

Área de Gestión de Información: Encargada de facilitar la gestión de los beneficios y/o Servicios que gestiona la Dirección, aplicando tecnologías para facilitar los procesos, como también, contribuyendo a generar información relevante a nivel institucional.

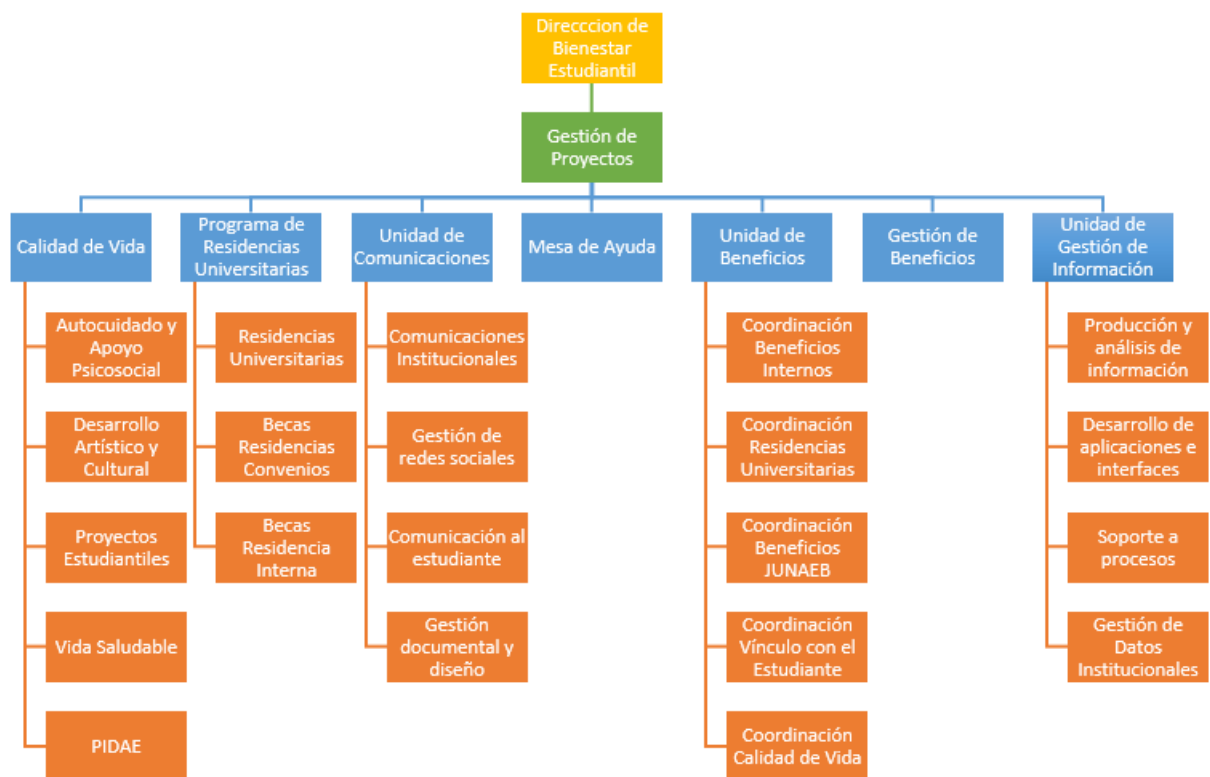


FIGURA 1.2: Organigrama DBE

1.2. Antecedentes del problema en la educación superior

Las Instituciones de Educación Superior (IES) se clasifican en en Institutos Profesionales (IP), Centros de Formación Técnica (CFT) y Universidades. En la figura 1.3 se observa que las universidades tienen en promedio tasas de retención superiores a los IP y CFT, y que la Universidad de Chile mantiene una tasa de retención mayor al sector universitario. A su vez, desde el año 2012 al 2016 se observan aumentos en la retención de estudiantes en los tres sectores, con un alza de un 3,9% para los IP, 3,4% para los CFT y 3,3% para las Universidades⁴

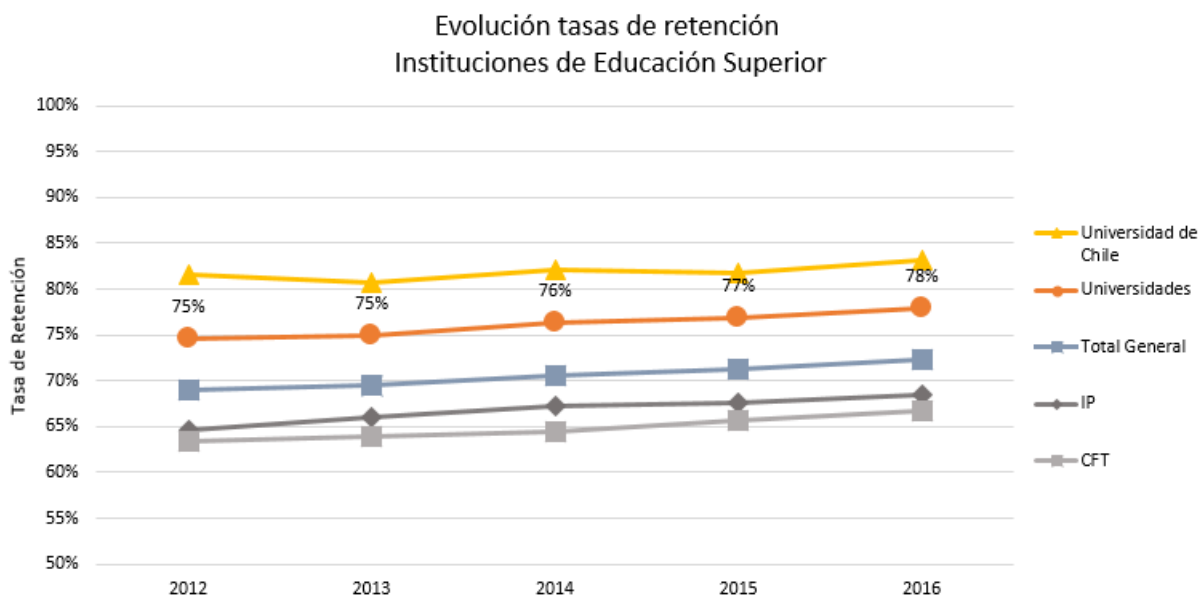


FIGURA 1.3: Evolución tasas de retención Instituciones de Educación Superior. Fuente Elaboración Propia, a partir de Informe de Retención de primer año de pregrado SIES Cohortes 2012- 2016.

Al comparar los valores de retención de la Universidad de Chile con aquellas universidades que han figurado al menos una vez dentro de los cinco primeros puestos del país⁵, la Institución se encuentra en el puesto número cuatro, tras la U. de los Andes, U. Adolfo Ibañez y Pontificia U. Católica (ver

⁴Fuente: Informe de Retención de primer año de pregrado SIES Cohortes 2012- 2016.

⁵ La universidad de Chile participa en la medición de diferentes rankings mundiales que califican el desempeño de las universidades (ScimagoIR, América Economía, ARWU, URAP y QS World University Rankings principalmente). Se utilizaron estos rankings para determinar qué universidades figuran dentro de los 5 primeros puestos a nivel nacional para así poder comparar sus tasas de retención, de lo cual se obtuvieron que las universidades que figuran dentro de dichos puestos corresponden a: Universidad de Concepción, Universidad Técnica Federico Santa María, Universidad de Santiago de Chile, Universidad Católica, Universidad Andrés Bello, Pontificia Universidad Católica de Valparaíso, Universidad de los Andes, Universidad Adolfo Ibañez y Universidad de Chile.

figura 1.4), observando una diferencia de 5,5% de la universidad que posee una mayor retención (U. de los Andes) y un 10,3% de la universidad que posee menor retención (U. Andrés Bello).

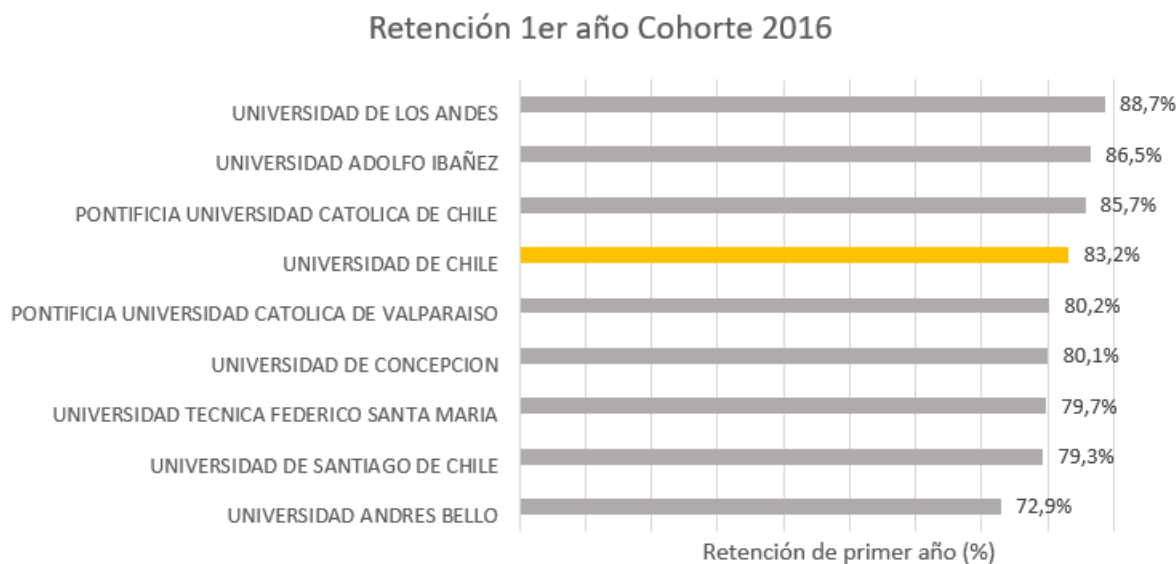


FIGURA 1.4: Retención 1er año Cohorte 2016. Fuente Elaboración Propia, a partir de datos de retención de primer año de pregrado SIES 2016

De forma interna, existe una variación amplia de retención según la facultad y carrera a la que pertenecen los alumnos (ver figura 1.5). El anuario 2017 de la Universidad de Chile indica que esta posee un promedio de 82,7% de retención en los programas académicos, existiendo 10 facultades que se encuentran por sobre promedio. Destacan la Facultad de Ciencias Físicas y Matemáticas, Facultad de Derecho y Facultad de Ciencias Veterinarias y Pecuarias las cuales poseen una tasa de retención sobre el 90%.

**Tasa de retención al primer año (%), por unidad académica
cohorta 2016- año 2017**

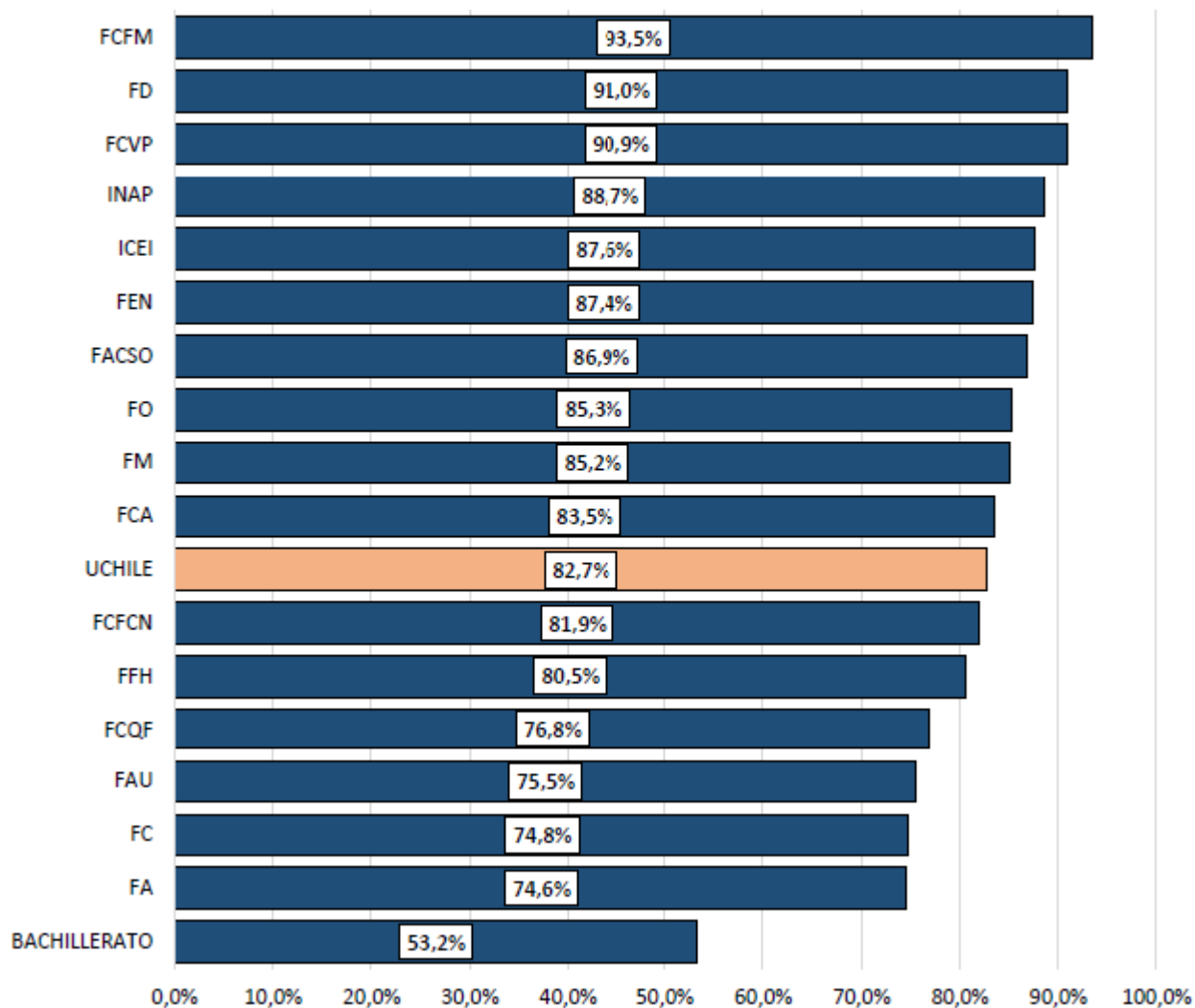


FIGURA 1.5: Tasa de retención al primer año por unidad académica. Fuente: Anuario 2017 Universidad de Chile

Capítulo 2

Proyecto

En este capítulo se detallan los antecedentes que justifican la realización de un proyecto y la descripción de este proyecto de tesis.

2.1. Antecedentes y justificación

Como se mencionó anteriormente, en el año 2014 la Universidad de Chile creó la “Política institucional de Equidad e Inclusión”¹, la cual sentó las bases para una universidad equitativa e inclusiva, garantizando de esta manera el acompañamiento de estudiantes en el acceso, permanencia y egreso oportuno.

Esta política interna de acompañamiento de estudiantes, también fue reforzada por políticas estatales promulgadas en 2018, como lo son “Ley sobre Educación Superior”² y la “Ley sobre Universidades Estatales”³, en las que se establece que para que una universidad pueda contar con gratuidad debe mantener políticas que permitan un acceso equitativo y programas de apoyo a estudiantes vulnerables para su retención.

Si bien la Universidad de Chile realiza esfuerzos por acompañar y retener a sus alumnos, esta sigue manteniendo una marcada heterogeneidad en la retención de sus distintas facultades, siendo 93,5 %

¹<http://www.uchile.cl/portal/presentacion/vicerrectoria-de-asuntos-estudiantiles-y-comunitarios/oficina-de-equidad-e-inclusion/documentos/110549/informes>

²<https://www.leychile.cl/Navegar?idNorma=1118991>

³<https://www.leychile.cl/Navegar?idNorma=1119253>

en la más alta (Facultad de Ciencias Físicas y Matemáticas) y 53,2% en la más baja (Bachillerato) (ver imagen 1.5).

Estos antecedentes suponen un gran desafío y responsabilidad para la Universidad, ya que debe garantizar el acompañamiento de sus alumnos durante su trayectoria universitaria, responder a un cuerpo estudiantil de diverso origen económico, social y cultural, y mejorar la retención de alumnos dentro de sus distintas facultades. Por lo tanto, es prioritario diseñar y sistematizar la retención para cubrir estas demandas.

2.2. Objetivo General

El objetivo general de este proyecto corresponde a generar una nueva capacidad ⁴ en la Universidad de Chile que permita detectar a alumnos en riesgo de deserción para así ofrecer beneficios estudiantiles de forma proactiva. La fecha de finalización del proyecto corresponde a diciembre de 2018.

2.3. Objetivos Específicos

Los objetivos específicos del proyecto para poder lograr el objetivo general son los siguientes:

1. Sistematizar la detección y asignación de beneficios estudiantiles para alumnos en riesgo de deserción, a través de la articulación de arquitectura de procesos, procesos de negocios y arquitectura de sistemas.
2. Determinar el riesgo de deserción de alumnos de primer año de la Universidad de Chile a partir de la evaluación y aplicación de modelos de minería de datos.

⁴Se utiliza la definición de capacidad del libro de Administración Estratégica de Michel A. Hitt, el cual indica que “Las capacidades existen cuando los recursos han sido integrados a propósito con el fin de desempeñar una o varias tareas específicas (...) Estas son fundamentales para crear ventajas competitivas y, con frecuencia, están basadas en el desarrollo, la transmisión y el intercambio de información y de conocimiento por medio del capital humano de la empresa.”

2.4. Resultados Esperados

Los resultados esperados a partir de los objetivos específicos son los siguientes:

1. Contar con una propuesta de arquitectura de procesos, procesos de negocio y soporte tecnológico que permitan materializar la nueva capacidad.
2. Que los modelos desarrollados puedan predecir los alumnos en riesgo de desertar, de forma de poder realizar acciones que permitan disminuir la tasa de deserción universitaria.

2.5. Alcance

Se detallan los aspectos que se encuentran dentro y fuera del alcance de este proyecto:

1. Dentro del alcance

- Proyecto acotado al estudio de la deserción universitaria de forma voluntaria e involuntaria, es decir, aquellos alumnos que abandonan la Universidad de Chile a través de la formalización del abandono o son expulsados de una carrera por desempeño académico insuficiente o razones disciplinarias, sin volver a matricularse en el siguiente periodo en una carrera de la Universidad de Chile.
- Predicción de deserción mediante algoritmos de minería de datos de alumnos de primer año. Para ello se utilizarán datos de alumnos nuevos de la Universidad de Chile del cohorte 2017 ⁵.
- Validar que modelo puede ser replicado a nuevas generaciones de alumnos de primer año a través de un análisis retrospectivo (validar que modelo es generalizable). Para ello se utilizarán datos de alumnos nuevos de la Universidad de Chile cohorte 2016.
- Diseño de arquitectura de procesos, procesos de negocio y arquitectura tecnológica.

⁵El motivo de escoger alumnos de primer año es que es en el primer año donde se puede generar el mayor vínculo entre el estudiante y la universidad, en tanto que se considera a alumnos de 2017 debido a que este representa el cohorte más cercano del cual se posee datos para poder estudiar al deserción y a que posee características específicas que los diferencian de los anteriores cohortes, como por ejemplo, que el 40% de este cohorte posee gratuidad, en contraste con años anteriores, donde no existía el financiamiento de gratuidad o era de un porcentaje menor.

-
- Evaluación económica de implementar la nueva capacidad en la Universidad de Chile con recursos propios.

2. Fuera del alcance

- Diseñar programas de ayuda estudiantil para mitigar la deserción.
- Predecir el rendimiento académico.

2.6. Criterios de Éxito

Los criterios de éxito de este proyecto son:

- Criterios de Éxito del Negocio

Haber detectado la deserción de al menos un 1 % del total de alumnos nuevos del cohorte 2017.

- Criterios de Éxito de Minería de Datos

Se deberá demostrar que el modelo seleccionado es generalizable, es decir, que es válido no sólo para los datos de alumnos nuevos con que se entrenaron los modelos (cohorte 2017), sino que también para nuevas generaciones de alumnos que ingresan a la Universidad de Chile. Además, los resultados deberán ser validados por expertos de la Dirección de Bienestar Estudiantil.

Se utilizarán dentro de este estudio las variables región, sexo, nacionalidad, carrera, procedencia educacional, vía ingreso, tramo socioeconómico, índice de vulnerabilidad escolar, enfermedad catastrófica, enfermedad permanente, enfermedad crónica, hacinamiento, nivel educacional del jefe de hogar, nivel educacional de la madre, créditos universitarios, becas de arancel, becas de mantención, puntajes PSU, NEM, Ranking y otras variables generadas a partir de los datos disponibles de alumnos.

2.7. Riesgos Potenciales

Los riesgos asociados al desarrollo de este proyecto, probabilidad de ocurrencia, impacto y estrategia de mitigación se detallan en la siguiente tabla:

TABLA 2.1: Riesgos Potenciales

N	Tipo	Riesgo	Probabilidad	Impacto	Estrategia de Mitigación
1	Político	Cambio de prioridad de la administración	Probable	Alto	Gestionar atentamente, mostrando avances en reuniones eficientes y efectivas.
2	Alcance	Definición imprecisa de alcance	Probable	Medio alto	Actualizar el alcance del proyecto a medida que se cuenta con más información
3	Datos	Demora en la entrega de datos o datos insuficientes para realizar estudio.	Probable	Alto	Tramitar solicitud de datos al comienzo del proyecto.
4	Equipo de Proyecto	Apoyo de personal de la DBE durante el desarrollo del proyecto	Probable	Alto	Gestionar atentamente, mostrando avances en reuniones eficientes y efectivas.
5	Proyecto	Baja adopción de la solución	Probable	Alto	Generar plan de gestión del cambio

Capítulo 3

Marco Teórico

En este capítulo se detallan las metodologías, conceptos y bibliografía consultada para el desarrollo de este trabajo de tesis.

3.1. Metodología de Ingeniería de Negocios y Diseño de Servicios

La metodología que se utilizó para llevar a cabo este proyecto corresponde a la "Metodología de diseño de servicios", desarrollado en el libro *Business Engineering and Service Design* [1], la cual se aplica en organizaciones que buscan desarrollar innovaciones en sus servicios para ser más competitivas. Esta metodología se observa en la figura 3.1.

1. **Estrategia y modelo de negocios:** Primero se analizan los aspectos estratégicos de la organización, de forma de comenzar con una estrategia y modelo de negocios innovativo.
2. **Capacidades y diseño de negocios:** A partir de la estrategia y modelo de negocios definido, se define la capacidad que debe ser desarrollada para materializar el modelo de negocios.
3. **Diseño de Arquitectura:** Para incluir la nueva capacidad en la estructura de negocios se utilizan como base para el diseño los Patrones de Negocio, los cuales corresponde a patrones encontrados en diferentes organizaciones para implementar una capacidad determinada.

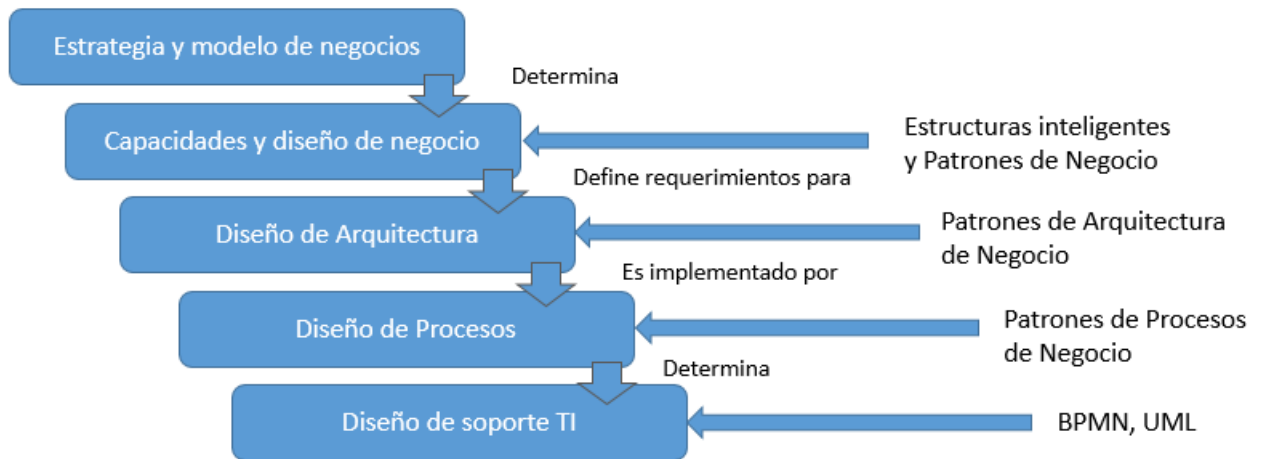


FIGURA 3.1: Metodología de diseño de Servicios. Fuente Business Engineering and Service Design.

4. **Diseño de Procesos:** A partir del diseño de la Arquitectura de Negocios, se diseñan los procesos que deben ser generados o mejorados dentro de la nueva arquitectura. Para ello, se utiliza como referencia para el diseño los Macroprocesos, que corresponden a un conjunto de procesos relacionados que son necesarios para conducir los servicios clave de la organización.
5. **Diseño de soporte de Tecnologías de Información:** Se diseñan las capacidades tecnológicas necesarias para soportar la operación de la Arquitectura y Procesos de Negocio.

3.1.1. Patrón de Negocio

En el libro de *Business Engineering and Service Design* [1], se define para organizaciones que prestan servicios, un modelo conceptual de las interacciones y componentes de negocio implicadas al momento de realizar un diseño sobre el negocio. La estructura básica de este modelo se observa en la figura 3.2.

En este modelo, la Cadena de Valor corresponde a todas las actividades y procesos que ocurren desde que un cliente genera una orden de servicio hasta la entrega.

El Sistema de Administración, corresponde a todas las actividades que se realizan para definir y dirigir la Cadena de Valor para poder cumplir con los requerimientos de sus clientes (desarrollo de nuevos servicios, planificación estratégica, desarrollo de nuevas capacidades, operaciones, logística y ventas).

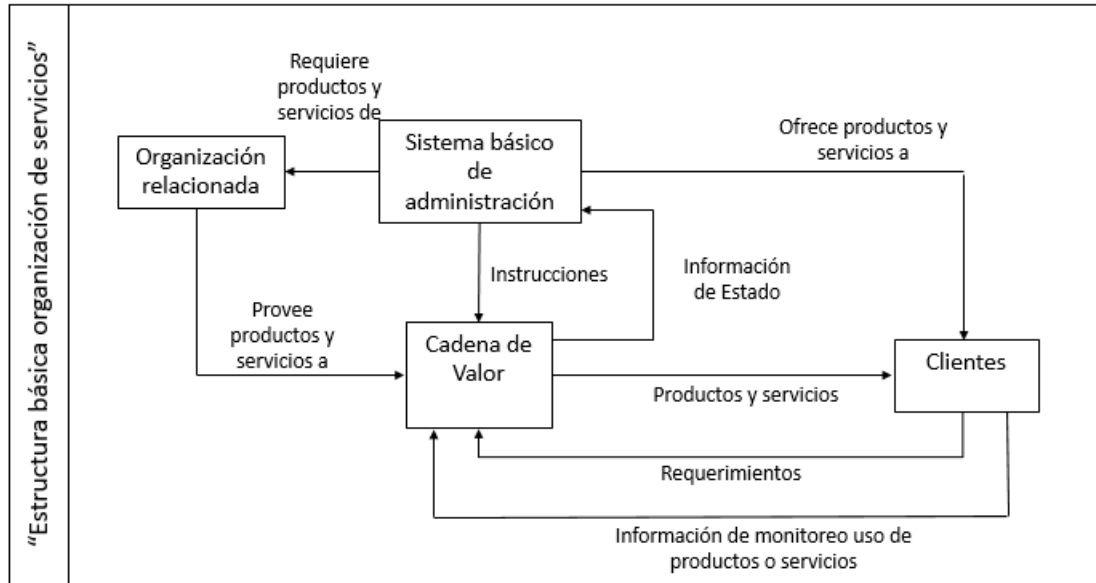


FIGURA 3.2: Estructura básica de organización de servicios. Fuente: Business Engineering and Service Design.

El resto de las componentes corresponde a clientes, proveedores y productos o servicios generados.

Considerar que dentro de este diagrama no se encuentran de forma implícita dentro de las componentes los recursos financieros, humanos y otros tipos de recursos utilizados para la administración.

Dada la estructura básica de organización de servicios (figura 3.2), y partir de estudios de organizaciones que han logrado desarrollar nuevas capacidades, se han reconocido patrones de las nuevas componentes de negocio y sus relaciones que se requiere para generar una nueva capacidad. Estos patrones toman el nombre de “Patrones de Negocio”, donde se han identificado 6 tipos:

1. Ventas Basadas en el conocimiento del Cliente
2. Creación de nuevas cadenas de valor
3. Aprendizaje interno para mejora de procesos
4. Evaluación de performance para planificación y mejora de procesos (KPIs)
5. Innovación de productos
6. Optimización de uso de recursos

El Patrón de Negocio de interés de esta tesis corresponde al Patrón de Negocio 1 “Ventas Basadas en el conocimiento del cliente”, el cual debe ser desarrollado en organizaciones que posean un posicionamiento estratégico, según lo definido por el modelo delta de Hax [2], de entregar servicios integrales al cliente y un modelo de negocio donde la retención del cliente es relevante. Las nuevas componentes de negocio que se requiere añadir a la estructura básica corresponden a “Analizar comportamiento de clientes” y “Generar ofertas”. Esto se observa en la figura 3.3 .

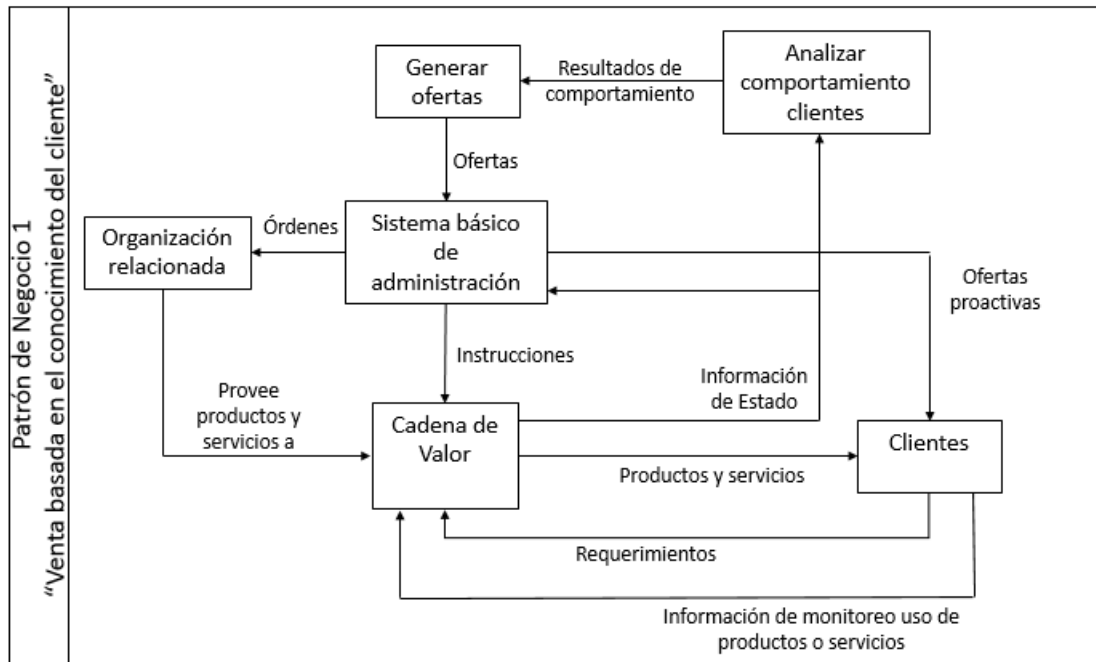


FIGURA 3.3: Patrón de Negocio 1. Fuente: Business Engineering and Service Design.

Para desarrollar estas componentes se requiere:

1. Capturar y organizar los datos de los clientes.
2. Procesar la información con modelos analíticos.
3. Generar ideas en base a los resultados para generar ofertas proactivas.

Estas componentes deben ser mapeadas en patrones de arquitecturas y de procesos de negocio.

3.1.2. Patrones de Arquitectura y Procesos de Negocio

Los Patrones de Arquitectura corresponden a patrones de estructuras organizacionales para llevar adelante la nueva capacidad deseada, mientras que los Patrones de Procesos de Negocio corresponden a una conceptualización de procesos comunes de diferentes organizaciones.

Los Procesos de Negocio se pueden agrupar en 4 grandes procesos, o para diferenciarlo de los procesos comunes llamados “macroprocesos”, debido a que contienen muchos procesos relacionados, subprocesos y actividades que son necesarias para llevar adelante los servicios clave, existiendo 4 macroprocesos:

1. **Macroproceso 1 o Cadena de Valor:** Corresponde a un conjunto de procesos para producir bienes o servicios que ofrece la organización a sus clientes, que comienza con la formulación de requerimientos y termina con la satisfacción de los pedidos. Incluye las actividades de marketing, ventas, abastecimiento, producción y logística.
2. **Macroproceso 2 o Desarrollo de Nuevas Capacidades:** Corresponde a un conjunto de procesos para el desarrollo de nuevas capacidades que requiere la organización para ser competitiva.
3. **Macroproceso 3 o Planificación de Negocio:** Corresponde a un conjunto de procesos que son necesarios para llevar a cabo la dirección de la organización, en la forma de estrategias, planes, programas y presupuesto.
4. **Macroproceso 4 o Administración de Recursos:** Corresponde a un conjunto de procesos necesarios para una operación correcta de los macroprocesos. Estos pueden ser recursos financieros, recursos humanos, infraestructura y materiales.

Estos macroprocesos pueden ser combinados en diferentes estructuras dependiendo del tipo de negocio. La estructura básica se observa en la figura 3.4, donde existe sólo un macroproceso de cada cada tipo (Cadena de Valor, Desarrollo de Nuevas Capacidades, Planificación de Negocio y Administración de recursos). En casos reales, pueden existir muchas cadenas valor, las cuales deben ser mapeadas en el diagrama, así como el detalle de sus interacciones.

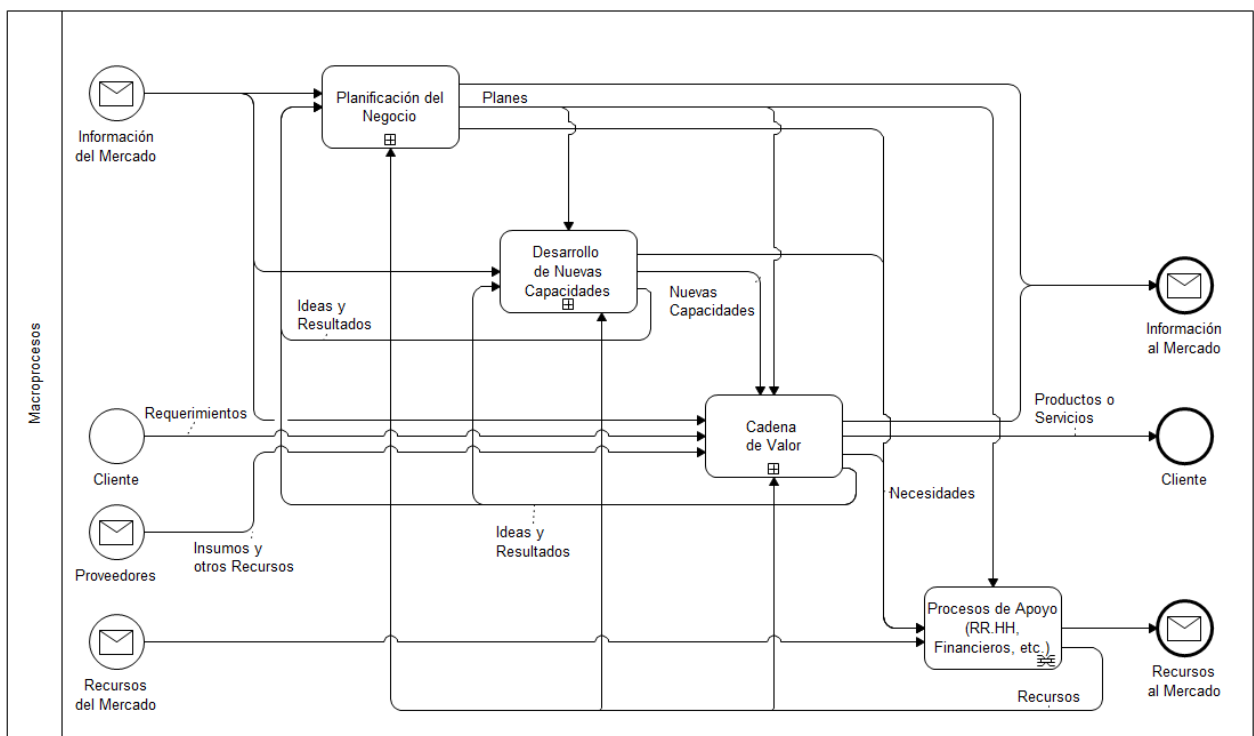


FIGURA 3.4: Macroprocesos. Fuente: Business Engineering and Service Design.

Para arquitecturas complejas, donde existe más de un nivel de negocio, como ocurre en el caso de holdings, complejos mineros (minas, plantas de procesamiento, fundición y servicios de agua), complejos forestales (aserraderos, celulosa y papeles), entre otras, existe una Arquitectura Multinivel, la cual se observa en la figura 3.5.

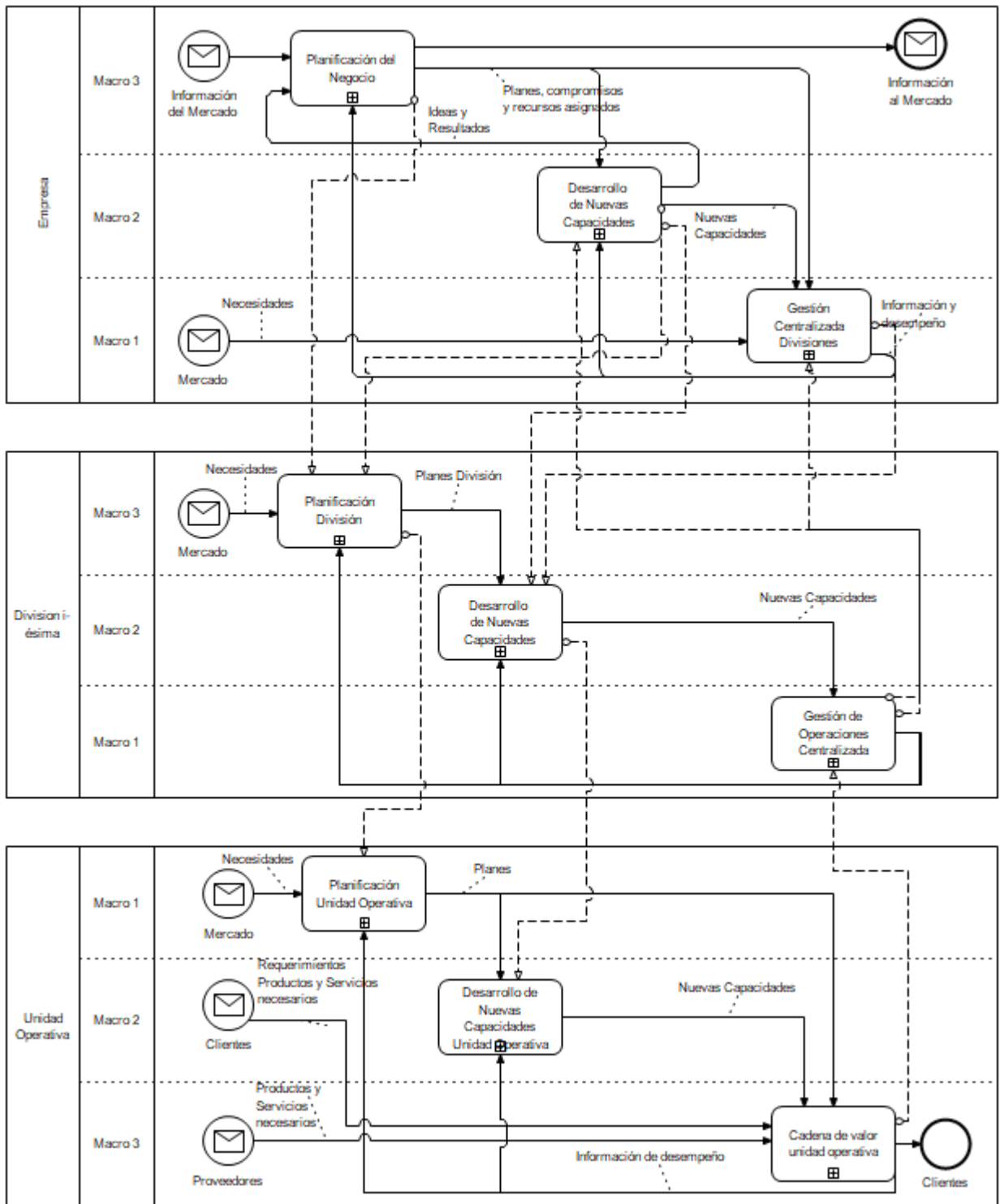


FIGURA 3.5: Arquitectura Multinivel Fuente: Business Engineering and Service Design.

En la Arquitectura Multinivel también se mapean los 4 macroprocesos, con la diferencia que existe

niveles centrales, intermedios y bajos de planificación, cadenas de valor, desarrollo de nuevas capacidades y administración de recursos (por simplicidad del diagrama se omite el macroproceso 4 administración de recursos, asumiendo que se encuentra implícito dentro de las actividades).

Para diseñar el modelo multinivel de arquitectura de procesos es relevante considerar los siguientes aspectos:

1. **Cómo distribuir el Plan de Negocio y el Desarrollo de Nuevas Capacidades**

Se debe decidir si estos aspectos deben ser centralizados o descentralizados. Una centralización de la planificación del negocio y el desarrollo de nuevas capacidades para ciertas operaciones puede generar un mejor rendimiento de los procesos al contar con áreas especializadas que permitan generar mayores beneficios a través de analítica, reduciendo los costos de desarrollar los procesos y sistemas.

2. **Compartir servicios**

Decidir si centralizar servicios es relevante, para así aprovechar economías de escala, por ejemplo, al adquirir productos o servicios que son requeridos por varias unidades dentro de la organización.

3.2. Metodología de Minería de Datos CRISP-DM

Para desarrollar la lógica de negocio se utilizó la metodología de *Cross Industry Standard Process for Data Mining* (CRISP-DM), la cual corresponde a una de las metodologías más utilizadas de minería de datos¹. Esta metodología consiste en un proceso jerárquico, consistente en 4 niveles de abstracción desde lo más general a lo más específico, las cuales corresponden a fase, tarea genérica, tarea especializada e instancia de procesos. Si bien la descripción de las fases es de carácter discreto, en la práctica muchas tareas pueden realizar al mismo tiempo, o avanzar en el proceso para luego realizar ajustes en etapas anteriores.

El detalle de cada una de las fases de la Metodología de CRISP-DM (ver figura 3.6) se detalla a continuación:

¹ <https://www.kdnuggets.com/polls/2014/analytics-data-mining-data-science-methodology.html>

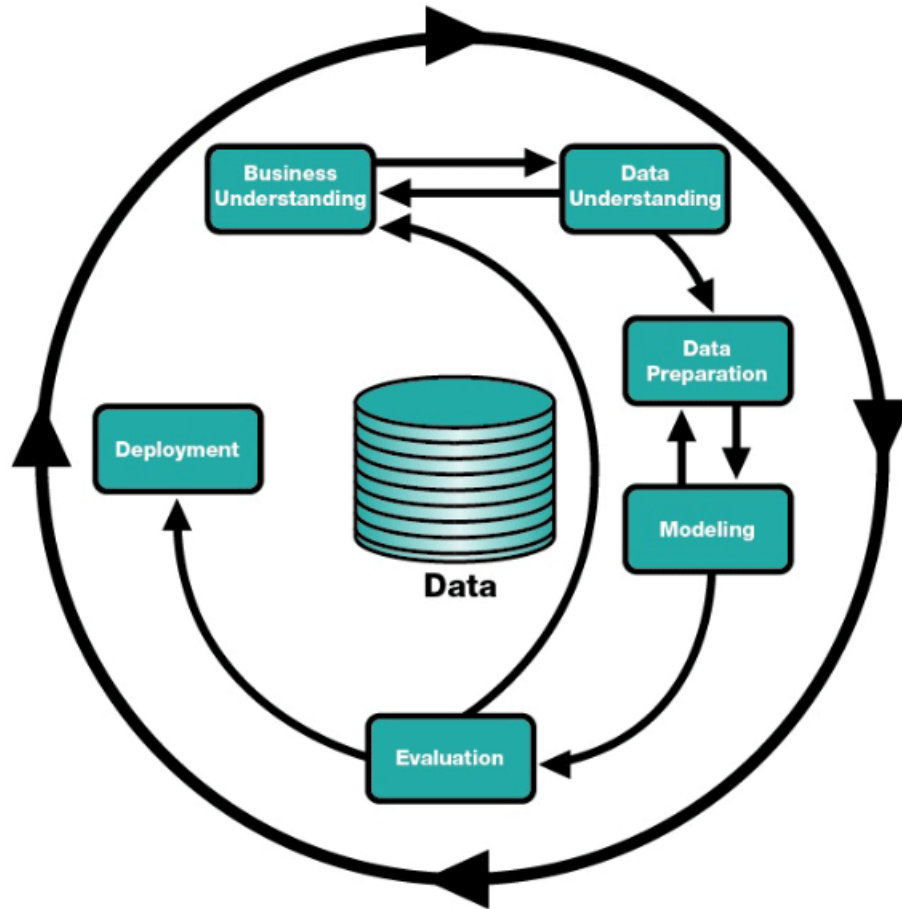


FIGURA 3.6: Metodología CRISP-DM. Fuente: Smart Vision Europe.

1. **Comprensión de negocio:** Esta fase se centra en comprender de forma correcta las necesidades de negocio que se quieren resolver, detallando objetivos, criterios de éxito, alcance, riesgos y plan de proyecto de minería de datos. Además se determina que herramientas de minería de datos será utilizado en el proyecto.
2. **Comprensión de datos:** Se recolecta los datos para ser utilizados en el proyecto, realizando una descripción de los datos encontrados (formato, categorías de variables, cantidad de registros, entre otros), análisis exploratorios (gráficos, correlaciones, distribución de variables, análisis estadísticos simples) y calidad de los datos (valores faltantes, datos desactualizados y otros errores en los datos). Como output se genera un reporte indicando como abordar cada uno de estos puntos y si la data es suficiente para poder cumplir con los objetivos de negocio.
3. **Preparación de datos:** Consiste en seleccionar los atributos (columnas) y registros (filas) a

ser utilizados, limpieza de información y reemplazo de datos faltantes, construcción de nuevas variables e integración de los datos, ya que generalmente se cuenta múltiples bases de datos.

4. **Modelado:** En esta fase se escogen los modelos a utilizar, considerando sus condiciones para operar de forma correcta (por ejemplo, algunos modelos operan bien con missing values, otros descartan los registros con datos faltantes, diseño de prueba (ej. separación de test de entrenamiento y de evaluación con método de validación cruzada), ajuste de parámetros y juzgar mediante indicadores el performance del modelo.
5. **Evaluación:** Se evalúan los resultados del modelo y si estos cumplen los criterios de éxito definido en los objetivos definidos en la primera fase. Además, se consideran los hallazgos obtenidos de datos que no estaban definidos como un objetivo de negocio, pero que si agregan información relevante para futuros proyectos.
6. **Puesta en marcha:** Se genera un plan de implementación, definiendo pasos a seguir, y planes de monitoreo y mantención de modelos.

3.3. Método de selección de variables *Forward Selection and Backward Elimination*

Para determinar que variables deben ser incluídas en los modelos, se utilizó el método combinado de *Forward Selection and Backward Elimination* [3], el cual corresponde a un algoritmo de selección de variables como se detalla a continuación:

Forward Selection: Algoritmo que inicia el modelo sin atributos, para luego agregar aquel que posea una mayor correlación con la variable explicada. En caso de no existir una variable significativa para el modelo, el proceso se termina, determinando que no hay ninguna variable relevante con poder predictor dentro del modelo. En caso de que si exista una, añade esta primera variable al modelo. Luego, el algoritmo corre nuevamente, teniendo como base la variable seleccionada, el cual es comparado con cada una de las variables restantes a partir de pruebas estadísticas (*F-test*). El algoritmo termina cuando ninguna variable mejora el rendimiento del modelo.

Backward Elimination: Algoritmo que inicia el modelo con todos los atributos, para luego ir eliminando aquellos que agreguen menor valor al modelo, la cual corresponderá a la variable que posea

un menor estadístico F (*F min*). En caso de que *F min* no sea estadísticamente significativo, el algoritmo elimina la variable y continúa su iteración. En caso de que sea estadísticamente significativo, el algoritmo se detiene.

Forward Selection and Backward Elimination (Stepwise): Este algoritmo opera como *Forward Selection*, pero cada vez que agrega una variable, examina si una variable debe ser eliminada con el método de *Backward Elimination*. El algoritmo termina cuando ninguna variable puede ser añadida o eliminada.

Esta metodología es práctica para la selección de atributos, ya que la alternativa sería probar todas las combinaciones posibles de modelos, lo cual crecería de forma exponencial a medida que se añaden nuevas variables, de forma $2^p - 1$, donde p es el número de variables. A modo de ejemplo, para un modelo de 10 variables predictoras, se tendrían que probar $2^{10} - 1 = 1.023$ modelos. Este método si bien encuentra óptimos locales y no globales, es práctico y utilizado ampliamente en la literatura para la selección de variables.

3.4. Modelos de Minería de Datos

Se define minería de datos como “el proceso de descubrir correlaciones, patrones y tendencias relevantes a través del tratamiento de grandes bases de datos, y utilizando tecnologías para el reconocimiento de patrones, así como técnicas estadísticas y matemáticas”².

Dentro de la minería de datos existen modelos supervisados y no supervisados, los cuales son utilizados según el problema que se busque resolver.

3.4.1. Modelos Supervisados

Corresponde a modelos donde se cuenta con variables explicativas y una variable explicada. A partir de la data disponible, se extrae un set de entrenamiento, el cual es utilizado para ajustar los parámetros de los modelos. De esta manera en un escenario óptimo, se espera que a partir de una nueva instancia la variable explicada sea predecida correctamente. Esto requiere que los algoritmos

²Gartner Inc., Data Mining definition

tengan una capacidad de predicción que sea generalizable, es decir, que permita predecir de forma razonable casos que no se encuentren dentro de la muestra.

3.4.1.1. Regresión Logística

La regresión logística es un modelo supervisado utilizada cuando la variable explicada es de carácter categórica (dos o más categorías), lo cual es útil para los fines de este estudio (predecir si un alumno deserta o no deserta).

$$P[Y = 1/x_1, x_2, \dots, x_k] = \frac{e^{g(x)}}{e^{g(x)} + 1}$$

donde

$$g(x) = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots + \beta_k * x_k$$

Los parámetros del modelo (β 's) son estimados según el método de máxima verosimilitud $l(\beta|x)$, la cual determina los valores para los beta según la máxima probabilidad de reproducir el conjunto de los datos observados. El supuesto de que las observaciones son independientes permiten expresar la máxima verosimilitud como el producto de cada una de las observaciones.

$$L(x_1, x_2, \dots, x_k, \beta) = \prod_{k=1}^k f(x_k, \beta)$$

Aplicando logaritmo a la ecuación, se puede realizar un cálculo con un mejor rendimiento computacional:

$$\ln(L(\beta|x)) = \sum_{k=1}^k \ln(f(x_k, \beta))$$

Las ventaja de este modelo es que los resultados de son sencillos de interpretar, ya que lo valores varían entre 0 y 1, lo cual corresponde a la probabilidad de ocurrencia de un suceso. Por ejemplo, calcular la probabilidad de que un alumno sea desertor.

Este modelo es sensible a los missing values, ya que los valores faltantes son interpretados como nulos en caso de las variables numéricas o equivalentes a la categoría definida como base en el caso de las variables categóricas, afectando la calibración de los parámetros y la capacidad predictora del modelo. Además, asume variables numéricas distribuyen como una normal, por lo cual los datos deben ser normalizados antes de ser utilizados [3].

3.4.1.2. Árboles de decisión

Los árboles de decisión son algoritmos predictivos que permiten clasificar la información en dos o más categorías. Estos están compuestos de ramificaciones, donde los puntos en que se divide la información es llamado nodo, donde un nodo corresponde a la variable escogida por el algoritmo para ramificar (*split*). El final de la ramificación en donde no se divide nuevamente el algoritmo es llamado hoja.

Algunas ventajas de los árboles de decisión:

- Son simples de comprender y visualizar, ya que los árboles pueden ser desplegados gráficamente y comprendidos por personas que no son expertos en la materia.
- Permite trabajar tanto con variables numéricas y categóricas, dependiendo del algoritmo de árbol de decisión utilizado.
- Tiene un buen performance para sets de datos grandes, utilizando un tiempo razonable para analizar la información.
- Robusto contra colinearidad de las variables.

Las desventajas de este modelo corresponden a:

- Los algoritmos de árboles de decisión no garantizan óptimos globales. Esto puede ser mitigado creando muchos árboles (*Random Forest*) o poda (*prunning*).
- Puede generar tendencias si hay una clase dominante. Es por ello que se recomienda balancear los datos.

-
- Sensibles a cambios de set de datos.

Existen diferentes algoritmos de árboles de decisión, los cuales se diferencian principalmente en el criterio que utilizan para decidir que nodo se debe utilizar para generar un split. Para efectos de este estudio se utilizarán los algoritmos *Iterative Dichotomiser 3* (ID3), *Classification and Regression Trees* (CART) y *Chi-squared Automatic Interaction Detection* (CHAID), los cuales poseen diferentes criterios y parámetros para generar los *split*. El detalle a continuación:

1. Algoritmo ID3

Este algoritmo utiliza las métricas de entropía y de ganancia de información para generar los *split*.

- **Función de entropía:** Caracteriza la impureza de un conjunto de datos. Este se calcula tanto de la variable explicada del set de datos S como para cada una de las variables del mismo set de datos, bajo la siguiente fórmula:

$$Entropia(n) = - \sum_{x=i}^{n=k} p(n, x) * \log_2(p(n, x))^2$$

Donde P(n,x) corresponde a la probabilidad de ocurrencia de la clase x del nodo n, donde n corresponde a la variable analizada, x corresponde a una categoría y k la cantidad de categorías de la variable.

La función de entropía perfectamente clasificado (todos positivos o todos negativos) toma valor 0. Para problemas de clasificación binaria, una entropía donde todos los valores son positivos o negativos la entropía es igual 0. En tanto que si la mitad corresponde a casos positivos y la otra mitad a negativos, la entropía es igual a 1.

- **Ganancia de información:** Permite medir cuánta incertidumbre se reduce al dividir el set de datos por una variable A. Para ello se calcula la entropía del set de datos S y se le resta la entropía de la variable A del mismo set de datos.

$$Ganancia\ de\ informaci\ on = Entropia(S) - Entropia(S, A)$$

Los pasos del algoritmo de ID3 corresponden a los siguientes:

-
- a) Calcular la entropía del data set.
 - b) Para cada variable:
 - 1) Calcular la entropía para todas sus categorías.
 - 2) Obtener el promedio ponderado de entropía de la variable.
 - 3) Calcular la ganancia de información (*information gain*) de la variable.
 - c) Escoger el atributo con mayor ganancia de información. En otras palabras, se escoge el atributo que reduce más la incertidumbre.
 - d) Repetir hasta obtener el árbol deseado

Ventajas de este algoritmo son:

- Favorece particiones más pequeñas.
- Permite la poda de árboles (*prunning*), reduciendo la cantidad de tests, lo cual es realizado a través de tamaño mínimo para *split*, tamaño mínimo de las hojas y ganancia mínima.

Desventajas:

- Sobreajuste si se utiliza un set de datos pequeños.
- Sólo un atributo a la vez es testeado para hacer la decisión de selección.
- Sólo utiliza variables categóricas.
- Tendencia a seleccionar variables para *split* que poseen muchas categorías.

2. Algoritmo CART

A diferencia de ID3 este algoritmo utiliza el Índice de Gini para generar los *splits*, en vez de la entropía.

Índice de Gini: Esta métrica indica que tan mezcladas se encuentran las clases en dos grupos creados por el *split*. Este se calcula como:

$$G = 1 - \sum p(n, x)^2$$

Un set perfectamente clasificado (todos positivos o todos negativos) toma valor de índice de gini igual a 0.

Para problemas de clasificación binaria, un índice de gini perfectamente clasificado toma valor 0. En tanto que si la mitad corresponde a casos positivos y la otra mitad a casos negativos, el índice de gini es igual a 0,5. Para el resto de las variables categóricas el máximo valor del índice de gini $=1-1/k$, donde k el número de categorías.

Algoritmo:

- a) Calcular el índice de gini para el set de datos.
- b) Para cada variable:
 - 1) Calcular el gini index para todas sus categorías.
 - 2) Obtener el promedio ponderado de entropía de la variable.
 - 3) Calcular la ganancia de gini (*gini gain*) de la variable.
- c) Escoger el atributo con mayor ganancia de gini.
- d) Repetir hasta obtener el árbol deseado

Ventajas del algoritmo:

- Fácil de interpretar y visualizar.
- Permite utilizar datos numéricos y categóricos.
- Relaciones no lineales en los parámetros no afectan el performance del árbol.

Desventajas:

- Puede tener problemas para generalizar los resultados (*overfitting*).
- Inestables con pequeñas variaciones en los datos, resultando en árboles diferentes.

3. Algoritmo CHAID

Este algoritmo utiliza la prueba de hipótesis chi-cuadrado para generar los *split*, lo cual lo realiza a través del estadístico no paramétrico chi cuadrado, el cual compara la distribución observada de los datos con la distribución esperada de estos.

Este test sólo utiliza variables categóricas.

Poda de árboles (*Pruning*)

En árboles de decisión se puede realizar una poda de árboles para restringir la cantidad de split que se realizan, existiendo dos tipos de poda: pre-podado y podado. El prepodado se encuentra

conformado por *minimal size for split*, *minimal leaf size* y *minimal gain*. En tanto que le podado se encuentra conformado por *confidence*. Además de lo anterior, se puede restringir por la profundidad máxima de los árboles (*maximal depth*). En anexo Parámetros árboles de decisión se detalla cada uno de estos parámetros y si es poseído por los algoritmos ID3, CART y CHAID.

3.4.2. Modelos No Supervisados

Estos modelos son utilizados para buscar relaciones y patrones en los datos que no son conocidos, sin la necesidad de definir variables explicativas y explicadas como en los modelos supervisados. Frecuentemente estos métodos son utilizados para encontrar patrones ocultos en los datos, de forma de comprender en mayor profundidad la información disponible y etiquetar los distintos grupos encontrados para posteriores análisis.

3.4.2.1. Análisis de Cluster

El análisis de cluster sirve como una herramienta exploratoria de información, ya que permite agrupar la información según su similaridad y generar hipótesis a partir de ella. Se utilizarán dentro de este proyecto los algoritmos de clusterización: X-Mean y DBSCAN. El primero debido a que permite encontrar el número óptimo de clusters y el segundo porque elimina el ruido de la muestra.

Para comprender el algoritmo de cluster X-Mean, se detallará en primer lugar el funcionamiento de cluster K-Mean, debido a que X-Mean utiliza el mismo algoritmo con la diferencia que determina el número óptimo de clusters (k óptimo).

1. Cluster K-means

Este algoritmo posee las siguientes etapas de ejecución:

1. Primero se escogen de forma aleatoria k centroides $\mu_1, \mu_2, \dots, \mu_k \in \mathbb{R}^n$
2. Luego, se mide la distancia de una entidad a cada uno de los centroides, asignándose la entidad al centroide con el cual posea la mínima distancia. Para efectos de estos estudios se utilizará la distancia euclidiana:

$$D_{ij} = \min \sqrt{\sum_{k=1}^n (X_{ki} - X_{kj})^2}$$

3. Se realiza lo mismo con cada una de las entidades de la muestra restantes.
4. Una vez que todas las entidades han sido asignadas a un cluster, se calcula el punto medio de cada cluster.
5. El proceso se repite de la misma forma, pero ahora tomando el punto medio de cada cluster como punto de inicio (en vez de los k centroides escogidas al azar al iniciar el algoritmo).
6. El algoritmo continua su iteración hasta que ninguna entidad es reasignada a otro cluster.

2. Cluster X-Means

Una mejora del algoritmo de cluster K-Means, corresponde al algoritmo de cluster X-means de Dan Pelleg y Andrew Moore [4], el cual encuentra el número óptimo de clusters de la muestra (K^*). El beneficio de este algoritmo es que posee una mayor velocidad de procesamiento, selecciona el número óptimo de clúster (evitando la imputación manual del número de clusters) y encuentra un mejor óptimo local que el análisis de cluster normal [4].

Cabe destacar, que para los algoritmos de cluster es necesario que las variables estén normalizadas y estandarizadas para así obtener un mejor rendimiento del modelo.

3. Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

Este es un algoritmo que se basa en la densidad para el descubrimiento de clusters, eliminando el ruido de la muestra. Posee tres métricas fundamentales:

- Epsilon (ϵ): Corresponde al radio de un punto.
- *Min points*: Cantidad mínima de puntos para forma un cluster.
- Densidad Cantidad de puntos en un radio específico (epsilon).

Además los puntos pueden ser clasificados de tres formas:

- Puntos Centrales: Puntos que poseen en su vecindario al menos la cantidad de puntos definida en *min points*.
- Puntos de Borde: Puntos que poseen en su vecindario menos de la cantidad de puntos definida en *min points* pero se encuentran en el vecindario de un punto central.
- Ruido: Puntos que no son puntos centrales ni de borde.

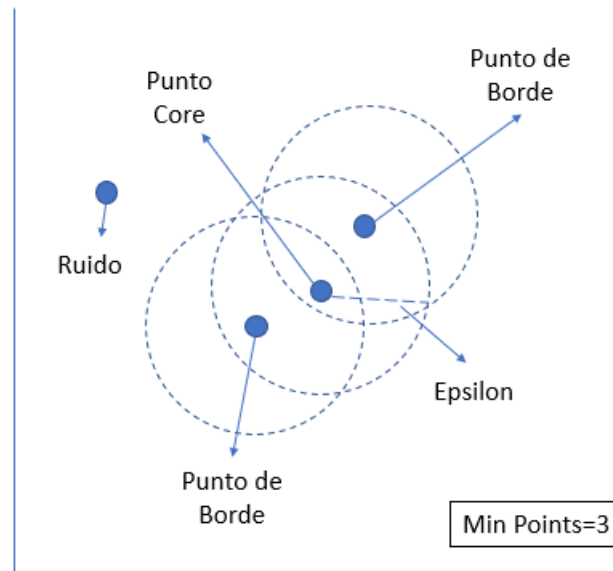


FIGURA 3.7: Puntos Core, Border y Ruido DBSCAN. Fuente: Elaboración propia.

El algoritmo de DBSCAN se puede resumir como:

Para cada punto P de un set de datos y considerando $min\ points = m$, identificar los puntos (pts) en una distancia ϵ :

1. Si $pts \geq m$, etiquetar P como un punto central.
2. Si $pts < m$ y un punto central esta a una distancia n , etiquetar P como un punto de borde.
3. Si $pts < m$, etiquetar P como ruido³.

Las ventajas de este modelo corresponden a que:

³<https://www.kdnuggets.com/2017/10/density-based-spatial-clustering-applications-noise-dbscan.html>

-
- son útiles cuando clusters poseen formas irregulares, están entrelazados o hay ruido/outliers en los datos.
 - no requieren definir un número k de clusters, ya que el algoritmo determina la cantidad de clusters.
 - a diferencia de K-Means y X-means no depende de las condiciones de inicio, ya que no requiere utilizar centroides para realizar los cálculos.

3.4.2.2. Random Forest

A diferencia de los árboles de decisión, Random Forest crea un conjunto de árboles a partir de diferentes subconjuntos de registros y atributos. El motivo de probar este modelo, es que sus resultados son más generalizables que los árboles de decisión, evitando de esta manera el sobreajuste.

Algoritmo

1. Crear un set de datos con bootstrap.
2. Crear un árbol de decisión del set creado a partir de bootstrap, pero seleccionando un set de variables de forma aleatoria.
3. Repite el proceso n veces donde n es igual al parámetro definido por "*number of trees*".
4. Para cada registro, se evalúa el resultado obtenido en cada árbol. Luego se procede a votación, obteniendo como etiqueta del registro aquel que haya obtenido una mayor cantidad de votos.

Ventajas

- Sirve tanto para clasificación como regresión.
- Permite generalizar los resultados, al evitar el sobreajuste, debido a que presenta una mayor variedad al escoger de forma aleatoria las variables y registros para cada árbol.

Desventajas

- A mayor cantidad de árboles, mejor el performance, pero con el costo de generar un procesamiento más lento.
- Es un modelo sólo de carácter predictivo y no descriptivo. Es decir, no se puede saber a partir de Random Forest como interactúan los distintos atributos.

Parámetros

En Random Forest se poseen los mismos parámetros de poda que los árboles de decisión, además de determinar la cantidad de árboles (*number of trees*), la estrategia de votación (*confidence* o *majority vote*) y *guess subset ratio*, el cual determina la cantidad de atributos a seleccionar para cada árbol. Además, se puede elegir el criterio de división de árboles (ganancia de información, índice de gini, *gain ratio*, entre otros). La descripción de cada uno de los parámetros se encuentra en Anexo Parámetros Random Forest.

3.4.3. Validación Cruzada

Para determinar la capacidad de predicción y ajustar los parámetros de los modelos se debe determinar un set de entrenamiento y otro de evaluación:

1. **Set de entrenamiento:** Corresponde a los datos que son utilizados para calibrar los modelos. Generalmente corresponden entre el 67 % y 90 % de la muestra.
2. **Set de evaluación:** Corresponde a los datos son utilizados para probar los modelos. Generalmente corresponde entre el 33 % y 10 % de la muestra.

Surge entonces el problema de decidir que set de datos elegir como set de entrenamiento y de evaluación. El método de validación cruzada, hace alusión a la separación de información en carpetas o secciones de datos. En este sentido, se separa la información en k secciones, donde cada sección es utilizada como set de evaluación, y el resto como set de entrenamiento. Generalmente se divide la muestra en 10 partes ($k=10$). A modo de ejemplo, se puede visualizar en la imagen 3.8 el proceso de validación cruzada para $k=4$.

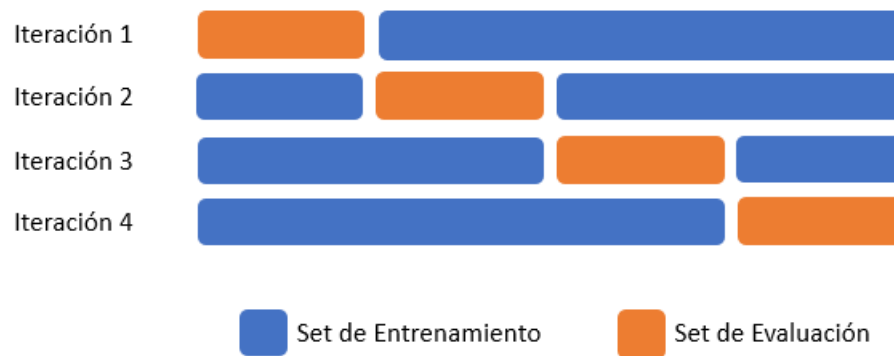


FIGURA 3.8: Validación Cruzada ($k=4$). Fuente: Elaboración propia.

El beneficio de utilizar este método es que se evita el sobreajuste de los datos a la muestra, lo cual provocaría modelos demasiado precisos a los datos disponibles, pero cuyos resultados no son generalizables. Por ejemplo, si hay sobreajuste en la predicción de deserción, los modelos no permitirán predecir de forma correcta la deserción de un alumno nuevo, ya que los parámetros del modelo son aplicables a la muestra y no al universo de estudiantes.

Dado el interés de este estudio que los resultados de los modelos sean generalizables a la población, es que se escoge esta metodología determinar la capacidad de predicción y calibrar los modelos.

3.4.4. Matriz de Confusión

Esta matriz es fundamental para evaluar el desempeño de algoritmos de clasificación, ya que nos permite identificar la cantidad de casos donde el algoritmo realizó una correcta predicción y la cantidad de casos en donde se realizó una predicción incorrecta. Esta matriz posee 4 casos distintos (ver tabla 3.1).

TABLA 3.1: Matriz de Confusión

	Predicción (<i>yes</i>)	Predicción (<i>no</i>)
Actual (<i>yes</i>)	Verdadero Positivo (VP)	Falso Negativo (FN) (Error Tipo II)
Actual (<i>no</i>)	Falso Positivo (FP) (Error Tipo I)	Verdadero Negativo (VN)

Para explicar la matriz de confusión se utilizará como ejemplo la situación donde un modelo predice de forma correcta o incorrecta la deserción de un alumno.

- **Verdadero Positivo (VP):** El modelo predice los casos positivos de forma correcta. En este caso se predice de forma correcta la deserción de un alumno.
- **Verdadero Negativo (VN):** El modelo predice los casos negativos de forma correcta. En este caso se predice de forma correcta que un alumno no desertará.
- **Falso Positivo (FP):** El modelo predice como positivo un caso que en realidad es negativo (Error Tipo I). En este caso se predice que un alumno desertará, pero en realidad no lo hace.
- **Falso Negativo (FN):** El modelo predice como negativo un caso que en realidad es positivo (Error Tipo II). En este caso se predice que un alumno no desertará, pero en realidad si lo hace.

A modo de guía, la nomenclatura utilizada de Verdadero/Falso y Positivo/Negativo corresponden a los valores actuales y predcidos respectivamente.

A partir de la matriz de confusión, se pueden elaborar indicadores para evaluar qué modelo es el mejor para resolver un determinado problema:

- **Accuracy** = $\frac{VP+VN}{N}$. Corresponde al porcentaje de aciertos (verdaderos positivos y verdaderos negativos) con respecto al total de la muestra.

- **Precision** = $\frac{VP}{(VP+FP)}$. Corresponde al porcentaje de verdaderos positivos con respecto al total de valores predecidos como positivos.
- **Recall** = $\frac{VP}{(VP+FN)}$. Corresponde al porcentaje de verdaderos positivos con respecto a los valores que son actuales positivos.
- **F-Measure** = $\frac{2*Recall*Precision}{Recall+Precision}$. Calcula la media armónica entre *Recall* y *Precision*.

Dependiendo del tipo de problema, se deberá determinar que es lo que se quiere priorizar, si tener modelos con alto *Recall*, alta *Precision* o de forma mixta (*F-Measure*). Si lo que se quiere evitar es tener error de tipo I, se utilizarán modelos con una alta *Precision*. Mientras que si lo que se busca es evitar tener error de tipo II, se utilizarán modelos con un alto *Recall*.

3.4.5. ROC Curve y AUC

La curva de ROC (*Receiver Operator Characteristic*) compara el porcentaje de verdaderos positivos (VP) con el porcentaje de Falsos Positivos (FP). Es utilizado para representar el performance del clasificador sin considerar la distribución de la clase o costos de error en la predicción [5]. Se utilizó esta técnica para poder comparar el performance de los distintos modelos de forma gráfica.

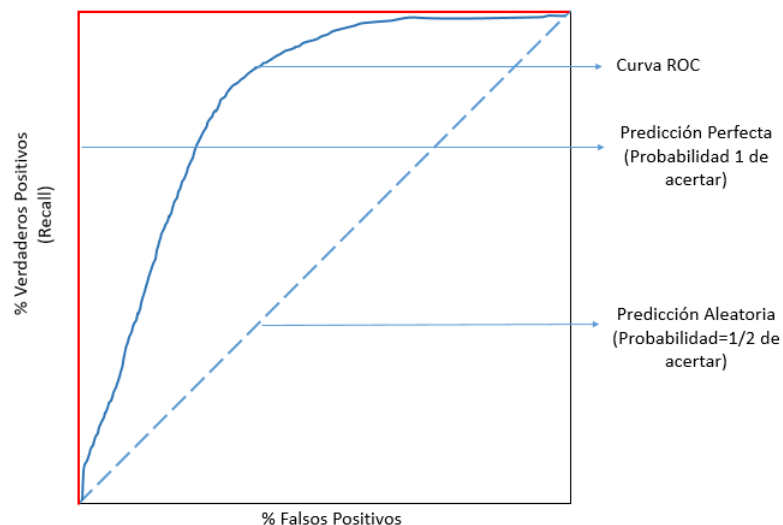


FIGURA 3.9: ROC Curve. Fuente: Elaboración propia.

$$\text{Eje } y = \% VP = \frac{VP}{TP + TN}$$

$$\text{Eje } x = \% FP = \frac{FP}{TN + FP}$$

De forma complementaria se utilizó el Área bajo la Curva ($AUC = \text{Area Under the Curve}$), la cual varía entre 1 y 0,5. Aquellos modelos que posean un AUC equivalente a 1 corresponden a modelos que predicen de forma perfecta y 0,5 a modelos que predicen los resultados de forma aleatoria.

3.5. Concepto de Deserción

Para efectos de este estudio la deserción se define como “El abandono prematuro de un programa de estudios antes de alcanzar el título o grado, y considera un tiempo suficientemente largo como para descartar la posibilidad de que el estudiante se reincorpore” [6]. Esta a su vez se clasifica en cuatro tipos de deserción, según si desertó de forma voluntaria o involuntaria, y si desertó de la carrera o de la institución (ver tabla 3.2).

TABLA 3.2: Tipos de Deserción

I. Deserción de Carrera Voluntaria	III. Deserción Institucional Voluntaria
II. Deserción de Carrera Involuntaria	IV. Deserción Institucional Involuntaria

I. Deserción de Carrera Voluntaria: Corresponde a aquellos alumnos que se cambian de carrera dentro de la institución por admisión nueva a través de cualquiera de sus vías de ingreso, no considerándose desertores aquellos alumnos que se cambian de carrera a través de las vías formales de la Universidad (transferencias internas).

II. Deserción de Carrera Involuntaria: Corresponde a todos aquellos estudiantes que se ven obligados a retirarse de su carrera como consecuencia de una decisión de la universidad debido a desempeño académico insuficiente o a razones disciplinarias, pero que luego ingresan a otra carrera de la Universidad de Chile realizando nuevamente el proceso de postulación.

III. Deserción Institucional Voluntaria: Corresponde a aquellos alumnos que abandonan la Universidad de Chile a través de la formalización del abandono o que no registran matrícula en el periodo siguiente en la Universidad.

IV. Deserción Institucional Involuntaria: Corresponde a aquellos alumnos que abandonan la Universidad de Chile, tras haber sido expulsados de una carrera como consecuencia de desempeño académico insuficiente o razones disciplinarias.

3.5.1. Estudios Relacionados

En esta sección se detalla la bibliografía consultada para el desarrollo de este trabajo de tesis en materias de deserción estudiantil e implementación de analítica en organizaciones.

3.5.1.1. Modelos de Deserción Estudiantil

El estudio titulado como “Modelo de análisis de la deserción estudiantil en la educación superior” de Erika Himmel año 2002 [6], aborda el tema de deserción y retención de estudiantes de la educación superior desde una perspectiva conceptual. En este estudio se propone un concepto de deserción (el cual fue utilizado dentro de este trabajo de tesis, ver Concepto de Deserción) y sintetiza diversos enfoques teóricos empleados en la investigación de este tema, los cuales pueden ser agrupados en 5 grandes categorías: psicológicos, económicos, sociológicos, organizacionales y de interacciones (si bien pueden ser agrupados de esta manera, existen modelos que integran más de una de estas categorías).

1. Enfoques psicológicos

Estos modelos plantean que existen rasgos de personalidad que influyen en la deserción. Uno de los primeros modelos corresponde a los de Fishbein y Ajzen (1975) quienes sugieren que las intenciones de una persona son el resultado de sus creencias, las que influyen sobre sus actitudes y la llevan a manifestar un determinado comportamiento. En este sentido, la decisión sobre desertar se ve influenciada por conductas previas, las actitudes hacia la deserción y/o persistencia y por norma subjetivas hacia estas acciones, llevando a la formación de una intención conductual, lo que lleva a adoptar un determinado comportamiento. Si estas conductas, actitudes y normas son contrarias a permanecer en la universidad, el alumno terminará

desertando. En contraste, si estas son positivas, el alumno se mantendrá dentro de sus estudios. Este modelo fue posteriormente modificado por Attinasi (1986) postulando que la deserción estudiantil se ve influida por las percepciones de los alumnos y los análisis que estos realizan de su vida universitaria después de su ingreso. Ethington (1990) construyó un modelo considerando todo lo anterior, incorporando una teoría más general sobre las conductas de logro de Eccles et al (1983), dentro de las cuales se encuentran atributos como la perseverancia, la elección y el desempeño. Este pudo observar que los valores como las expectativas de éxito influyen sobre la persistencia en la universidad.

2. Enfoques Sociológicos

Enfatizan la influencia de factores externos al individuo en la retención, de forma adicional a los enfoques psicológicos. El modelo de Spady (1970), uno de los más citados en cuanto a deserción universitaria, se basa en la teoría del suicidio de Durkheim (187-1951), quien plantea que el suicidio es consecuencia de la ruptura del individuo y el sistema social por su imposibilidad de integrarse a la sociedad. Esto ocurre cuando existe una baja conciencia moral y bajo apoyo de las relaciones sociales. En este sentido, Spady relaciona a la deserción con la baja integración de los estudiantes con el entorno de educación superior. Este mismo también afirma que el medio familiar es una de las muchas fuentes que expone a los estudiantes a influencias, expectativas y demandas, las que afectan su nivel de integración social en la universidad.

3. Enfoques económicos

Esta categoría de modelos adopta una perspectiva de costo beneficio. De esta manera, se postula que cuando los beneficios sociales y económicos asociados a los estudios en educación superior son percibidos como mayores que realizar actividades alternas, como por ejemplo el trabajo, el estudiante decidirá permanecer en la universidad.

En esta misma línea se encuentran los enfoques que examinan la asignación de subsidios para los estudiantes, como rebajas en la matrícula, becas y créditos con bajo interés, los cuales buscan equiparar la capacidad de pago de los estudiantes e incrementar la apreciación de los beneficios de la educación universitaria (Cabrera, Nora y Asker 1999).

4. Enfoques organizacionales

Esta categoría de modelos se centra en la deserción según las características de la institución de educación superior, en cuanto a los servicios que esta ofrece a los estudiantes que ingresan

a ella. Es de relevancia en este enfoque la calidad de la docencia (Braxton, Millen, Sullivan, 2000) y de la experiencia de los estudiantes en el aula. También se ha incorporado dentro de este tipo de estudios efectos de beneficios de salud; actividades complementarias como deportes, actividades culturales y apoyos académicos; disponibilidad de recursos bibliográficos; laboratorios; e indicadores como cantidad de alumnos por profesor (Tillman, 2002).

5. Enfoques de interacciones

En este tipo de enfoque resaltan los estudios de Vincent Tinto, uno de los autores con mayores influencias sobre deserción y retención. Tinto expande el modelo de Spady con la teoría de intercambio de Nye (1979). La teoría de intercambio se fundamenta en que los seres humanos evitan las conductas que les produce un costo de algún tipo, y buscan las recompensas en las relaciones, interacciones y estados emocionales (Nye, 1979). En este sentido, Tinto plantea que los estudiantes actúan de acuerdo a la teoría de intercambio en la construcción de su integración social y académica. Si los beneficios de permanecer en la institución son mayores que los costos personales (esfuerzo, dedicación, entre otros) el estudiante permanecerá en la institución. En cambio, si hay otras actividades que le reporten mayores recompensas, este desertará. El modelo considera que existen diversos factores que refuerzan el permanecer en la institución que seleccionó, los cuales comprenden antecedentes familiares (ej. nivel socio-económico y cultural de la familia, valores que sustenta), atributos personales y experiencia académica preuniversitaria.

3.5.1.2. Un modelo analítico para la predicción del rendimiento académico de estudiantes de ingeniería

Estudio realizado en la Facultad de Ingeniería Universidad de Chile en el año 2015 [7], considerando alumnos desertores por eliminación debido a bajo rendimiento académico.

El objetivo del estudio fue mostrar como las herramientas de minería de datos pueden ser utilizadas para generar modelos predictivos que sirvan para apoyar a aquellos estudiantes en riesgo de deserción o insuficientes desempeño académico. Para ello se utilizaron datos de alumnos de primer año del Plan Común de Ingeniería y Ciencias de la Universidad de Chile de los cohortes 2010, 2011, 2012, 2013 y 2014.

Las variables utilizadas fueron variables individuales de los alumnos (género, tiempo desde el egreso de enseñanza media y región de procedencia), variables de preingreso (tipo de establecimiento educacional enseñanza media, experiencias previas en educación superior, puntajes en la psu, vías de ingreso especiales), ranking, promedio de notas de enseñanza media y la creación de 11 variables adicionales a partir de las notas de los alumnos.

Se utilizó la metodología de minería de datos para predecir la caída en causal de eliminación por motivos académicos, utilizando el modelo de regresión logística en combinación con método de selección de atributos *Forward Selection and Backward Elimination*. Los modelos fueron entrenados con datos de 2010 a 2013 y probados con datos de 2014. El poder predictivo fue evaluado con las medidas de *Recall* y *Precision*.

El modelo desarrollado clasificó correctamente a más del 86% de los casos, con niveles bajos de error tipo II y una precisión de 38%, permitiendo desarrollar intervenciones focalizadas sobre los alumnos en el segundo semestre. La variable con mayor poder explicativo corresponde al ratio de créditos inscritos aprobados, es decir, a mayor cantidad de cursos reprobados en el primer semestre, mayores son las probabilidades de reprobado por segunda vez. Otra variable significativa corresponde a género, donde un estudiante hombre posee un 88% más probabilidades de reprobado dos veces el mismo ramo que una mujer (y por ende caer en causal de eliminación académica). La tercera variable significativa indica que aquellos estudiantes que superen la nota de aprobación (4.0) en los primeros controles tienen menores probabilidades de volver a reprobado una asignatura que aquellos que no.

Se utilizó este estudio como referencia para este proyecto para determinar qué variables son importantes considerar para estudiar la deserción de alumnos de la Universidad de Chile, qué método de selección de variables emplear (*Forward Selection and Backward Elimination*) y qué indicadores utilizar para medir el rendimiento de los modelos (*Recall* y *Precision*).

3.5.1.3. El impacto del tipo de financiamientos sobre la probabilidad de retención de primer año en la educación superior - El caso de la Universidad de Chile.

Este estudio fue realizado en la Dirección de Bienestar de la Universidad de Chile en el año 2017 [8], considerando alumnos desertores tanto de carrera como institucionales.

El objetivo de esta investigación fue medir el impacto que causan las ayudas financieras de educación superior sobre la probabilidad de retención de los estudiantes de la Universidad de Chile para lo cual se utilizaron datos de alumnos de primer año pertenecientes a los cohortes 2013, 2014, 2015 y 2016.

Para analizar los datos se construyeron tres tipos de modelos probabilísticos, el primero para estudiar la retención estudiantil, el segundo para notas y el tercero para avance académico. En el primero se utilizó el método de máxima verosimilitud para calcular los parámetros, mientras que en el segundo y el tercero el método de Mínimos Cuadrados Ordinarios (MCO).

Los principales resultados fueron que aquellos alumnos que poseen becas o Crédito con Aval del Estado (CAE) cuentan con una mayor probabilidad de retención. En particular, aquellos alumnos que poseen gratuidad o que poseen una combinación de becas de mantención y de arancel estatales poseen una probabilidad de retención de mayor magnitud.

Este estudio fue utilizado como referencia para comprender el impacto de las becas de arancel, becas de mantención y créditos estatales en la deserción de alumnos de la Universidad de Chile, con el objetivo de determinar qué variables incluir en los modelos predictivos.

3.5.1.4. Impacto de los distintos tipos de ayuda financiera gubernamental e institucional en la perseverancia de los estudiantes de la Universidad de Chile.

Este estudio fue realizado en la Universidad de Chile en año 2015 [9], considerando alumnos desertores tanto de carrera como institucionales.

El objetivo de este proyecto fue determinar de forma cuantitativa la influencia de los distintos tipos de ayuda financiera en la probabilidad de persistencia de los estudiantes de la Universidad de Chile. Para ello se utilizaron datos pertenecientes a la Universidad de Chile, a saber, admisión y matrícula, ayuda financiera y admisión PSU (DEMRE) pertenecientes a los cohortes de estudiantes de primer año 2009 y 2010.

Se utilizó el modelo cuantitativo de supervivencia de Chen y DesJardins (2010), la cual permite determinar si los factores tienen efectos que cambian en función del tiempo y cómo varios factores se relacionan con la aparición de un evento en diferentes momentos.

Los resultados obtenidos fueron que las combinaciones de becas y préstamos estatales, institucionales

y/o externas no estatales, aumentan la probabilidad de permanecer en la Universidad. En particular, las ayudas de mantención poseen una mayor preponderancia en la persistencia de los estudiantes que las ayudas de arancel. Resalta también la primacía del Crédito con Aval del Estado (CAE) por sobre las becas en reducir la probabilidad de deserción. Destacan las ayudas brindadas por entidades externas no estatales a la universidad (fundaciones, municipios, entidades bancarias, etc), seguido de las ayudas institucionales (internas) que corresponden a becas de mantención no cubiertas por el estado y becas que exigen mérito académico para acceder a ellas, como Andrés Bello y Beca Universidad de Chile. También se destaca que a mayor quintil de ingreso, disminuye el riesgo de persistencia. Por último, que los resultados Puntaje Promedio PSU (Promedio PSU Lenguaje y Matemáticas) y Puntaje NEM no contribuyen a la persistencia en la universidad.

Este estudio fue utilizado para comprender la influencia en la retención de alumnos de las distintas tipos de ayuda financiera, de los quintiles de ingreso y la Prueba de Selección Universitaria (PSU), con el objetivo de determinar las variables a considerar en los modelos predictivos.

3.5.1.5. Determinantes de deserción en la educación superior chilena, con énfasis en efecto de becas y créditos.

Estudio realizado en la Universidad Católica de Chile en año 2013 [10], considerando la persistencia a nivel de sistema de educación superior, por tipo de institución (Universidad Privada, Universidad del Cruch, Centro de Formación Técnica e Instituto Profesional) y en la misma institución.

El objetivo del estudio es estudiar el rol de distintas variables en la deserción del sistema de educación superior en Chile. Para ello se utilizaron las siguientes fuentes de datos: DEMRE, Servicio de Información de Educación Superior (SIES), ayudas estudiantiles del Ministerio de Educación y Comisión Ingreso de alumnos que se inscribieron para rendir la PSU en los procesos 2007, 2008, 2009 y 2010.

Se realizaron análisis descriptivos, de regresión y de propensity score matching, donde la última metodología busca controlar el fenómeno de endogeneidad presente en la asignación de becas, ya que se plantea que la asignación de estas no corresponde a un fenómeno aleatorio, sino que al cumplimiento de requisitos mínimos para acceder a estas (académicos y sociales principalmente).

Uno de los principales resultados corresponde a que la deserción se encuentra relacionada con el nivel

socioeconómico, observándose mayor deserción en grupos de menores ingresos, de menor educación de los padres y entre egresados de establecimientos subvencionados o particulares. Además, se observa una relación positiva entre PSU y NEM con una mayor persistencia. El Crédito con Aval del Estado (CAE), Crédito de Fondo Solidario (CFS) y becas de excelencia académica (BEA y Juan Gómez Millas) están correlacionados con un mayor nivel de persistencia, siendo el CAE el que posee una mayor relación con la persistencia de todas las variables. No se encontró evidencia suficiente de que las Becas Vocación de Profesor y Beca hijo del profesor estén asociados a una mayor persistencia.

Este estudio fue utilizado para comprender la influencia en la deserción y endogeneidad presente en becas arancel, becas de mantención y créditos estatales, con el objetivo de determinar qué becas considerar dentro de los modelos predictivos y cómo tratar aquellas que poseen sesgo de selección para ser asignadas, ya sea por rendimiento académico o nivel socioeconómico.

3.5.1.6. Revisión estudio “*Competing on Analytics*”

El centro de investigación de *Babson College’s Working Knowledge Research Center* [11], realizó un estudio a 32 organizaciones que han implementado métodos analíticos complejos, dentro de las cuales 11 de ellas eran competidores completamente basados en analítica, es decir, que declaraban como una de sus estrategias claves el desarrollo de la analítica en la organización. El objetivo de este estudio era determinar qué características y prácticas son claves para competir bajo este escenario. Los resultados arrojaron que existen 3 atributos claves que deben ser llevados a cabo:

1. Uso extendido de optimización y modelos matemáticos

Las empresas que han implementado analítica dentro de sus organizaciones desarrollan modelos más allá de estadísticas descriptivas, como modelos que permitan pronosticar la demanda de productos o servicios, para así mantener una dotación de personal y un stock de productos adecuados, o modelos que permitan predecir el riesgo de perder un cliente para así crear estrategias para retenerlo. Esto va de la mano de continua experimentación de estrategias que permitan intervenir y alcanzar los resultados deseados.

2. Visión transversal de la analítica en la organización

Comprensión de que la mayoría de las funciones de negocio pueden ser abordadas con analítica, ya sea

las áreas de recursos humanos, marketing, operaciones, postventa, entre otros, generando una puesta en práctica global en el uso de datos para la toma de decisiones. Es importante que las iniciativas analíticas de relevancia para la organización sean tratadas bajo un “único carril”, manteniendo un liderazgo, herramientas y tecnologías en común. De esta manera se busca evitar y eliminar los esfuerzos aislados en las unidades, ya que esto provoca la generación de múltiples versiones de indicadores, ya que cada unidad tiende a mantener sus propias bases de datos, herramientas y personal para trabajar la información. Lo que se apunta es a asegurar que los datos críticos y otros recursos sean bien administrados, de forma de que estas sean compartidas de forma fácil, y sin los impedimentos de inconsistencia de formato, definiciones y estándares.

3. Prioridad por parte de los altos ejecutivos

Se requiere que altos ejecutivos comprendan y transmitan la importancia de los métodos cuantitativos para la toma de decisiones, ya que se requiere generar cambios en la cultura, procesos, comportamientos y habilidades de los empleados. Si bien no es necesario que los altos ejecutivos posean una formación de estadística, si es necesario que comprendan los distintos métodos analíticos y conozcan sus limitaciones. De todas maneras, en caso de no poseer una mayor comprensión en métodos analíticos, siempre está la posibilidad de asesorarse por compañías externas a través de consultorías.

Es relevante dentro de estas organizaciones, además del desarrollo de la analítica:

1. Mantener el foco correcto

Es importante destinar energías a resolver problemas que son relevantes para la organización, evitando la creación de iniciativas que provoquen que los esfuerzos se tornen difusos y que se pierda el foco de negocio al profundizar en cada una de ellas.

2. Cultura correcta

Se debe mantener una cultura que valore el medir, probar y evaluar en base a evidencia cuantitativa, instando a los empleados a tomar decisiones en base a datos. Esto debe ser potenciado en todo momento por los ejecutivos principales. De todas maneras, cuando se deban tomar decisiones rápidas o cuando no se puede recopilar información de calidad, se debe priorizar la toma de decisiones con la información disponible.

3. Personas correctas

Se debe velar por emplear a las personas con amplios conocimientos en analítica y de negocios, sobre todo a aquellos que ocuparán altos cargos. Además, estos deben ser capaces de explicar de forma sencilla ideas complejas, y poseer las habilidades necesarias para interactuar con los tomadores de decisión. Si bien estos cargos son difíciles de encontrar, se debe velar en todo momento por traer a los mejores independiente del tiempo que tomen los procesos de selección.

4. Tecnología correcta

Se debe velar por mantener sistemas que permitan recopilar información de cada movimiento de la organización. Para ello existen sistemas como ERP, CRM, puntos de ventas, entre otros. Además, se debe contar con repositorios de datos que permitan integrar la información de distintos sistemas de forma correcta, facilitando que el acceso a información única a todos los empleados.

Capítulo 4

Planteamiento Estratégico

En este capítulo se presenta la información estratégica de la Universidad de Chile y la Dirección de Bienestar Estudiantil (DBE). Además se realiza un análisis PEST y se detalla el modelo de negocios de la DBE.

4.1. Balanced Scorecard

Actualmente la Universidad de Chile no cuenta con una herramienta de Balanced Scorecard para su gestión, pero si posee declarados sus objetivos estratégicos dentro de su Plan de Desarrollo Institucional (PDI) 2017-2026. Estos corresponden a:

- **Docencia:** Desarrollo y crecimiento de un cuerpo académico, estudiantil y funcionario de destacada vocación, calidad, responsabilidad y compromiso, para la formación integral de profesional, científicos, investigadores y creadores artísticos de la más alta calidad y pertinencia.
- **Investigación y creación artística:** Desarrollo de la investigación, creación y producción artística e innovación al más alto nivel, articuladas con los problemas estratégicos y necesidades del país.
- **Gobierno, gestión y presupuesto:** Consolidación del gobierno, impulso de la gestión transversal y cambio en el paradigma de la distribución presupuestaria de la Universidad de Chile.

-
- **Extensión y comunicaciones:** Desarrollo de la extensión y las comunicaciones, que valores y fortalezca la interacción y el compromiso entre la universidad y la sociedad.
 - **Internacionalización:** Desarrollo de una estrategia institucional de internacionalización de la Universidad de Chile.
 - **Educación:** Potenciar el área de educación y su vinculación con establecimientos pedagógicos a través de una política integrada.

A modo de ejercicio, se realiza un Balanced Scorecard de la Universidad de Chile en base a los objetivos planteados por el PDI.

1. **Perspectiva financiera**

- Generar sustentabilidad financiera con el fin de asentar las bases de autonomía, calidad y competitividad académica.
 - Aumentar tasa de crecimiento de matrícula en programas de pregrado, postgrado y doctorado.
 - Aumentar fondos percibidos por financiamiento de investigación.
 - Aumentar fondos percibidos por extensión del conocimiento.

2. **Perspectiva de estudiantes**

- Mejorar la calidad de los profesionales egresados.
 - Aumentar la tasa de crecimiento de egresados de magíster y doctorados de la Universidad de Chile.
 - Aumentar la proporción de profesores con magíster y doctorado.
 - Aumentar la tasa de empleabilidad de egresados de la Universidad de Chile.
- Aumentar la matrícula en programas universitarios.
 - Aumentar la matrícula de estudiantes nacionales.
 - Aumentar la matrícula de estudiantes extranjeros.
- Disminuir la deserción en programas universitarios
 - Disminución tasa de deserción de pregrado.

-
- Disminución tasa de deserción postgrado.
 - Disminución tasa de deserción doctorado.

3. Perspectiva interna

- Sistematizar procesos críticos relacionados a alumnos, académicos y gestión administrativa.
- Incentivar y mejorar el desarrollo de programas académicos.
- Fortalecer la vinculación, alianzas y convenios.

4. Perspectiva de aprendizaje y desarrollo

- Implementar tecnologías de información en actividades docentes, de alumnos y de gestión administrativa.
- Mejorar las competencias de personal de colaboración.
- Fortalecer la efectividad y compromiso del cuerpo docente.
- Mejorar infraestructura universitaria y equipamiento.

4.2. Posicionamiento Estratégico Universidad de Chile

Según el Modelo Delta de HAX [2], la Universidad de Chile posee un posicionamiento estratégico de Diferenciación y de Solución Integral al Estudiante.

■ Diferenciación

La Universidad de Chile busca ofrecer a sus alumnos una educación universitaria con parámetros internacionales de excelencia académica, para lo cual realiza esfuerzos por contar con un cuerpo académico de excelencia, fortalecer la calidad y pertinencia de las carreras de pregrado a través de políticas que permitan mantener altos estándares de docencia, potenciar programas de doctorado para que sean competitivos internacionalmente, contribuir a las necesidades de fortalecimiento académico a través de la formación de doctores, programas para atraer y retener a académicos con grado de doctor, entre otros¹.

¹Proyecto de Desarrollo Institucional 2006

■ Solución Integral al Estudiante

La Universidad de Chile asume la responsabilidad de acompañar al estudiante en toda su trayectoria en la universidad (acceso, progreso y egreso), la cual se encuentra enmarcada bajo la política de Equidad e Inclusión Estudiantil del año 2014. En particular, la DBE contribuye en esta estrategia identificando las necesidades de estudiantes que provienen de contextos vulnerables, generando programas que permitan satisfacer sus requerimientos.

4.3. Política de Equidad e Inclusión Estudiantil

En el año 2014 el Senado Universitario aprobó la política de Equidad e Inclusión Estudiantil, creando de esta forma un marco regulatorio para las estrategias, programas y acciones de equidad e inclusión en la Universidad de Chile. Esta política considera firmemente que la formación en diversidad juega un rol relevante en los procesos educativos de calidad y que el valor del mérito que caracteriza a la Universidad de Chile debe ser protegido, estableciendo distintos criterios de excelencia según la realidad de cada estudiante.

Uno de los lineamientos más relevantes de esta política corresponde al Modelo de Desarrollo Integral del Estudiante (MDIE).

4.3.1. Modelo de Desarrollo Integral del Estudiante

Este modelo fue creado con el objetivo de contribuir a la igualdad de oportunidades durante la permanencia del alumno en la Institución. Este considera al estudiante como un ser complejo con múltiples necesidades y potencialidades, y está centrado en tres dimensiones que impactan en el logro académico, a saber: calidad de vida, aprendizaje y enseñanza (ver figura 4.1).



FIGURA 4.1: Modelo de Desarrollo Integral del Estudiante. Fuente: Política de Equidad e Inclusión Estudiantil 2014.

Dimensión Calidad de Vida

Abarca la caracterización socioeconómica, asignación de beneficios estudiantiles como el Programa de Residencias Universitarias, Programa Preescolar de apoyo a estudiantes madres o padres, Programa de Apoyo Económico (PAE), Beca de Apoyo al Estudiante (BAE), Programa Estudiantil para la Retención (PER), etc), atención médica, dental, psicología clínica y fomento de la participación estudiantil. Esta dimensión se encuentra a cargo de la Dirección de Bienestar Estudiantil.

Dimensión Aprendizaje

Considera iniciativas de apoyo al logro académico, realizadas fuera de la sala de clases y no dependientes de forma directa de los profesores. Incluye acciones de diagnóstico y caracterización académica, programas de reforzamiento grupal, tutorías y mentorías que promueven la motivación, modelos de estudio, estándares de actuación disciplinar y relaciones interpersonales, entre otros.

Dimensión Enseñanza

Se enfoca en el desarrollo docente e incluye programas de formación pedagógica para la educación superior e iniciativas de acompañamiento en la implementación de innovaciones, fomentando que el docente conozca a sus estudiantes, utilice métodos e instrumentos de evaluación pertinentes y diversos, y retroalimente oportunamente sobre logros y dificultades en el aprendizaje.

4.3.2. Ámbitos de acción del Modelo de Desarrollo Integral del Estudiante

Para el MDIE se establecen ámbitos de acción en las distintas etapas del estudiante en la universidad (acceso, ingreso y egreso). Los ámbitos de acción que se vinculan con el accionar de la DBE corresponden a:

TABLA 4.1: Ámbitos de acción MDIE

Etapa	Acciones
Ingreso	Crear programas que incentiven y permitan el acceso a estudiantes de establecimientos de alta vulnerabilidad social, asegurando el acceso a aquellos estudiantes que cumplen de forma exitosa los programas, y que la elección de la carrera sea realizada de forma informada y con previa orientación vocacional.
	Adaptar las estructura de ingreso a la universidad en reconocimiento del mérito de los estudiantes en contexto, en concordancia con la equidad en la admisión al pregrado.
	Fortalecer medidas complementarias de acceso (como SIPPE u otros sistemas).
Progreso	Desarrollar mecanismos que integran la experiencia y los conocimientos de los estudiantes de atención prioritaria.
	Fomentar la participación de los estudiantes en la vida universitaria.
	Resguardar el acceso a las herramientas necesarias para actividades académicas, como el uso de laboratorio de computación y servicios de impresión de documentos, contando con becas que permitan solventar estos gastos en caso de ser necesario.
	Velar por el acompañamiento efectivo de los estudiantes de atención prioritaria en cuanto a postulaciones y entrega de beneficios junto a la detección y resolución de conflictos.
	Velar por la protección de la maternidad y paternidad generando programas para este fin.
	Incentivar el trabajo colaborativo entre estamentos de la comunidad universitaria.
	Considerar cupos preferentes en salas cunas y programas de apoyo preescolar, para hijos de estudiantes de atención prioritaria.
<i>Continúa en la siguiente página</i>	

Continuación de Tabla	
Etapa	Acciones
	Operar bajo un principio de universalidad (servicios deben ser útiles para todos los estudiantes de la universidad).
Egreso	Velar por la igualdad de oportunidad de egreso de sus estudiantes, con particular atención en aquellos que tengan menor capital social o necesidades específicas que complican la inserción en el medio laboral.
	Fomentar la realización de acciones necesarias para hacer efectiva la titulación oportuna y el acceso a oportunidades de empleo en su área de formación.

4.4. Visión, Misión y Objetivos Estratégicos DBE

Actualmente la DBE tiene declarado sus objetivos estratégicos, pero no así su misión y visión. Es por ello que en base a los objetivos estratégicos, y a información adicional sobre los desafíos y acciones realizadas en el área, se propone a modo de ejercicio una misión y visión para la dirección.

4.4.1. Visión

Ser un referente en la formación de profesionales con un alto estándar de calidad bajo un contexto de equidad e inclusión al 2020.

4.4.2. Misión

Proveer programas de apoyo a estudiantes de contextos vulnerables con el objetivo de que puedan desarrollar su máximo potencial social, intelectual y espiritual durante el desarrollo de sus estudios, contribuyendo de esta manera a la formación de profesionales con una educación de calidad independiente de su origen social y económico.

4.4.3. Objetivos Estratégicos

1. Generar lineamientos para avanzar en la equidad interna y atención de la diversidad en materia de beneficios, en sinergia con las facultades y la política pública.
2. Modernizar, articular y aumentar la pertinencia en la asignación y gestión de las ayudas, apoyando la retención y el progreso académico de los estudiantes.
3. Producir información sistemática de asignaciones y perfilar a quienes recurren a las ayudas estudiantiles, para contribuir al desarrollo de las políticas y programas institucionales y nacionales de soporte a la inclusión en la vida universitaria.

4.5. Análisis PEST DBE

Esta herramienta es utilizada para analizar el entorno externo de una organización, unidad de negocio o proyecto, facilitando el detectar amenazas y oportunidades para anticiparse a eventos que podrían ser perjudiciales y captar las oportunidades de mercado. El análisis PEST considera 4 factores de estudio: Político, Económico, Socio-Cultural y Tecnológico. Estos aspectos son detallados a continuación con respecto a la Dirección de Bienestar Estudiantil.

4.5.1. Político y Económico

Se detalla a continuación las políticas externas por las cuales se rige la Dirección de Bienestar Estudiantil, las cuales traen consigo consecuencias económicas para el financiamiento de la universidad. Dentro de las consecuencias económicas de estas leyes se encuentran la definición de aranceles regulados, término de limitaciones para el crecimiento de la matrícula y financiamiento estatal para crear o fortalecer planes de apoyo para la permanencia de estudiantes.

4.5.1.1. Gratuidad

En 2018 se promulgó la Ley 21.091 sobre Educación Superior² en Chile, la cual en su título V establece el Financiamiento Institucional para la Gratuidad, garantizando de esta manera estudios gratuitos a aquellos alumnos de mayor vulnerabilidad socioeconómica.

Para la entrega de este beneficio, las instituciones adscritas a gratuidad deben cumplir con los siguientes requisitos:

1. Contar con acreditación institucional avanzada o de excelencia.
2. Estar Constituidas como personas jurídicas de derecho privado sin fines de lucro.
3. Estar adscritas al Sistema de Acceso a las Instituciones de Educación Superior y cumplir con los criterios de selectividad para las universidades que reciban este financiamiento, el cual está basado en desempeños mínimos que deben tener los estudiantes matriculados en primer año.
4. Aplicar políticas que permitan el acceso equitativo de estudiantes y contar con programas de apoyo a estudiantes vulnerables que promuevan su retención, fomentando que al menos el 20 % de la matrícula total de la institución corresponda a estudiantes de hogares pertenecientes a los cuatro primeros deciles de menores ingresos del país³.

En cuanto a financiamiento, la gratuidad establece aranceles regulados, los cuales corresponden al valor monetario que se le asignará a las carreras que serán financiadas por gratuidad, dinero que será entregado por parte del Estado a las Instituciones de Educación Superior para financiar los estudios de aquellos alumnos que hayan sido beneficiados.

Para determinar el valor de estos aranceles regulados, la Subsecretaría de Educación Superior evalúa y crea propuestas de los costos directos e indirectos asociados a cada una de las carreras. Estas propuestas son discutidas con las instituciones de educación superior adscritas a gratuidad y sus respectivas federaciones de estudiantes. Una vez llegado a una propuesta final, esta es evaluada por una Comisión de Expertos creada para estos fines, quienes tomarán la decisión final sobre los montos de los aranceles regulados.

²<https://www.leychile.cl/Navegar?idNorma=1118991>

³<https://www.leychile.cl/Navegar?idNorma=1118991>

Esta ley también establece vacantes máximas de estudiantes de primer año para las instituciones de educación superior que reciban el financiamiento institucional para la gratuidad. Esto es determinado para cada institución considerando los niveles y años de acreditación institucional, cobertura regional, entre otros. De todas maneras, estas pueden solicitar aumentos en las vacantes máximas, siempre que sus solicitudes estén fundamentadas en necesidades estratégicas del país y que ya haya sido contemplada en la elaboración de los Planes de Desarrollo Institucional.

La cobertura de gratuidad corresponde la duración nominal de la carrera del estudiante, y en caso de que este exceda dicho tiempo, durante el primer año de exceso la Institución de Educación Superior sólo podrá cobrar la mitad del arancel regulado y los derechos básicos de matrícula, y en el segundo año de exceso se limitará a cobrar el valor del arancel regulado al estudiante (y no el establecido por la institución de educación superior). En caso de cambio de carrera, se considera la duración nominal de la carrera en curso, descontándose el tiempo que el estudiante ya haya cursado con gratuidad en la carrera anterior.

4.5.1.2. Ley de Universidades Estatales

En el mismo año de promulgación de la Gratuidad a través de la Ley sobre Educación Superior, se publica la Ley 21.094 sobre Universidades Estatales ⁴, la cual establece un marco común para universidades del estado para el cumplimiento de su quehacer. De esta ley, es relevante rescatar para el quehacer de la Dirección de Bienestar Estudiantil los siguientes puntos:

■ Coordinación

La ley indica que las universidades del estado deben actuar bajo el principio de coordinación, con el propósito de fomentar una labor conjunta de forma de contribuir al progreso del país y elevar los estándares de la educación pública (artículo 50).

■ Crecimiento de Matrícula

La ley establece que las universidades del estado podrán crear planes de crecimiento de su oferta académica o de su matrícula, los cuales deberán basarse en objetivos estratégicos de las necesidades del país, indicadores objetivos, considerar mecanismos de equidad e inclusión

⁴<https://www.leychile.cl/Navegar?idNorma=1119253>

para los nuevos estudiantes y estar contemplados en los planes de desarrollo institucional. Estos planes no se rigen según las vacantes máximas establecidas por la política de gratuidad, siempre que sean aprobados previamente por decreto del Ministerio de Educación y suscrito además por el Ministerio de Hacienda (artículo 62).

■ **Plan de Fortalecimiento Institucional**

Con el objetivo de apoyar el desarrollo institucional de las universidades del estado, se creó un plan de fortalecimiento transitorio, cuya duración es de diez años, destinados a usos estratégicos de las universidades del estado (Artículo 58). Este plan contempla recursos por un monto de \$300 mil millones de pesos, donde se deberá destinar al menos la mitad dentro de los 5 primeros años (artículo 60). Las iniciativas que serán financiados con estos fondos corresponderán a:

1. Mejorar y actualizar el Plan de Desarrollo Institucional.
2. Fortalecer la la gestión institucional.
3. Financiar el crecimiento de la oferta académica y de matrícula.
4. Fortalecer la calidad académica y fomentar la incorporación de académicos con grado de doctor.
5. Fortalecer la investigación e incidencia en la elaboración e implementación de políticas públicas, como desarrollo sustentable, cambio climático, sismología, cuidado y protección de niños y adultos mayores, inclusión y no discriminación, planificación urbana sostenible, entre otras.
6. Elaborar programas y acciones de vinculación con el medio que promuevan el desarrollo regional, la interculturalidad, el respeto de pueblos originarios y el cuidado del medio ambiente.
7. Conservar y mejorar la infraestructura.
8. Crear o fortalecer planes de apoyo para la permanencia o titulación de estudiantes.
9. Apoyar la obtención de la acreditación institucional. (artículo 62)

4.5.1.3. Sistema de Financiamiento Solidario

El actual gobierno posee el proyecto de ley que crea el Sistema de Financiamiento Solidario, el cual es un crédito otorgado por un sociedad anónima estatal que financia a través de crédito los estudios

de alumnos de Instituciones de Educación Superior de los primeros 9 deciles, reemplazando el actual método de financiamiento del Crédito con Aval del Estado (CAE) y el Fondo Solidario de Crédito Universitario (FSCU).

La cobertura del crédito corresponde al financiamiento del arancel regulado más un porcentaje del arancel real (el cual generalmente es superior al arancel regulado), lo cual dependerá del nivel de vulnerabilidad del alumno y el nivel de acreditación de la carrera. El tope de financiamiento de este crédito corresponderá a 1,5 veces el arancel regulado.

Los alumnos que pertenezcan a los 6 primeros deciles y que posean el crédito, estarán exentos de pago durante la realización de sus estudios, siendo la propia institución de educación superior la que se encargue de financiar la diferencia con el arancel real por medio de becas y/o créditos propios⁵.

4.6. Modelo de Negocios DBE

Un modelo de negocios sirve para expresar de forma lógica quiénes son los clientes, qué valoran y cómo se articulan los procesos y recursos que tengan resultados económicos positivos. A continuación se detalla el modelo de negocios de la DBE.

4.6.1. Propuesta de Valor

La DBE facilita las condiciones para que estudiantes de contextos vulnerables puedan completar sus estudios, considerando la permanencia y el egreso oportuno de la carrera. Para esto, focaliza y gestiona la ayudas dependiendo de la situación detectada contribuyendo por ejemplo a padres y madres universitarios en el cuidado de sus hijos, a estudiantes de regiones a través de programas de residencia universitaria, a estudiantes de colegios vulnerables con programas alternativos de ingreso y acompañamiento, con becas alimenticias y de mantención de alumnos con necesidades económicas o en otras ocasiones facilitando el acceso a ingresos extra a través de compensaciones laborales.

4.6.2. Recursos Clave

- Infraestructura para hogares universitarios y salas cuna

⁵<https://www.chileatiende.gob.cl/fichas/54854-proyecto-de-ley-que-crea-el-sistema-de-financiamiento-solidario>

-
- Recursos monetarios para asignación de becas.
 - Profesionales administrativos a cargo de realizar las actividades de caracterización socioeconómica.
 - Profesionales a cargo de identificar las necesidades de los estudiantes, y que generen programas efectivos y eficientes para poder abordarlas.
 - Fuentes de información externa e interna para la asignación de becas internas (PSU, DEMRE, JUNAEB, MINEDUC, Ficha Social y Sistema Guía Curricular).

4.6.3. Procesos Clave

- Procesos de calificación socioeconómica SIPEE para estudiantes provenientes de colegios vulnerables.
- Procesos de caracterización a través de la Ficha Social para todos los estudiantes que postulen a beneficios estudiantiles de la U. de Chile.
- Identificar las necesidades que poseen los estudiantes más vulnerables.
- Generar programas que permitan abordar las necesidades identificadas.
- Evaluar la eficiencia y efectividad de los programas implementados.
- Coordinar con facultades la entrega de beneficios.

4.6.4. Fórmula de utilidades

- Ingresos de arancel a partir de la retención de los alumnos en la universidad.
- Beneficios sociales de que un alumno complete sus estudios universitarios.
- Gastos para financiar becas de arancel y de mantención internas.
- Gastos de administración SIPEE y en Programa de Residencia Universitaria.
- Gastos en profesionales a cargo de extraer información desde sistemas informáticos, realizar estudios para caracterizar a estudiantes, generar nuevos métodos de acompañamiento y evaluar los nuevos programas implementados.

Capítulo 5

Situación Actual

En este capítulo se describe la situación actual de la Universidad de Chile y la Dirección de Bienestar Estudiantil, para lo cual se utiliza la Metodología de Ingeniería de Negocios y Diseño de Servicios (ver Marco Teórico). A partir de la aplicación de la metodología, se realiza un diagnóstico de la situación actual, para luego finalizar con una cuantificación de la problemática, la cual corresponde a una cuantificación de la deserción institucional de la Universidad de Chile¹.

5.1. Patrón de Negocio a utilizar

A partir del modelo del posicionamiento estratégico de la Universidad de Chile de Servicio Integral al Estudiante y del modelo de negocios de la DBE, surge la necesidad de generar una nueva capacidad que permita formalizar y sistematizar la retención de los alumnos de contexto más vulnerables.

El patrón de negocio correspondiente a la metodología de Ingeniería de Negocios que aplica en este caso corresponde al patrón de negocio 1 “Venta basada en el conocimiento del cliente” [1]. Este patrón, instanciado para la Universidad de Chile, toma el nombre de “Atención basada en el conocimiento del estudiante”, e indica que las componentes de negocio que deben ser desarrolladas para generar la capacidad de formalizar y retener a los alumnos de contextos más vulnerables, las cuales corresponden a capturar y organizar la data de los estudiantes, procesar dicha información

¹Alumnos que abandonan o son expulsados de sus programas de estudios, ya sea por motivos académicos o disciplinarios, y que no vuelven a matricularse en algún programa de estudio impartido por la Universidad de Chile.

con métodos analíticos y generar ideas, en base a los análisis, para generar ayudas proactivas a los estudiantes. La generación de esta nueva se visualiza en la figura 5.1 (recuadros destacados).

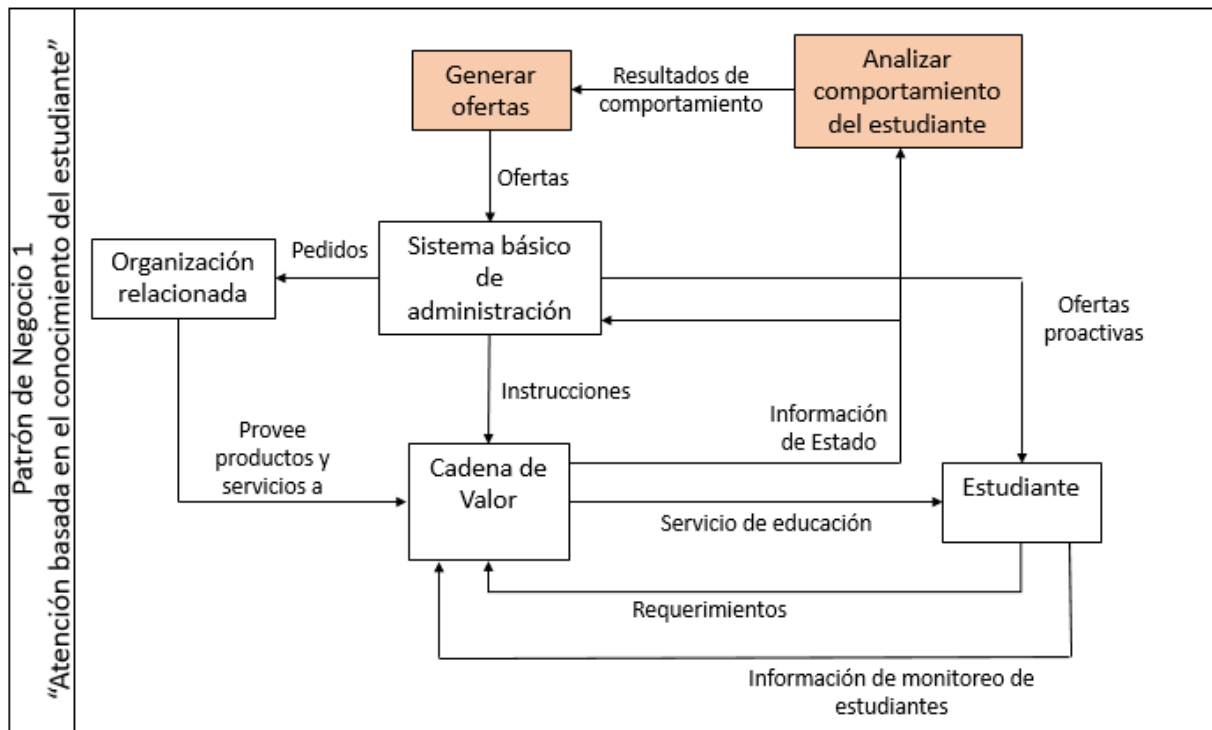


FIGURA 5.1: Patrón de Negocio 1 - Universidad de Chile. Fuente: Elaboración propia.

Según este Patrón de Negocio, para poder materializar la capacidad de retener a alumnos de contextos vulnerables, se requiere el rediseño del Macroproceso 1 "Cadena de Valor" [1], siempre que no haya una necesidad de cambios estructurales en la Cadena de Valor, que ocurre cuando la analítica puede ser rápidamente incorporada a los procesos y que los modelos predictivos son estables y que no requieren una revisión constante[1].

Actualmente en la Universidad de Chile la analítica puede ser incorporada rápidamente a los procesos, ya que se cuenta con fuentes de datos para realizar estudios predictivos (Ficha Social, DEMRE, JUNAEB, MINEDUC, entre otras) y datos históricos de alumnos entre los años 2011 a 2018. Si bien se puede incorporar más datos de alumnos, como los almacenados en el sistemas de administración académica (U-Cursos de la Facultad de Ingeniería y Web Docencia de la Facultad de Economía), los datos disponibles son suficientes para incorporar analítica en los procesos.

También se pueden incorporar modelos predictivos estables, ya que la data disponible posee un

formato estándar y la generación de nuevos atributos o cambios de formato son documentados por los administradores de la información (DEMRE, MINEDUC, JUNAEB e información interna).

Dado que el Patrón de Negocio 1 puede ser llevado a cabo, se procede a mapear estas capacidades en la Arquitectura de Procesos de Negocio, profundizando el estudio en macro 1 “Cadena de Valor”, como lo indica este patrón de negocio 23 [1]. Como la Universidad de Chile posee diferentes niveles de negocio (Administración Central, Facultades y Departamentos), su arquitectura debe ser presentada bajo una Arquitectura de Procesos de Negocio Multinivel.

5.2. Arquitectura de negocio Universidad de Chile

Actualmente la Universidad posee tres niveles de negocio, los cuales corresponden en el nivel superior a la Administración Central (Rectoría y Vicerrectorías) que presta servicios de orientación, coordinación, apoyo y desarrollo a la labor universitaria. En segundo nivel se encuentran las Facultades, las cuales poseen como máxima autoridad el decano. Y en tercer nivel, se encuentran los departamentos, los cuales materializan las actividades de investigación, extensión y educación. Dada esta estructura organizacional, es que se deben diseñar los macroprocesos dentro de una arquitectura de negocio multinivel (ver marco teórico sección 3.1.2 Patrones de Arquitectura y Procesos de Negocio). Como se indica 3.1.2, uno de los factores clave de diseño para estas estructuras corresponde a definir:

1. Cómo distribuir el “Plan de Negocio”.
2. Cómo distribuir el “Desarrollo de Nuevas Capacidades”.
3. Determinar qué servicios deben ser centralizados.

A continuación, se detalla la situación actual con respecto a estos tres aspectos.

1. Plan de Negocio

Actualmente, los Organismos Superiores de la Universidad (Rectoría, Consejo y Senado Universitario) elaboran y aprueban el Plan de Desarrollo Institucional (PDI), el cual contiene los lineamientos generales de la universidad. En paralelo al PDI, existen políticas que se generan a nivel central para apoyar las labores universitarias de las facultades, las cuales deben ser

aprobadas por el Senado Universitario. A partir de esos planes y políticas centrales, cada facultad crea sus planes estratégicos de acuerdo a sus necesidades y se rige según las políticas universitarias.

2. Desarrollo de Nuevas Capacidades

El desarrollo de Nuevas Capacidades en la Universidad de Chile posee un esquema mixto, ya que se desarrollan nuevas capacidades tanto a nivel central como en las mismas facultades y departamentos. A nivel central, se abordan necesidades que son transversales a la Universidad, generando economías de escala en la gestión de proyectos, administración de hardware, licencias de software, entre otros. En cuanto a facultades y departamentos, estos desarrollan capacidades específicas según sus requerimientos de gestión (siempre que posean la capacidad financiera para hacerlo) a través de recursos propios o fondos concursables.

3. Compartir servicios

En los estatutos de la U. de Chile se encuentra establecido que las Vicerrectorías apoyarán la gestión de las actividades universitarias. Actualmente, la forma de apoyar las actividades de las facultades e institutos por parte de las Vicerrectorías, corresponde al apoyo en materias financieras y de gestión (VAEGI), de investigación (VID), académicas (VAA), de bienestar estudiantil, equidad e inclusión (VAEC), de comunicaciones (VEXCOM) y de género (Dirección de Igualdad de Género).

El diagrama de la Arquitectura de Procesos de Negocio Multinivel actual de la Universidad de Chile se visualiza en la figura 5.2.

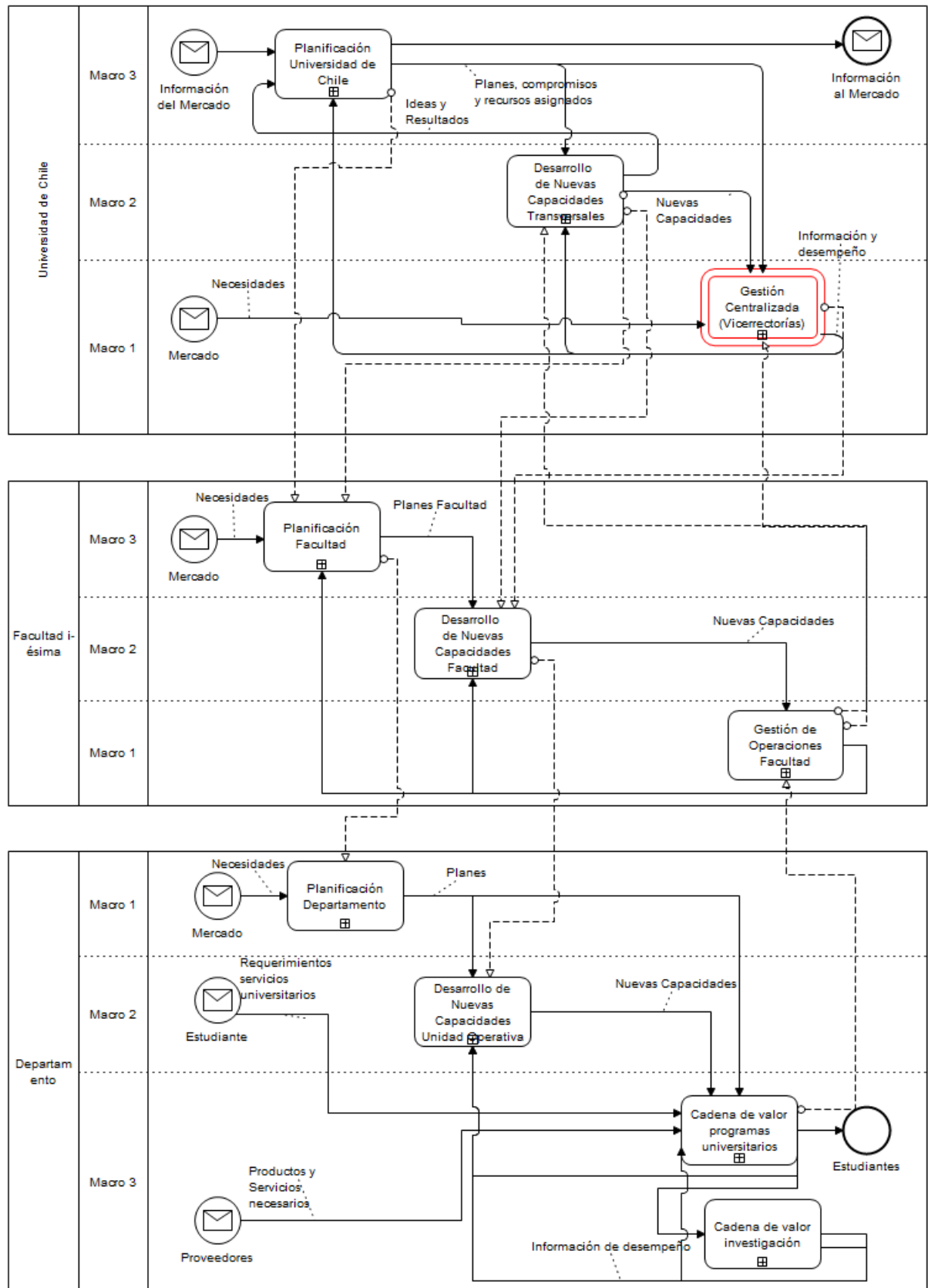


FIGURA 5.2: Arquitectura Multinivel, Universidad de Chile. Fuente: Elaboración propia.

Como se indicó anteriormente, para que el Patrón de Negocio 1 “Atención basada en el conocimiento del estudiante” pueda ser llevado a cabo, se debe focalizar el estudio en Macro 3. En particular, en esta tesis se centra el estudio en la Macro 3 de la Universidad de Chile, llamada “Gestión Centralizada” (ver recuadro destacado 5.2).

Aperturando el macroproceso “Gestión Centralizada”, se identifican 5 macroprocesos (ver figura 5.3): Administración de relación con el Estudiante, Administración de Relación con Proveedores, Gestión de Producción y Entrega Programas Universitarios, Producción y Entrega de Programas Universitarios y Mantenimiento de Estado. Como el enfoque de este proyecto de tesis va direccionado a generar una nueva capacidad que permita formalizar y retener a los alumnos de contextos más vulnerables en base a analítica, se realiza la apertura del proceso “Administración de Relación con el Estudiante”, donde se realizan todas las actividades de analítica, generación de programas de ayuda estudiantil, planificación, procesamiento y selección de estudiantes a beneficios.

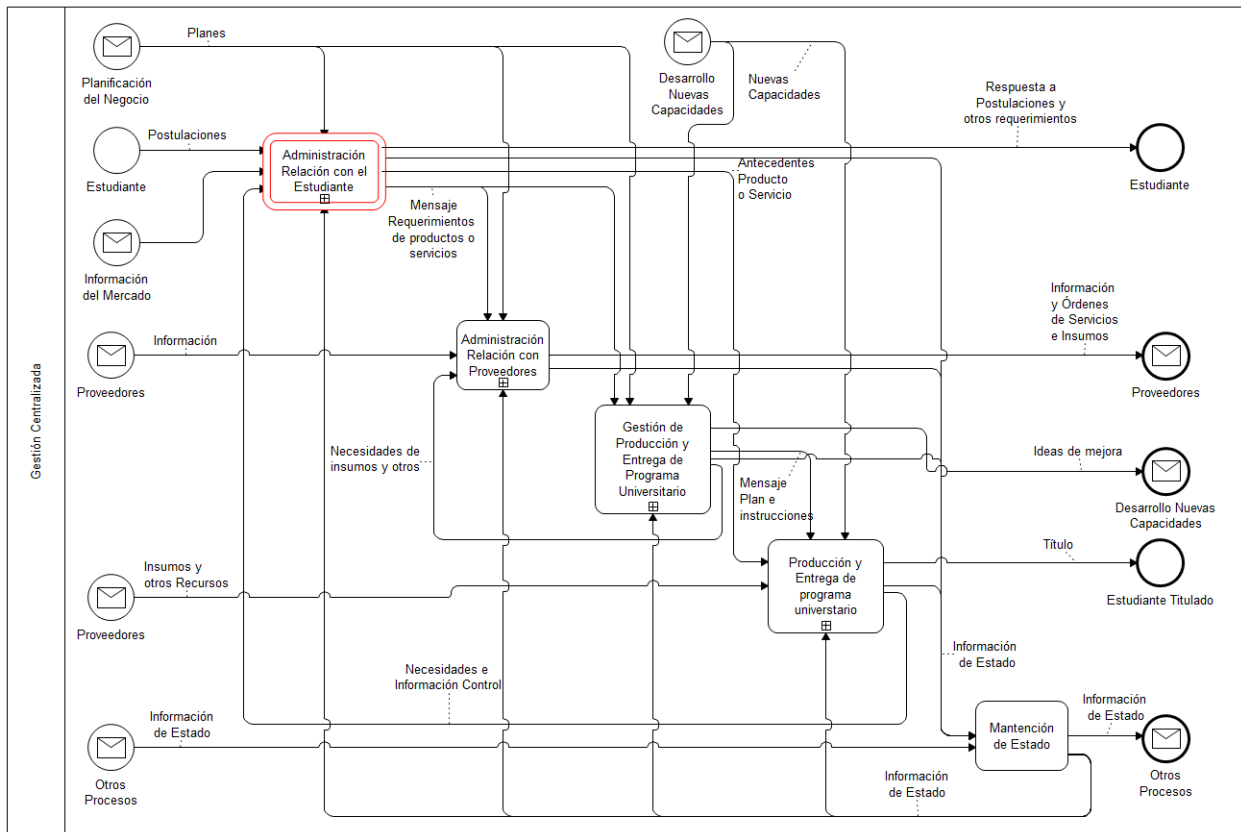


FIGURA 5.3: Gestión Centralizada, Universidad de Chile. Fuente: Elaboración propia.

Dentro del proceso “Administración de Relación con el Estudiante”, se encuentran los procesos de

“Atención Basada en el Conocimiento del Estudiante”, “Procesamiento de postulantes a beneficios” y “Selección de Postulantes a Beneficios” (ver figura 5.4).

■ **Atención basada en el conocimiento del estudiante**

Actualmente, se estudia el comportamiento de los alumnos a través de análisis descriptivos y estadísticos, para luego generar planes y programas que permitan retener a los alumnos más vulnerables.

■ **Procesamiento de postulantes a beneficios**

Recopilación de información de los postulantes por parte de las oficinas de bienestar de las facultades. En este proceso los estudiantes completan la Ficha Social, la cual es elaborada por la Dirección de Bienestar Estudiantil.

■ **Selección de postulantes a beneficios**

En base al puntaje obtenido en la ficha social, los programas de ayuda estudiantil establecidos y a la planificación de asignación de beneficios se determina qué estudiantes deben recibir beneficios.

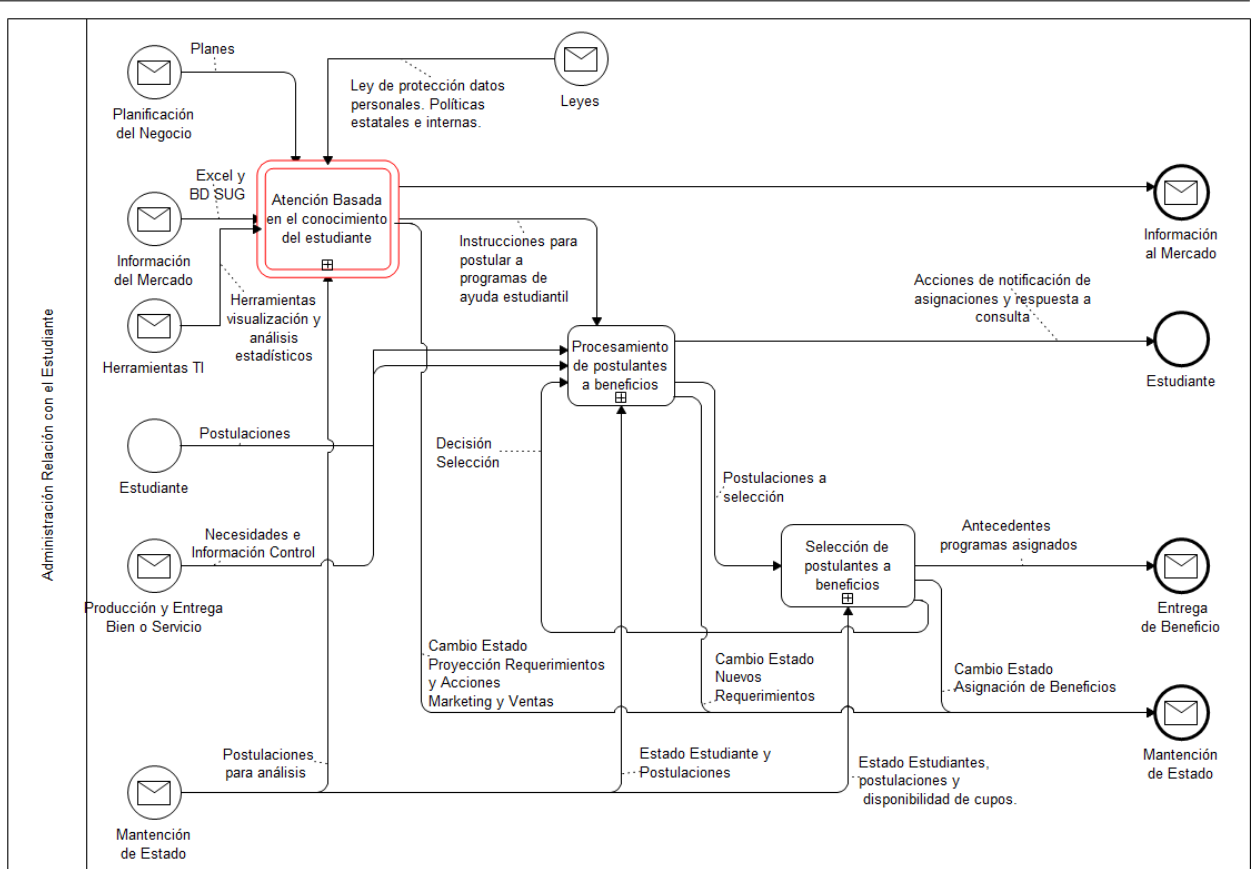


FIGURA 5.4: Administración de la Relación con el Estudiante, Universidad de Chile. Fuente: Elaboración propia.

Aperturando el proceso “Atención basada en el Conocimiento del Estudiante” se encuentran los procesos de “Introducción de Nuevos Análisis y Modelos”, “Analizar comportamiento y segmentación de estudiantes”, “Definir programas de ayuda estudiantil” y “Planificar asignación de beneficios” (ver figura 5.5).

■ Introducción de nuevos análisis y modelos

En este proceso se estudian qué modelos estadísticos y herramientas de visualización son útiles para analizar la información. Si bien esto se realiza en la práctica, no existe un proceso formal para evaluar las ventajas, desventajas y costos de las herramientas disponibles en el mercado.

■ Analizar comportamiento y segmentación de estudiantes

En este proceso ingresa del mercado información de estudios y análisis sobre temáticas de

deserción universitaria, datos de alumnos en planillas excel desde DEMRE, JUNAEB, MINE-DUC y el Departamento de Pregrado (notas de alumnos) y datos extraídos desde la Base de Datos SUG de la universidad. Luego la información es analizada a través de herramientas de visualización de datos que facilitan la visualización de tendencias, gráficos, tablas e indicadores, y herramientas para realizar análisis estadísticos. Estos análisis se realizan cuidando en todo momento los datos personales de los alumnos (ley 19.628 sobre protección a la vida privada) y ateniéndose a los planes estratégicos establecidos por la Universidad de Chile, la Vicerrectoría de Asuntos Comunitarios y Estudiantiles y la misma Dirección de Bienestar estudiantil. Como resultado se obtienen los resultados de los análisis, una proyección de beneficios que deberán ser asignados y registros de los análisis realizados.

■ Definir programas de ayuda estudiantil

A partir de los análisis descriptivos y estadísticos realizados sobre los alumnos, se evalúan los programas vigentes para abordar las necesidades de los alumnos y se modifican o generan nuevos programas para abordar sus necesidades. Esto es realizado tanto por las personas a cargo de realizar análisis estadísticos y personas expertas en el negocio. Para generar nuevos programas se deben atender en todo momento a la ley 19.628 sobre protección a la vida privada, y deben considerar los nuevos programas de becas y créditos generados por el gobierno, leyes vigentes (como por ejemplo, ley sobre universidades estatales y ley sobre educación superior) y políticas internas. Los nuevos programas generados deben contar con los requisitos para postular y especificaciones de montos, los cuales son enviados a a implementar a las direcciones de bienestar de las facultades. Además se publica en la web Universidad de Chile los nuevos programas de ayuda estudiantil disponibles y se responde a los requerimientos de información del Sistema de Información de Educación Superior (SIES) y Ministerio de Educación (MINEDUC).

■ Planificar asignación de beneficios

A partir de los programas de ayudas estudiantiles definidos y la proyección de asignación de beneficios (aumentos de matrícula, nuevos cupos especiales de acceso a la universidad, aumentos de financiamiento estatal para la gratuidad, entre otros) se crea un plan de asignación de beneficios. Para ello se requiere el personal a cargo de evaluar los planes y programas en base al presupuesto disponible y el estado de asignación de beneficios estudiantiles estatales,

ya que la asignación de beneficios interna se realiza una vez que son asignados los beneficios de MINEDUC y DEMRE (ej. Gratuidad y Beca Vocación Profesor).

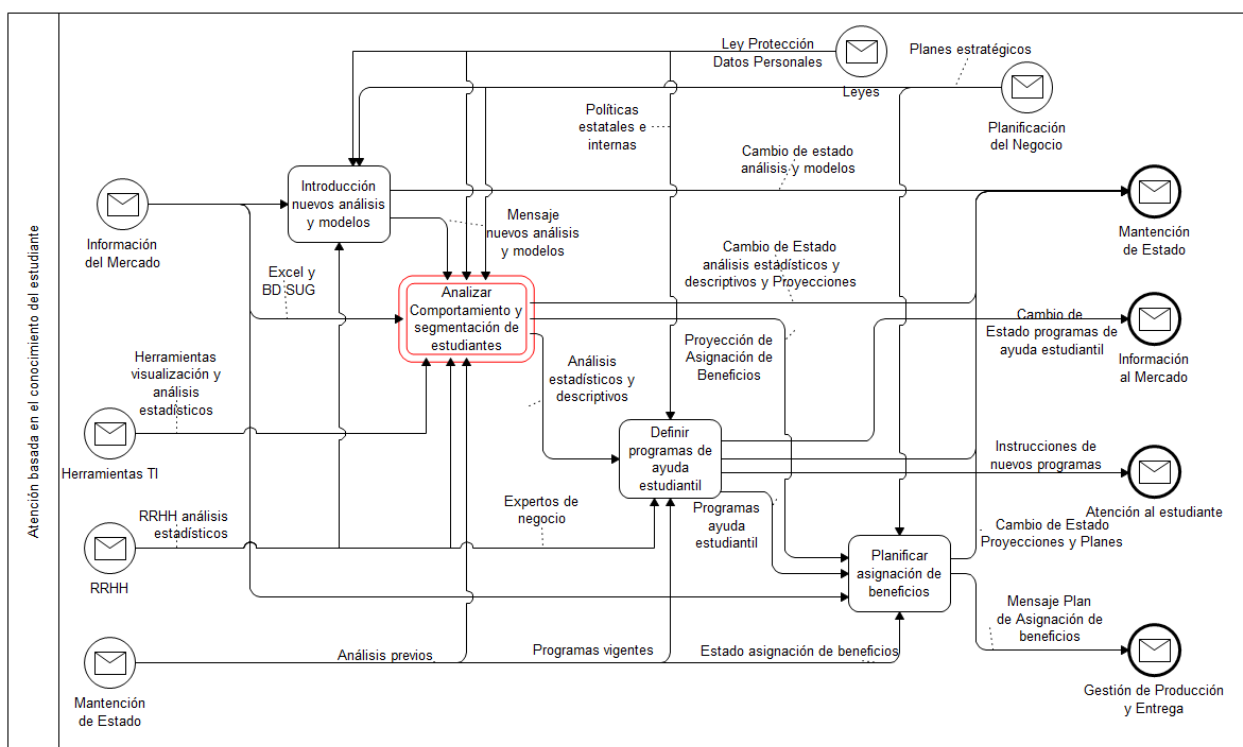


FIGURA 5.5: Atención Basada en el Conocimiento del Estudiante, Universidad de Chile. Fuente: Elaboración propia.

Aperturando el proceso “Analizar Comportamiento y Segmentación de Estudiantes” se encuentran los procesos de “Preparar Datos alumnos” y “Realizar análisis descriptivos y estadísticos” (ver figura 5.6).

■ Preparar Datos Alumnos

En este proceso ingresa del mercado información de estudios y análisis sobre temáticas de deserción universitaria, datos de alumnos en planillas excel desde DEMRE, JUNAEB, MINE-DUC y el Departamento de Pregrado (notas de alumnos) y datos extraídos desde la Base de Datos SUG de la universidad, los cuales son consolidados en una única planilla excel. Además, en caso de que los análisis descriptivos y estadísticas requieran de nueva información, se preparan los nuevos datos para su incorporación. Estos se rigen en todo momento por la ley de protección de datos 19.628 y ateniéndose a los planes estratégicos vigentes. El output de este

procesos son datos preparados para ser utilizados por herramientas de visualización y análisis estadísticos.

■ Realizar análisis descriptivos y estadísticos

En este proceso ingresan los datos preparados de los alumnos, para ser imputados en las herramientas de visualización de datos y de análisis estadísticos. Como resultado se obtienen los resultados de los análisis, una proyección de beneficios que deberán ser asignados y registros de los análisis realizados.

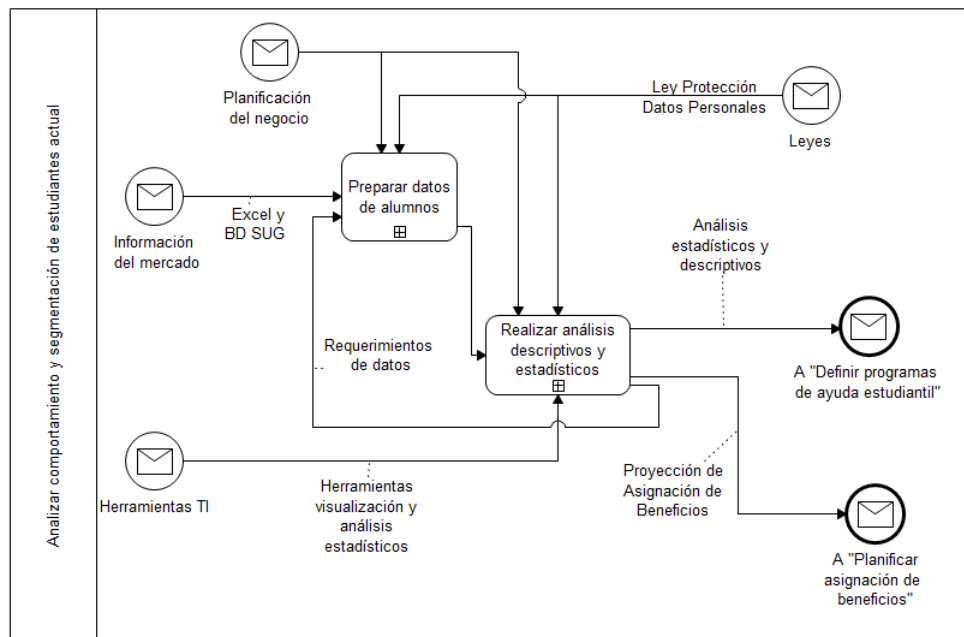


FIGURA 5.6: Analizar comportamiento y segmentación de estudiante, Universidad de Chile. Fuente: Elaboración propia.

5.3. Diagnóstico Arquitectura Multinivel

A partir de la situación actual de los factores clave para el diseño de Arquitecturas Multinivel (que corresponde a cómo distribuir el plan de negocio, cómo distribuir el desarrollo de nuevas capacidades y a determinar qué servicios deben ser centralizados), se realizó el siguiente diagnóstico:

1. Plan de Negocio

La forma de administrar actual es idónea para la diversidad de actividades de la Universidad, debido a que desde nivel central resultaría complejo y costoso detallar planes de desarrollo para cada una de las facultades, ya que cada una de ellas posee una especialización de acuerdo al área de estudios a la que se dedica, por ejemplo, las actividades de la facultad de ingeniería forestal difieren de forma completa a las actividades realizadas por la facultad de Medicina. Además, al diseñar planes estratégicos para las facultades de forma centralizada se estarían incumpliendo los estatutos universitarios al no respetar la autonomía de las facultades. Lo que sí se puede hacer desde nivel central (y actualmente se encuentran realizando esfuerzos para ello a través de la Unidad de Análisis Institucional, perteneciente a VAEGI), es detallar Programas de Desarrollo Institucional con objetivos SMART, es decir, objetivos estratégicos específicos, medibles, alcanzables, relevantes y basados en un horizonte temporal, que permitan traducirse en planes de acción en las facultades para cumplirlos.

2. Desarrollo de Nuevas Capacidades

No todas las facultades poseen la capacidad financiera para desarrollar nuevas capacidades, existiendo grandes brechas de entre aquellas facultades que generan ingresos adicionales por concepto de venta de bienes y servicios, y aquellas que se financian principalmente por aranceles y aportes estatales. De esta manera, el desarrollo de capacidades de nivel central permite a las facultades con escasos recursos recibir servicios de uso transversal a la universidad a bajos costos de mantención, y nula o baja inversión. En este sentido, las facultades se ven beneficiadas por parte de la Dirección de Bienestar Estudiantil al no tener que invertir en generar capacidades que permitan brindar apoyo a sus estudiantes, ya que esta se encarga de estudiar el cuerpo estudiantil para generar acciones, programas para su acompañamiento y planificar la entrega de beneficios, mientras que las facultades se centran en procesar a los postulantes a beneficios.

También resulta útil que las facultades mantengan su autonomía en el desarrollo de nuevas capacidades para que así puedan abordar las necesidades específicas que requieren, las cuales no pueden ser cubiertas por la administración central. A modo de ejemplo, una facultad puede iniciar un proyecto para instalar nuevos instrumentos tecnológicos para soportar sus actividades de laboratorio, lo cual requiere de conocimientos técnicos y de negocio específicos, necesitando de asesoría externa para su implementación. De la misma forma aplica para la relación entre

la facultad y sus departamentos, dejando grados de autonomía dependiendo del desarrollo alcanzado por estos, lo cual es y debe ser definido por las mismas facultades.

Ahora bien, debido a que se posee un esquema mixto para generar nuevas capacidades y a que en la universidad no existe una política de gobierno de datos y administración de sistemas, cada departamento tiende a implementar herramientas tecnológicas según sus propias necesidades generando múltiples bases de datos, y por ende, indicadores con diferencias en su definición, formato y calidad. Además, al poseer los sistemas sus propios dueños, acceder a la información se vuelve un proceso administrativo largo al tener que solicitar la información vía formal previa justificación, y en el caso de darse el acceso, la información suministrada generalmente se encuentra desactualizada por el tiempo requerido para su preparación y envío. Según Davenport [11] este es un elemento clave que debe ser desarrollado en las organizaciones para competir en base a analítica, indicando que “las iniciativas de analítica de relevancia para la organización deben mantener un liderazgo, herramientas y tecnologías en común”.

3. **Compartir Servicios**

La configuración actual de prestación de servicio por parte de las vicerrectorías permite concentrar el conocimiento y experiencia adquirida en un solo lugar, manteniendo de esta forma estándares de calidad de forma transversal en la universidad. Además, facilita que las facultades puedan mantener el foco en realizar sus actividades académicas y de investigación.

5.4. **Diagnóstico Situación Actual**

A partir del posicionamiento estratégico de la Universidad de Chile, los objetivos estratégicos y el modelo de negocios de la Dirección de Bienestar Estudiantil, el Patrón de Negocio a implementar y la situación actual de la Arquitectura de Procesos y Procesos de Negocio se realizó el siguiente diagnóstico:

- **Necesidad de un gobierno de datos y de sistemas**

Como se detalló anteriormente, debido a que la universidad posee un esquema mixto para generar nuevas capacidades y que no se posee una política de gobierno de datos y administración

de sistemas, cada departamento tiende implementar sus propias herramientas tecnológicas según sus propias necesidades generando múltiples bases de datos no integradas e indicadores. Es por ello que se requiere que ‘las iniciativas de analítica de relevancia para la organización deben mantener un liderazgo, herramientas y tecnologías en común’.

- **Necesidad de incorporar análisis predictivos robustos para detectar alumnos en riesgo de deserción**

La analítica actual para estudiar el comportamiento de alumnos corresponde a métodos analíticos descriptivos y estadísticos. Los análisis estadísticos son útiles caracterizar la deserción universitaria y responder a preguntas específicas sobre la influencia de ciertas variables en la deserción universitaria, utilizando ecuaciones matemáticas que mejor estiman los resultados de la muestra. En contraste, las técnicas de minería de datos requieren realizar menos supuestos sobre las variables, utilizan diversos tipos de algoritmos en vez de ecuaciones, y poseen un alto poder de predicción al obtener resultados que son generalizables, es decir, que poseen un alto poder de predicción a registros diferentes a la muestra de estudio. Además, permite la utilización de algoritmos tanto supervisados como no supervisados, permitiendo encontrar patrones ocultos en los datos.

- **Necesidad de informar a los alumnos los beneficios a los que son titulares**

Se requiere informar a alumnos los beneficios que son titulares de forma proactiva, recomendando programas que les pueden ser útiles a su situación. Para ello se requiere contar con algoritmos predictivos lo suficientemente robustos que permitan detectar el riesgo de deserción, para así saber a quienes dirigir las recomendaciones.

- **Diversidad de fuentes de datos y sistemas no integrados.**

En cuanto a la dirección de bienestar, esta maneja datos de alumnos en planillas excel que provienen desde DEMRE, JUNAEB, MINEDUC y el Departamento de Pregrado, además de los datos que son extraídos desde la Base de Datos SUG de la universidad. Si bien estos datos son suficientes para incorporar la analítica en los procesos, no permiten realizar un continuo seguimiento de alumnos, ya que no se encuentran integrados. Además, existen fuentes de información útiles para el seguimiento de alumnos, como los son los datos académicos U-Cursos, U-Campus, Web Docencia y Sistema de Administración Docente, los cuales pueden ser incorporados para una mejor caracterización de los alumnos. Es por ello que se requiere un

repositorio de información de alumnos que se mantenga actualizado en todo momento bajo un formato, definición y calidad.

■ Economías de escala

Como se indico anteriormente, la configuración actual de las Vicerrectorías de prestar servicios de forma centralizada permite concentrar el conocimiento y experiencia adquirida en un solo lugar, manteniendo de esta forma estándares de calidad de forma transversal en la universidad. Además, como no todas las facultades poseen la capacidades financiera para desarrollar nuevas capacidades, el desarrollo de nuevas capacidades a nivel central les permite acceder a servicios de bajos o nulos costos de mantención e inversión.

En este sentido, el rol de la Dirección de Bienestar beneficia a las facultades al no tener que invertir en generar nuevas capacidades que permitan brindar apoyo a sus estudiantes, ya que de forma centralizada planifica la asignación de beneficios, crea nuevos programas y determina los lineamientos globales para la Universidad, mientras que las facultades a través de sus direcciones de bienestar se concentran en procesar a los postulantes a beneficios para asignarles las ayudas correspondientes.

Si bien lo anterior ya es positivo, se identifica la oportunidad de mejorar las economías de escala ya existentes, a través de la sinergia de los procesos de analítica de la Vicerrectoría de Asuntos Estudiantiles y Comunitarios (VAEC) y la Vicerrectoría de Asuntos Académicos (VAA), ya que ambas realizan estudios para poder abordar las necesidades de los alumnos. Si bien la VAA se centra en los ejes de Aprendizaje y Enseñanza del Modelo de Desarrollo Integral del Estudiante y la VAEC al eje de Calidad de Vida (ver Modelo de Desarrollo Integral del Estudiante página 51), existen puntos en común que pueden ser potenciados al poder llevar a cabo procesos de analítica en conjunto.

5.5. Cuantificación del Problema u Oportunidad

Los costos de un alumno desertor corresponden a costos sociales y económicos. En términos sociales, se produce un costo para el estudiante y/o su familia al tener que hacerse cargo de deudas que pudiera haber adquirido al solicitar un crédito universitario. Ello sumado a que Chile es un país que exhibe una alta desigualdad socioeconómica, donde sólo un 20% de quienes postulan a la educación

superior provienen de familias con capacidad para financiar sus estudios [12] y que a que los ingresos al no tener una carrera profesional en Chile se ven considerablemente disminuidos, genera un impacto social enorme para al tener personas que deben hacerse cargo de financiar una deuda alta con bajos ingresos.

Si bien la estimación de costos que genera la deserción universitaria considera costos sociales, debido a la dificultad de contar con datos confiables que puedan representar de forma correcta los costos sociales de un alumno desertor, se considerarán sólo los costos económicos que le genera a la Universidad de Chile la deserción de alumnos.

Para calcular los costos económicos de un alumno desertor se consideró:

- Que los ingresos promedios por concepto de arancel que percibió la Universidad de Chile por sus alumnos de primer año del cohorte 2017 fue de \$3.914.092 ².
- Que la permanencia promedio de los alumnos en la Universidad de Chile de 6 años ³.
- Una tasa de descuento anual del 10 %.
- Que la tasa de deserción institucional de alumnos de primer año en 2017 fue de un 13 % (828 alumnos).
- Que la tasa de deserción se ha mantenido estable en la Universidad de Chile entre los años 2011 y 2017 ⁴.

Dado lo anterior, se tiene la universidad percibe \$17.046.891 por cada alumno que se mantiene dentro de la institución durante toda su carrera.

Considerando que el total de alumnos de primer año desertores del cohorte 2017 fueron 828 alumnos, se tiene que para dicho año se tuvo un costo anual por concepto de deserción de \$ 3.240 millones.

Considerando además que la deserción se ha mantenido estable entre 2011 y 2017, se puede estimar que cada año le cuesta a la universidad \$ 3.240 millones anuales por concepto de alumnos desertores de primer año.

²Fuente: Sistema de administración de alumnos Universidad de Chile “Guía Matrícula”

³ Base de datos “Duración real y sobreduración de las carreras 2017” SIES. Si bien se indica que la duración real para universidades en carreras de pregrado para 2016 es de 12,5 semestres, se consideraron para 2017 12 semestres (6 años) debido a que en esta misma base de datos se indica que la duración de las carreras va a la baja, disminuyendo para universidades 0,2 semestres entre 2012 y 2016.

⁴<http://web.uchile.cl/archivos/anuario/2017/38/>

Capítulo 6

Propuesta de Diseño de Procesos

En este capítulo se presentan las direcciones de cambio y alcance del rediseño, los cambios requeridos en los procesos de la Universidad de Chile, las tecnologías necesarias para el rediseño y la aplicación de minería de datos para predecir la deserción de alumnos.

6.1. Direcciones de Cambio y Alcance

En base a la situación actual y el diagnóstico realizado, se utilizaron las Direcciones de Cambio definidas en el libro Ingeniería de Negocios: “Diseño Integrado de Negocios, Procesos y Aplicaciones TI” [13] para sistematizar el diseño de procesos.

a. Estructura de empresa y mercado

TABLA 6.1: Dirección de cambio: Estructura de empresa y mercado

VARIABLES DE DISEÑO	ACTUAL	PROPUESTO
a.1 Servicio Integral al cliente	Sí	Conocer mejor las necesidades de alumnos mediante técnicas de minería de datos
a.2 Lock-in sistémico	No	No

a.3 Integración con proveedores	No	No
a.4 Estructura interna: centralizada o descentralizada.	<p>Operación entre Vicerreectorías de forma independiente.</p> <p>Sistemas y bases de datos dependen de cada organismo.</p>	<p>Se propone unificar los esfuerzos de la Vicerreectorías de Asuntos Estudiantiles y Comunitarios (VAEC) y la Vicerreectoría de Asuntos Académicos (VAA), con el objetivo de concentrar el conocimiento de la nueva capacidad de minería de datos en un único lugar y generar economías de escala para mantenerla. Según Barros, esto “requiere diseñar las interrelaciones entre cadenas de valor y los procesos de servicio” [1], que en este caso, corresponde a diseñar las interrelaciones de cadenas de valor y procesos de servicio de ambas Vicerreectorías.</p> <p>En la nueva situación, las bases de datos siguen dependiendo de cada organismo, pero bajo un esquema de gobierno de datos y de sistemas, ya que como indica Davenport, “las organizaciones que han tenido éxito en la implementación de analítica, mantienen sus iniciativas bajo un liderazgo, herramientas y tecnologías en común”[11].</p>
a.5 Toma de decisiones: Centralizada o descentralizada	Toma de decisiones independiente por Vicerreectoría	VAEC y VAA toman decisiones de forma independiente previa coordinación en temáticas de analítica.

b. Anticipación

TABLA 6.2: Dirección de cambio: Anticipación

Variables de Diseño	Actual	Propuesto
b.1 Planificación	Proyección de alumnos que requerirán ayuda socioeconómica, considerando aumentos de matrícula, nuevos cupos especiales y aumentos de financiamiento estudiantil.	A lo anterior, se añade la predicción de alumnos desertores que permitirá dirigir las ayudas estudiantiles.
b.2 Modelo predictivo de retención de alumnos	No	Modelo basado en técnicas de minería de datos que determinan el riesgo de un alumno de desertar para decidir asignación de beneficio.

c. Coordinación

TABLA 6.3: Dirección de cambio: Coordinación

Variables de Diseño	Actual	Propuesto
c.1 Reglas	Actualmente se decide brindar un beneficio de ayuda estudiantil en base a puntajes obtenidos en la Ficha Social.	Se agregará a la Ficha Social el riesgo de deserción, el cual permitirá priorizar la entrega de beneficios a aquellos alumnos que posean un alto riesgo de desertar.

c.2 Jerarquía	Programa de Emergencia para la Retención	En caso de necesidad de ayuda crítica del estudiante, se mantendrá el mecanismo de ayuda estudiantil del Programa de Emergencia para la Retención, el cual consiste en una ayuda económica para alumnos que posean una situación de emergencia.
C.3 Colaboración	Compartir información y experiencia con otros departamentos para generar programas de retención de estudiantes.	Se añade el compartir los resultados de análisis predictivos en las instancias de colaboración para así nutrir la discusión de que programas crear para la retención de alumnos en base a los resultados de los algoritmos de minería de datos.
C.4 Partición	Estudios sobre alumnos son llevados a cabo por cada departamento de forma independiente	Mantener capacidad de minería de datos con otros departamentos que trabajen atendiendo necesidades de alumnos.

d. Prácticas de Trabajo

TABLA 6.4: Dirección de cambio: Prácticas de Trabajo

VARIABLES DE DISEÑO	Actual	Propuesto
d.1 Lógica de Negocio automatizada o semi-automatizada	No automatizada	Semi-Automatizada. Los algoritmos de minería de datos detectan a alumnos con alto riesgo de deserción, generándose como output un listado de alumnos en riesgo de desertar. Este listado es entregado a encargados de bienestar de facultades, para que contacten al alumno para poder evaluar sus necesidades y decidir un curso de acción.

d.2 Lógica de apoyo a actividades tácitas.	Presentación de resultados análisis estadísticos	Presentación de resultados de modelos de minería de datos. Consiste en presentar a los directivos los resultados de los algoritmos para predecir la deserción, detallando beneficios y precisión de los modelos.
d.3 Procedimientos de comunicación e integración.	Se comunica a asistentes sociales la información que deben solicitar a los alumnos para luego asignar los beneficios de forma centralizada.	Además de lo anterior, se comunica a asistentes sociales los alumnos que se encuentran en riesgo de desertar para su contacto y evaluación.
d.4 Lógica y procedimientos de medición de desempeño y control.	No	Se mide desempeño según la precisión de los modelos en evitar la deserción de alumnos.

e. Integración de Procesos Conexos

TABLA 6.5: Dirección de cambio: Integración de Procesos Conexos

VARIABLES DE DISEÑO	Actual	Propuesto
e.1 Proceso aislado	Sí	No
e.2 Todos o la mayor parte de los procesos de un macroproceso	No	Los procesos de la cadena de valor (introducción de nuevos análisis y modelos, analizar compartimento y segmentación de estudiantes, definir programas de ayuda estudiantil y planificar asignación de beneficios) se encuentra integrados.

e.3 Dos o más macros interactúan	No	Integración entre cadena de valor, generación de nuevas capacidades y planes estratégicos.
----------------------------------	----	--

f. Mantenimiento de Estado

TABLA 6.6: Dirección de cambio: Mantenimiento de Estado

VARIABLES DE DISEÑO	ACTUAL	PROPUESTO
f.1 Datos propios	Sí	Sí
f.2 Integración con datos otros sistemas de la empresa	No	Sí. Data Mart que permita centralizar información relacionada a alumnos desde sistemas curriculares y académicos (U-cursos, web docencia, u-campus, Guía Matricula) y otras fuentes de información (Ficha Social, Becas, PSU, entre otras).
f.3 Integración con datos de sistemas de otras empresas	No	No se requiere.

6.2. Arquitectura de Procesos *to be*

El rediseño de procesos se realiza dentro de la macro Administración de la Relación con el Estudiante (ver figura 6.1), donde se modifican los procesos “Atención basada en el conocimiento del estudiante”, “Procesamiento de postulantes a beneficios” y “Selección de postulantes a beneficios”, los cuales incorporan la predicción de la deserción a través de modelos de minería de datos, para luego proceder a contactar a aquellos alumnos con riesgo de deserción de forma proactiva. Además, se incorpora la creación de un Data Mart de alumnos para contar con datos de alumnos actualizados y bajo un único estándar.

- **Atención basada en el conocimiento del estudiante**

Se rediseña este proceso, al segmentar y predecir mediante algoritmos de minería de datos la deserción de alumnos, ofreciendo ayudas estudiantiles de forma proactiva, y generando planes y programas que se basen en estos análisis. Además, se reemplaza la forma en que actualmente se extrae y depura la información de diferentes fuentes (que principalmente es realizado en planillas excel y a través de requerimientos con los dueños de los sistemas o de la información), por la creación de un Data Mart de alumnos que pueda contener información relevante de los estudiantes para ser utilizados dentro de los modelos predictivos y de clusterización.

- **Procesamiento de postulantes a beneficios**

Este proceso se rediseña ya que las asistentes sociales de las facultades, además de recopilación de información de los postulantes, se encargarán de otorgar recomendaciones de becas (en base a los análisis) de forma proactiva a aquellos alumnos que posean un alto riesgo de desertar.

- **Selección de postulantes a beneficios**

Además de la forma actual de otorgar beneficios de acuerdo al puntaje obtenido en la ficha social, los programas de ayuda estudiantil establecidos y a la planificación de asignación de beneficios, se considerará el riesgo de desertar detectado por los modelos de minería de datos.

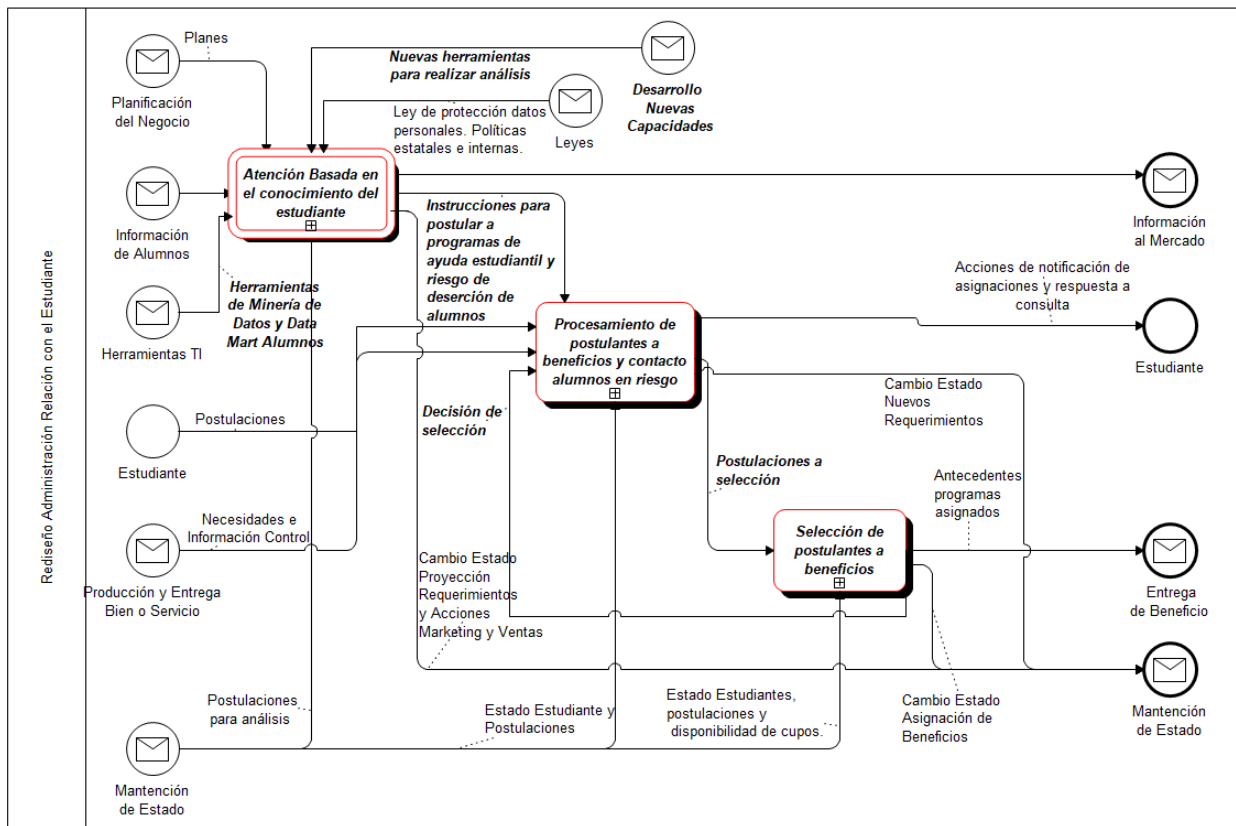


FIGURA 6.1: Rediseño Administración de la Relación con el Estudiante, Universidad de Chile.
Fuente: Elaboración propia.

Aperturando el proceso “Atención basada en el conocimiento del estudiante” (ver figura 6.2), se encuentran los procesos “Introducción de nuevos análisis y modelos”, “Analizar comportamiento y segmentación de estudiantes”, “Definir programas de ayuda estudiantil” y “Planificar asignación de beneficios”. Los modelos de minería de datos para predecir la deserción son incorporados dentro de “Analizar comportamiento y segmentación de estudiantes”, mientras que “Definir programas de ayuda estudiantil” y “Planificar asignación de beneficios” incorporan los outputs de los modelos de minería de datos para la elaboración de los respectivos programas y planificación de asignación de beneficios.

■ Introducción de nuevos análisis y modelos

Se propone la evaluación sistemática de herramientas de minería de datos, de forma de asegurar que estas se puedan integrar Data Mart de Alumnos, y que cumplan con políticas de gobierno de datos y administración de sistemas.

- **Analizar comportamiento y segmentación de estudiantes**

Este es un proceso que actualmente no existe en la organización y uno de los más relevantes para generar la nueva capacidad de retener a alumnos en riesgo de deserción mediante minería de datos. Este incluye recopilar, ordenar y transformar los datos, segmentar a alumnos mediante algoritmos de clusterización y predecir la deserción de alumnos.

- **Definir programas de ayuda estudiantil**

Este proceso se modifica al incorporar los resultados de minería de datos (segmentación y predicción de deserción), para generar programas que puedan satisfacer las necesidades de los alumnos.

- **Planificar asignación de beneficios**

Además de proyectar la asignación de beneficios actuales en base a aumentos de matrícula, nuevos cupos especiales de acceso a la universidades, aumentos de financiamiento para la gratuidad, entre otros, se incorporan como variable para realizar la proyección de asignación de beneficios las predicciones de alumnos desertores realizadas por los algoritmos de minería de datos.

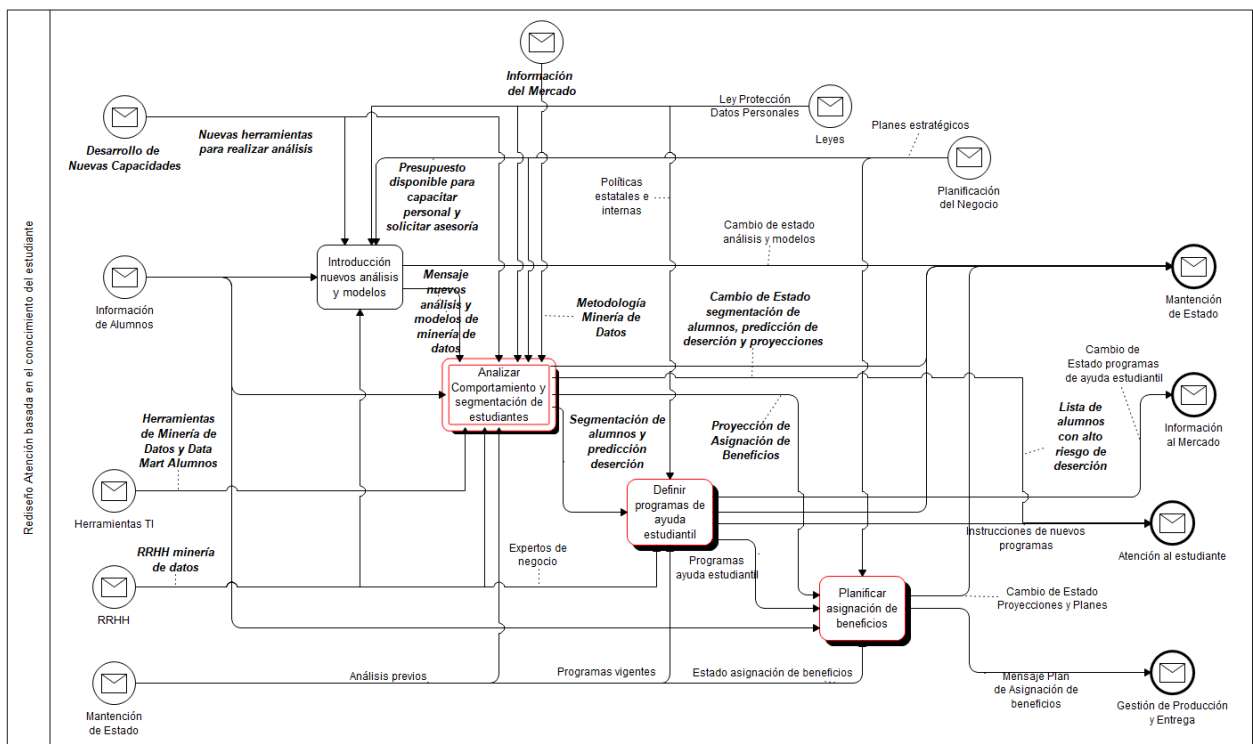


FIGURA 6.2: Rediseño Atención Basada en el Conocimiento del Estudiante, Universidad de Chile.
Fuente: Elaboración propia.

6.3. Diseño de Proceso “Analizar Comportamiento y Segmentación de Estudiantes”

A continuación se detalla el diseño del proceso “Analizar Comportamiento y Segmentación de Estudiantes”, el cual corresponde al proceso más relevante para generar la nueva capacidad de detectar a alumnos en riesgo de deserción para ofrecer ayudas de beneficios estudiantiles de forma proactiva.

El primer paso de este proceso es preparar los datos en el Data Mart de Alumnos. Para ello, se deben extraer datos de alumnos de sistemas internos y externos, y cargarlos en el *Data Staging Area* (DSA), el cual corresponde a un repositorio de datos preliminar antes de cargar los datos en el Data Mart. Luego, se corren scripts de limpieza y se escogen las variables que serán cargadas en el Data Mart, mediante un proceso de *Extract* (Extraer), *Transform* (Transformar) and *Load* (Cargar) (ETL). Se irán realizando ajustes a los datos a medida que el negocio vaya requiriendo la carga de datos de otras fuentes de información, construir nuevas variables o actualizar las ya existentes.

Luego se procede a probar diferentes algoritmos de minería de datos, ya sean estos predictivos o de clusterización, hasta llegar a resultados deseables y seleccionar uno de los modelos.

Con el modelo seleccionado, se pone en práctica los algoritmos predictivos, a partir de los cuales se definen nuevos programas de ayuda estudiantil, se planifica la asignación de beneficios, se contacta a aquellos estudiantes con alto riesgo de deserción y se complementa la ficha social al momento de decidir a qué postulantes se deberán asignar los beneficios.

Todas estas actividades van enmarcadas dentro de la metodología de CRISP-DM, donde se debe definir el problema a resolver, preparar los datos, modelar con diferentes algoritmos, evaluar los resultados y decidir la puesta en marcha.

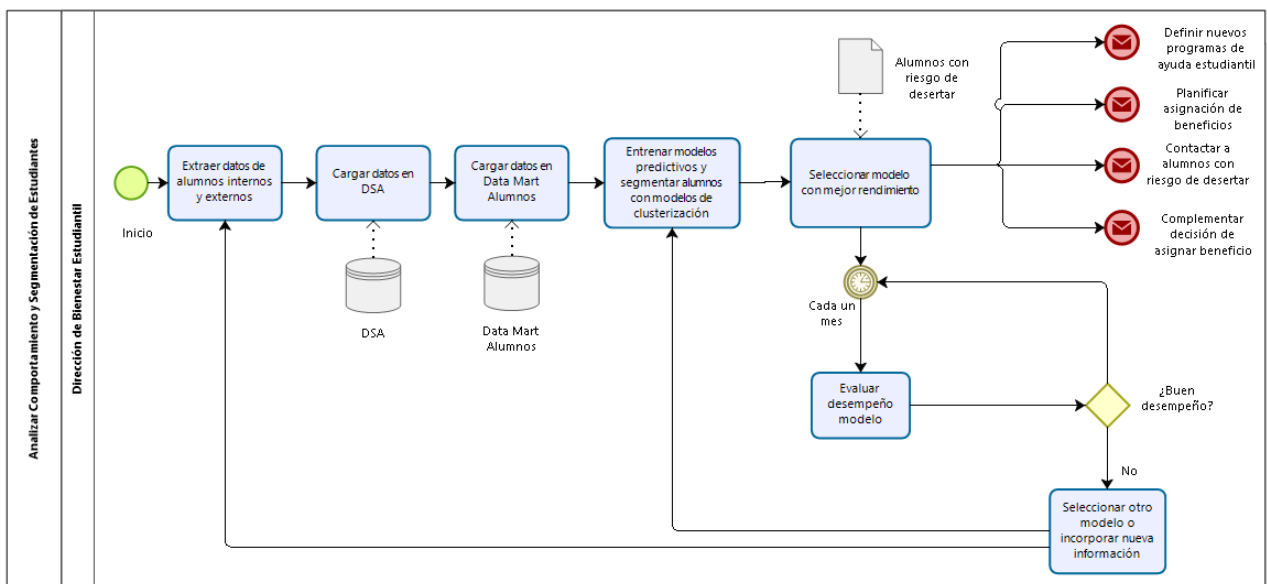


FIGURA 6.3: Rediseño Proceso Analizar Comportamiento y Segmentación de Estudiantes
Fuente: Elaboración propia.

6.4. Diseño de Lógica de Negocios

En esta sección se detalla la lógica de negocio inserta en el proceso “Analizar Comportamiento y Segmentación de Estudiantes”, específicamente en las actividades “Extraer datos de alumnos de fuentes internas y externas”, “Entrenar modelos predictivos y segmentar alumnos con modelos de clusterización” y “Seleccionar modelo con mejor rendimiento”.

Si bien las actividades figuran en un orden lineal en el diagrama, en la práctica corresponde a un proceso iterativo, partiendo por determinar el problema de negocio a resolver, para luego comprender los datos que se poseen disponibles, preparar los datos, entrenar los modelos de minería de datos, evaluar los resultados y puesta en marcha.

En determinar el problema de negocio se estableció que el objetivo del proyecto es disminuir la deserción universitaria.

En la comprensión y preparación de datos se extrajeron datos de múltiples fuentes de información, se realizaron análisis exploratorios de los datos y se procedió a realizar un tratamiento sobre los datos, ya sea reemplazando valores faltantes o creando nuevas variables. Luego de este tratamiento de datos se procedió a seleccionar las variables a través del método de *Forward Selection y Backward Elimination* (ver chap:Forward Selection and Backward Elimination).

En entrenar modelos de minería de datos, se entrenaron modelos de clusterización como predictivos. En algoritmos de clusterización, se utilizó X-Means y DBSCAN. Mientras que en algoritmos de predicción, se utilizó Random Forest, Decision Trees (3 tipos diferentes) y Regresión Logística.

La muestra utilizada corresponde a los alumnos de primer año del cohorte 2017 que rindieron la PSU ¹, los cuales corresponden a 6.254 alumnos, es decir, el 95 % del total de alumnos de primer año del cohorte 2017.

Los software utilizados fueron:

- **RapidMiner:** Software de minería de datos. Este fue utilizado para preparar los datos, correr los modelos predictivos y aplicar las diferentes técnicas de minería de datos (validación cruzada, balanceo de datos, ajustes de parámetros, entre otros).
- **Stata:** Software estadístico. Se utilizó para aplicar el algoritmo de selección de variables *Forward Selection and Backward Elimination*, determinar la normalidad de la distribución de las

¹El motivo de seleccionar los registros de alumnos que hayan rendido la PSU, es que esta variable tiene un alto poder predictor de deserción en alumnos, siendo esta estadísticamente significativa bajo modelos de regresión logística. Considerar a los alumnos que no rindieron la PSU para entrar a la universidad (alumnos extranjeros principalmente), significaría realizar reemplazo de los valores faltantes por algún método de reemplazo aleatorio, lo cual no aseguraría contar con datos confiables al momento de correr los modelos. Es por ello que en este estudio se opta por trabajar la muestra de primer año sólo de alumnos que hayan rendido la PSU, dejando para futuros trabajos incluir el estudio de alumnos que no hayan rendido la PSU. Se destaca que los programas que buscan equidad en el acceso como SIPEE, BEA, PACE, entre otras, se mantienen dentro del estudio, ya que estos poseen dentro de sus requisitos mínimos haber rendido la PSU.

variables de tipo real y entera, y para aplicar test de hipótesis (chi-cuadrado y test de medias) en el análisis exploratorio.

6.4.1. Comprensión y preparación de datos

En esta sección se identifican las fuentes de datos de alumnos con sus respectivos métodos de extracción, descripción de las variables, tratamiento de datos faltantes, construcción de nuevas variables y análisis exploratorios (gráficos, correlaciones, distribución de variables y análisis estadísticos simples). Este proceso si bien tiene un orden de lógico para su realización, en la práctica es un proceso iterativo donde a medida que se va obteniendo un mayor conocimiento de los datos, se van seleccionando, modificando o creando variables para ser utilizadas en los modelos.

6.4.1.1. Fuentes de Datos

Las bases de datos utilizadas para la realización de este estudio provienen tanto de bases de datos internas como de bases de datos externas que son entregadas directamente a la DBE. En la tabla 6.7 se detalla las instituciones del estado que entregan información a la DBE (JUNAEB y MINEDUC) y el organismo de la Universidad de Chile que se encuentra a cargo de realizar el proceso PSU (DEMRE).

TABLA 6.7: Descripción de Instituciones

Institución	Descripción
DEMRE	El Departamento de Evaluación, Medición y Registro Educacional (DEMRE) es un organismo de la U. de Chile que se encarga de desarrollar, analizar y publicar los resultados del proceso PSU.
JUNAEB	La Junta Nacional de Auxilio Escolar y Becas (JUNAEB) es un organismo estatal que se encarga de evaluar y entregar becas de mantención a escolares y universitarios.
MINEDUC	El Ministerio de Educación (MINEDUC) es un organismo estatal, que dentro de sus actividades, se encarga de evaluar y entregar becas de arancel, créditos universitarios y el beneficio de gratuidad a universitarios.

En la tabla 6.8 se detallan las bases de datos que fueron utilizadas indicando la fuente, datos utilizados y método de extracción.

TABLA 6.8: Extracción de Datos

Fuente	Descripción
Guía Curricular	Sistema interno que contiene información académica y financiera de los estudiantes y ex-alumnos de la universidad. La información de este sistema es almacenada en la base de datos denominada "SUG". A través de una consulta a la base de datos se extrajeron las variables región, nacionalidad, carrera, facultad, procedencia educacional, vía de ingreso, tramo socio-económico, situación académica de la carrera y rut de los alumnos de primer año del cohorte 2017 y de enero a mayo de 2018. Además se extrajo la columna rut y situación académica de los alumnos de los cohortes 2011 a 2016, con el objetivo de identificar a aquellos alumnos que cursaron otra carrera en la universidad sin haberla completado.
Ficha Social	Formulario interno que debe ser completado por los alumnos que postulan a los beneficios estudiantiles internos de la Universidad de Chile. De este formulario se extrajeron las variables enfermedad catastrófica, enfermedad permanente, enfermedad crónica, número de dormitorios, número de integrantes hogar, nivel educación jefe de hogar, nivel educación madre, asignación y montos de becas de arancel, asignación y montos de becas de mantención internas de los alumnos de primer año del cohorte 2017. Se utilizó sólo la información de los registros con estado "cerrada", ya que este estado indica que la información presentada fue acreditada por el alumno con los encargados de bienestar de su facultad.
Resultados proceso PSU	Se utilizaron las variables de Puntaje PSU, Ranking y NEM del cohorte 2017. Esta información es enviada a la DBE por el DEMRE en diciembre de cada año

Becas de arancel, créditos y gratuidad Estatales	Se utilizaron las becas, créditos (Crédito con Aval del Estado y Fondo Solidario de Crédito Universitario) y gratuidad asignada por el MINEDUC a los alumnos de primer año del cohorte 2017. Esta información es enviada a la DBE en marzo de cada año.
Becas de Mantención Estatales	Se utilizaron las asignaciones de becas de mantención asignadas a los alumnos por parte de la JUNAEB a los alumnos de primer año del cohorte 2017. Esta información es enviada a la DBE en marzo de cada año.
Índice de Vulnerabilidad Escolar (IVE)	Se extrajo el índice de vulnerabilidad escolar de enseñanza media de colegios municipales y subvencionados, exceptuando a los colegios particulares, ya que estos no son clasificados bajo el IVE. Esta información es de carácter público y se descarga directamente de la web de JUNAEB.

6.4.1.2. Descripción de Variables

En esta sección se describen aspectos relevantes de las variables que serán utilizadas en el proceso de minería de datos.

1. Variables Guía Matrícula y Ficha Social

TABLA 6.9: Atributos Guía Matrícula e Índice de Vulnerabilidad Escolar (IVE)

Nombre	Tipo	Valores Faltantes
Situación Académica (Años 2017 y 2018)	Categórica	0
Región	Categórica	0
Sexo	Binaria	0
Nacionalidad	Binaria	0
Carrera	Categórica	0
Procedencia Educacional	Categórica	39
Vía Ingreso	Categórica	0
Tramo Socioeconómico	Categórica	0

Índice de Vulnerabilidad Escolar (IVE)	Real	2.005
Enfermedad Catastrófica	Binaria	4.124
Enfermedad Permanente	Categórica	6.211
Enfermedad Crónica	Binaria	4.124
Número de dormitorios	Numérica	4.501
Número de integrantes del hogar	Numérica	4.124
Nivel de educación jefe de hogar	categórica	4.210
Nivel de educación madre	Categórica	4.411

Para ver los valores de cada variable y fuente de información ver anexo Detalle atributos Guía Matrícula e Índice de Vulnerabilidad Escolar (IVE).

2. Créditos MINEDUC

Durante el año 2017 se entregaron 2 créditos universitarios por parte del MINEDUC, los cuales corresponde al Crédito con Aval del Estado (CAE) y Fondo Solidario de Crédito Universitario (FSCU). Con estos créditos fueron beneficiados 924 alumnos con CAE y 287 con FSCU. Estas dos variables son de tipo binario y no poseen registros con valores faltantes. Para mayor detalle ver anexo Créditos Universitarios.

3. Becas de arancel internas y externas

Se entregaron 10 tipos de becas de arancel internas y externas a alumnos de primer año. La mayor cantidad de becas de arancel fueron entregadas por el MINEDUC con un total de 3.488 beneficiados en contraste con la Universidad de Chile que entregó 100 becas de arancel a alumnos de este cohorte. Gratuidad fue el beneficio con mayor alcance, con un total de 2.555 alumnos y un monto total de \$9.210 millones. Las 10 variables de arancel consideradas en el estudio son de tipo binaria y sin registros con datos faltantes. Para mayor detalle ver anexo Becas de Arancel Internas y Externas.

4. Becas de Mantención Internas y Externas

Se asignaron 16 tipos de becas de mantención internas y externas a alumnos de primer año. La beca con mayor cantidad de alumnos asignados corresponde a la Beca de Alimentación

para la Educación Superior (JUNAEB) con un total de 2.951 alumnos. Le sigue la Beca de Atención Económica (U. de Chile) con 2.576 beneficiados. Se observa un gasto total en becas internas de la Universidad de Chile de \$505 millones. Todas las variables de mantención son de tipo binaria y no poseen registros con valores faltantes. Para mayor detalle ver anexo Becas de Mantención Internas y Externas.

5. DEMRE

Se consideraron las pruebas de selección universitario (PSU) de Matemáticas, Lenguaje, Ciencias e Historia del año actual y anterior. Estas variables tienen un alto porcentaje de datos faltantes, debido a que los resultados de la prueba son válidos por dos años desde su realización. Se incluyeron además las variables Puntaje Ranking y NEM. Todas estas variables son de tipo real. Para mayor detalle ver anexo Resultados proceso PSU.

6.4.1.3. Tratamiento de datos faltantes

En esta sección se indica como se trataron los datos faltantes de cada una de las variables. Para ello se utilizaron reglas para reemplazo de valores, construcción de nuevas variables y categorización de variables continuas.

Las variables región y sexo en un inicio poseían valores faltantes, los cuales fueron completados de forma manual utilizando las columnas de ciudad de origen y nombre del alumno.

En tanto que los valores faltantes de las variables Procedencia Educacional, Puntaje PSU Matemáticas, Lenguaje, Ciencias e Historia, e Índice de Vulnerabilidad Escolar (IVE) se trabajaron como se detalla continuación:

■ Procedencia Educacional

Se reemplazaron los 29 datos faltantes de forma aleatoria entre colegio particular, subvencionado y municipal, manteniendo las probabilidades de ocurrencia de acuerdo a la proporción de cada una en la muestra (30,8 %; 40,7 % y 28,6 % respectivamente).

■ **PSU Matemáticas, Lenguaje, Ciencias e Historia, Ranking y NEM**

441 de un total de 6.254 alumnos de la muestra en estudio (7,1 %), ingresaron a una carrera utilizando el puntaje PSU que obtuvieron el año anterior (ver valores faltantes Puntaje Matemáticas y Lenguaje Actual en anexo Resultados proceso PSU). Esto es permitido ya que la prueba de selección universitaria posee una duración de dos años de validez para postular a las distintas carreras ². Dado este hecho, se construyó a partir de las variables PSU Actual y Anterior de cada rama (Matemáticas, Lenguaje, Ciencias e Historia) las variables binarias “PSU Final” de cada una de ellas (ver tabla 6.10).

TABLA 6.10: Nuevas variables PSU

Nombre	Tipo	Valores	Missing Values	Fuente
Puntaje Matemáticas Final	real	264 a 850	0	DEMRE
Puntaje Lenguaje Final	real	201 a 850	1	DEMRE
Puntaje Ciencias Actual	real	262 a 850	1.928	DEMRE
Puntaje Historia Anterior	real	292 a 836	2.722	DEMRE

En la tabla 6.10 se visualiza que la PSU de Ciencias e Historia siguen manteniendo un alto porcentaje de datos faltantes (31 % y 44 % respectivamente). Esto se debe a que estas pruebas son de carácter optativo (el alumno debe escoger al menos una ellas, a diferencia de las pruebas de Lenguaje y Matemáticas, que son obligatorias). Una opción es reemplazar los valores de las PSU de Ciencias e Historia con valores aleatorios pero este podría generar errores en los modelos debido a la gran cantidad de datos faltantes. Dado lo anterior y a que se cuenta con predictores relacionados más robusto en comparación (PSU de Lenguaje y Matemáticas), no se considerarán estas variables en el estudio.

El valor faltante de la variable PSU Lenguaje Final (ver tabla 6.10) fue reemplazado por el valor promedio de la misma variable, el cual corresponde a 661 puntos.

Para evitar efectos de multicolinealidad en los modelos, se descarta la variable Puntaje NEM,

²<http://www.psu.demre.cl/postulacion/como-postulo-a-una-universidad/uso-puntajepsu-admision-consecutivo>

ya que se encuentra directamente relacionada a la construcción de la variable Puntaje Ranking (ver Anexo Puntaje Ranking).

■ Índice de Vulnerabilidad Escolar (IVE)

A partir de la tabla 6.11 “Atributos Guía Matrícula e Índice de Vulnerabilidad Escolar” se identifica que el puntaje IVE posee 2.005 datos faltantes de una muestra total de 6.254 alumnos (32 %). Esto se debe a que la JUNAEB califica con puntaje IVE sólo a los colegios municipales y subvencionados (y no a particulares pagados). Para reemplazar los datos faltantes se utilizaron dos métodos:

1. IVE (1): Reemplazo de missing values por valores aleatorios entre 0 y 13%, dando como resultado una variable de tipo real (se utiliza el supuesto de que los colegios pagados en Chile poseen un menor porcentaje de alumnos vulnerables en comparación a colegios subvencionados y municipales, y dado que el porcentaje mínimo de IVE de estos colegios obtenidos en la muestra corresponde a un 13% (ver tabla 6.11), se utiliza esta cota).

2. IVE (2): Categorizar la variable según norma establecida por DBE:

- IVE RANGO 1: Entre 76,7% y 100% de Vulnerabilidad Escolar.
- IVE RANGO 2: Entre 53,4% y 76,6% de Vulnerabilidad Escolar.
- IVE RANGO 3: Entre 30% y 53,3% de Vulnerabilidad Escolar.
- SIN RANGO IVE: Alumnos provenientes de colegios particulares pagados.

TABLA 6.11: Nuevas variables “IVE (1) e IVE (2)”

Nombre	Tipo	Valores	Missing Values	Fuente
IVE (1)	Real	0 a 95 %	0	JUNAEB
IVE (2)	Catagórica	IVE RANGO 1 (354) IVE RANGO 2 (1.452) IVE RANGO 3 (1.957) SIN RANGO IVE (2.491)	0	JUNAEB

El motivo de utilizar estos dos métodos, es utilizar la variable de tipo real en modelos de clusterización que utilicen la distancia euclidiana como medida de separación, y utilizar la variable categórica en métodos de clusterización basados en similitud.

6.4.1.4. Construcción de nuevas variables

Luego del reemplazo de los datos faltantes, se procedió a construir las siguientes variables:

1. **Deserción:** Variable binaria construida a partir de la situación académica del alumno la cual toma valor 1 si el alumno es un desertor de la Universidad de Chile y 0 si no.
2. **Enfermedad Permanente o crónica:** Variable binaria construida a partir de las variables enfermedad permanente y enfermedad crónica, que toma valor 1 si el alumno posee un enfermedad permanente o crónica.
3. **Hacinamiento:** Variable categórica construida a partir de número dormitorios y número de integrantes de hogar, donde si el índice $\frac{\text{Número Dormitorios}}{\text{Número Integrantes Hogar}}$ es menor que 2 se considera “sin hacinamiento”, si es menor 3 pero mayor o igual a dos se considera “hacinamiento medio”, y mayor o igual a 3 se considera “hacinamiento alto”.
4. **Nivel de educación familiar:** Variable categórica construida a partir de las variables nivel educacional jefe de hogar y nivel educacional de la madre, la cual posee valores básica, media completa, técnico completa y universitaria completa.
5. **Región:** Variable categórica que agrupa la región de origen del alumno en alumnos provenientes de la Región Metropolitana (RM), Región de Valparaíso y Bernardo O’Higgins, y otras regiones.
6. **Carrera año anterior:** Variable binaria construida a partir de datos históricos de alumnos de los años 2011 a 2016, donde se corroboró si los alumnos de primer año del cohorte 2017 habían pertenecido a otra carrera de la universidad sin haberla terminado, tomando valor 1 si el alumno posee una carrera no terminada anterior, y 0 si no.
7. **CAE-FSCU:** Variable binaria construida a partir de las variables CAE y FSCU, donde si el alumno posee ambas becas la variable toma valor 1, y 0 de lo contrario.

-
8. **Becas de Arancel Tipo I:** Corresponde a becas que exigen haber tenido un promedio de PSU en Matemáticas y Lenguaje de al menos 700 puntos, NEM superior a 6,0 o pertenecer al 10% con resultados más altos de su establecimiento. Si el alumno posee Beca Universidad de Chile, Andrés Bello, Puntaje PSU o Excelencia Académica, esta variable toma valor 1. De lo contrario, toma valor 0.
 9. **Becas de Arancel Tipo II:** Corresponde a becas que exigen un promedio de PSU en Matemáticas y Lenguaje de al menos 500 puntos y NEM 5,5. Si el alumno posee beca Bicentenario, Beca Hijo de Profesionales de Educación, Becas las Condes o Reparación Valech toma valor 1. De lo contrario, toma valor 0.
 10. **Porcentaje de cobertura créditos:** Variable real que indica el porcentaje de arancel cubierto por créditos.
 11. **Porcentaje de cobertura becas:** Variable real que indica el porcentaje de arancel cubierto por becas.
 12. **Beca de residencia:** Variable binaria que identifica si el alumno posee una beca de residencia. Toma valor 1 si el alumno posee Beca de Residencia Interna, Beca de Residencia Sistema Prioritario de Equidad Educativa (SIPEE) u Hogares Universitarios. De lo contrario, toma valor 0.
 13. **Beca de libre disposición y de alimentación:** Variable binaria que identifica si el alumno posee una beca de mantención para cubrir gastos de transporte, alimentación u otros que pudiera requerir el alumno para el desarrollo de sus estudios. Si el alumno posee Beca de Atención Económica (BAE), Beca de Alimentación para la Educación Superior (BAES), Beca de Mantención Educación Superior (BMES), Beca de Mantención Vocación del Profesor o Beca Presidente del a República, toma valor 1. De lo contrario, toma valor 0.
 14. **Beca de pueblos originarios:** Variable binaria que identifica si el alumno posee becas destinadas a inclusión indígena. Si el alumno posee Beca Indígena o Beca Residencia Indígena, toma valor 1. De lo contrario, toma valor 0.
 15. **Beca de integración territorial:** Variable binaria que identifica si alumnos de zonas extremas del país poseen becas para cubrir gastos de viajes u otros que pudieran requerir. Si

el alumno posee Beca Integración Territorial o Beca Patagonia Aysén, toma valor 1. De lo contrario, toma valor 0.

El detalle de los motivos y métodos de construcción de cada una de estas se detallan en el anexo Creación de Nuevas Variables.

6.4.1.5. Distribución de variables

Ciertos modelos y algoritmos requieren que sus variables distribuyan como una normal y sean simétricas. En el caso de de la regresión logística, se requiere que las variables numéricas distribuyan como una normal, y si no lo son, que al menos sean simétricas [3]. En el caso de los algoritmos de clusterización basados en la distancia euclidiana, se requiere que las distancias estén medidas bajo el mismo estándar de medición, ya que de lo contrario la aglomeración de los datos puede verse afectada. Por ejemplo, las escalas de “Cobertura Arancel”, la cual varía entre 0 y 1, y la variable “PSU Matemáticas Final”, la cual varía entre 264 a 850, deben ser estandarizadas para alcanzar resultados correctos con métodos de clusterización.

Para normalizar las variables numéricas se puede recurrir a transformaciones según la distribución presentada por la variable. En el caso de variables que posean un sesgo o una “cola” hacia la izquierda (*left skewness*), se recomienda utilizar transformaciones del tipo cuadrática (x^2), raíz cúbica ($\sqrt[3]{x}$) o logarítmica ($\log(x)$). Para distribuciones con un sesgo o “cola” hacia la derecha (*right skewness*), se recomienda utilizar transformaciones del tipo raíz cuadrática (\sqrt{x}), raíz cúbica ($\sqrt[3]{x}$) o logarítmica ($\log(x)$)³.

Se obtuvo como resultado que las variables PSU Matemática Final, PSU Lenguaje Final y Puntaje Ranking distribuyen como una normal.

La variable Puntaje IVE no distribuye como una normal, pero aplicando una transformación raíz cúbica logra cumplir con los criterios de normalidad. Pese a ello se utilizará por preferencia su versión categórica (IVE 2) para obtener resultados más robustos.

Las variables Porcentaje Cobertura Crédito y Porcentaje Cobertura Arancel fueron descartadas del

³<https://medium.com/@TheDataGyan/day-8-data-transformation-skewness-normalization-and-much-more-4c144d370e55>

estudio por no cumplir requisitos de normalidad y por poseer una mejor representación a través de variables binarias.

Las variables PSU Lenguaje, PSU Matemáticas, PSU Ranking e IVE (1) fueron estandarizadas con una distribución normal de media 0 y varianza 1 ($N(0, 1)$).

El detalle de los análisis de cada una de las variables se encuentra en el anexo Distribución de Variables.

6.4.1.6. Relación entre variables

Se utilizó el análisis de correlación para explorar la relación entre las variables predictoras con la variable predecida (deserción) y entre variables predictoras. Esta es una herramienta que permite simplificar el análisis exploratorio, ya que permite focalizar el estudio en aquellas variables que se encuentren más correlacionadas, en vez de realizar contrastes con cada uno de los atributos, recordando que se trata sólo de análisis exploratorio, no indicando una relevancia estadística con deserción.

La correlación es de tipo real, varía entre -1 y 1, y opera sólo con variables de tipo real y binarias (no categóricas). Una correlación positiva significa que a medida que una variable aumenta su valor, la otra también aumenta. En tanto que una correlación negativa, significa que a medida que una variable aumenta su valor, la otra la disminuye.

■ Correlación entre variables predictoras y deserción

Las variables que estuvieron más correlacionadas con la variable deserción institucional fueron:

TABLA 6.12: Correlación con variable deserción

Variable	Correlación
PSU Matemática Final	-0.115
PSU Lenguaje Final	-0.061
CAE	-0.058
Carrera Anterior No Terminada	-0.039
CAE - FSCU	-0.032

Para determinar si el promedio de PSU de Matemáticas de alumnos que no desertan es superior al promedio PSU de Matemáticas que desertan, se aplicó un test t de student de diferencia de medias en stata (ttest). Este determinó que el promedio de PSU de Matemáticas de alumnos que no desertan es superior de forma estadísticamente significativa ($p=0.000$). Lo mismo se aplicó para PSU de Lenguaje, obteniendo también que el promedio de PSU de Lenguaje de alumnos que no desertan es superior al promedio de alumnos que desertan de forma estadísticamente significativa ($p=0.000$).

Para determinar si existe asociación entre la variable CAE y deserción, se aplicó un test de Pearson chi-cuadrado en stata (chi2). El resultado fue que la variable CAE se relaciona con deserción de forma estadísticamente significativa ($p=0.000$). Lo mismo se aplicó para Carrera Anterior No Terminada y CAE FSCU, dando como resultado que estos se encuentran relacionados de forma estadísticamente significativa ($p=0.002$ y $p=0.010$ respectivamente). Como todas estas variables poseen una correlación negativa con deserción (ver tabla 6.12) y se encuentran además relacionadas con deserción de forma estadísticamente significativa, se puede decir que alumnos con CAE, que provengan de una carrera anterior de la universidad o que posean la combinación de créditos CAE- FSCU desertan en menor proporción. Las proporciones de cada una de estas variables se detallan en las figuras 6.4, 6.5 y 6.6.

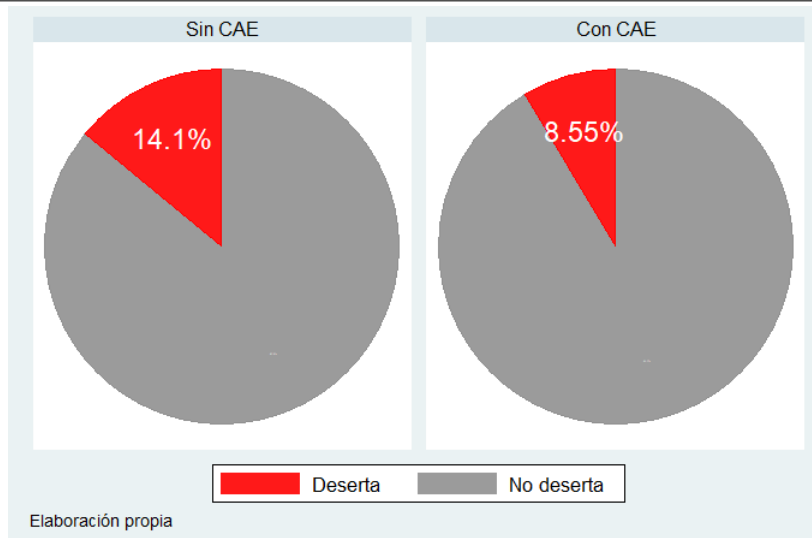


FIGURA 6.4: CAE y Deserción. Fuente: Elaboración propia.

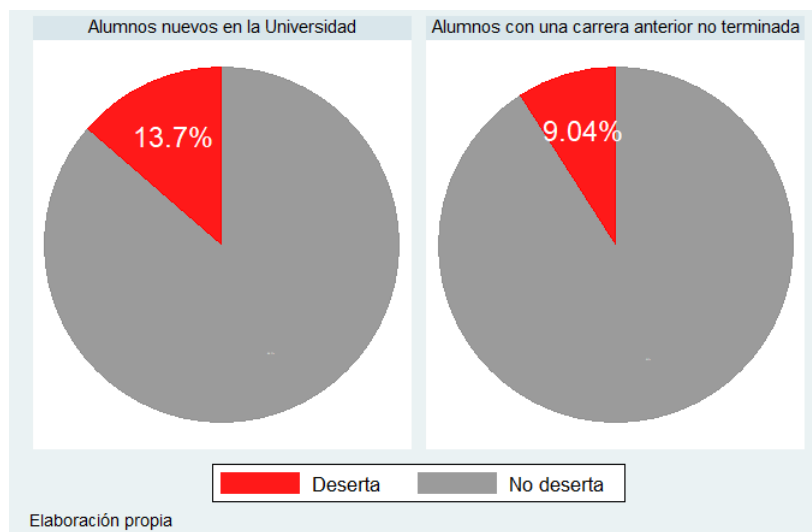


FIGURA 6.5: Carrera Anterior No Terminada y Deserción. Fuente: Elaboración propia.

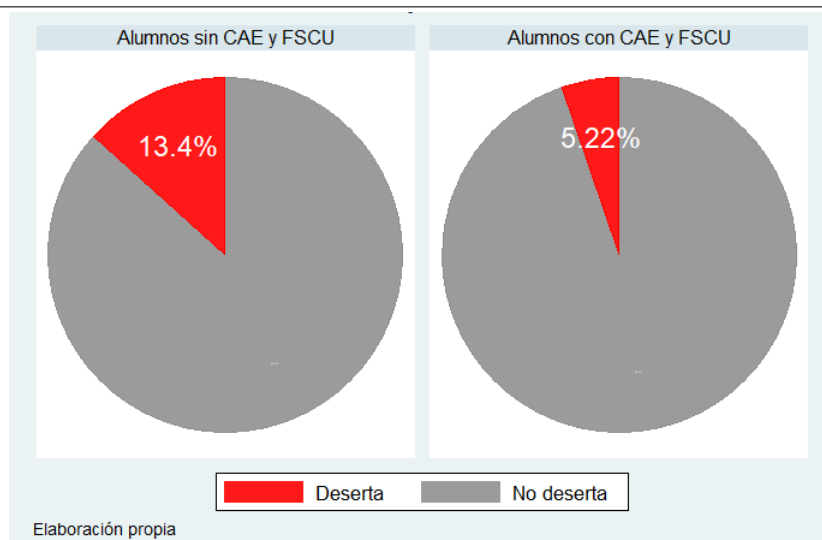


FIGURA 6.6: CAE-FSCU y Deserción. Fuente: Elaboración propia.

■ Correlación entre variables predictoras

Las 10 variables predictoras que se encontraban más correlacionadas entre si se detallan en la tabla 6.13.

TABLA 6.13: Correlación entre variables predictoras

Variable 1	Variable 2	Correlación
Gratuidad	B. Libre Disp. y Alim.	0.75
CAE-FSCU	FSCU	0.62
IVE (1)	B. Libre Disp. y Alim.	0.56
IVE (1)	Gratuidad	0.48
PSU Ranking	Flag PSU Ranking	0.48
Gratuidad	CAE	-0.35
CAE	CAE-FSCU	0.33
IVE (1)	PSU Matemática Final	-0.32
PSU Lenguaje Final	PSU Matemática Final	0.31
Arancel Tipo II	Gratuidad	-0.30

- Gratuidad y Becas de Libre Disposición y Alimentación

Estas variables se encuentran fuertemente correlacionadas (0.75) debido a que todos los alumnos que poseen gratuidad reciben por parte de la universidad la beca de mantención BAE, siendo esta beca representativa dentro de los alumnos con becas de libre disposición y alimentación (2.555 de 3.446 alumnos, es decir, el 74 % de alumnos con becas de libre disposición y alimentación).

- CAE-FSCU y FSCU / CAE-FSCU y CAE

Estas variables se encuentran correlacionadas (0.62 y 0.33 respectivamente) debido a que CAE-FSCU corresponde a una construcción de la variables FSCU y CAE.

- IVE (1) y Becas de Libre Disposición y Alimentación / IVE (1) y Gratuidad

Estas variables se encuentran correlacionadas debido a que estos beneficios son asignados a estudiantes de contextos vulnerables.

- PSU Lenguaje Final y PSU Matemática Final

Alumnos que obtienen un alto puntaje en Matemática obtienen generalmente un alto puntaje en Lenguaje, y viceversa.

6.4.1.7. Selección de variables

Para hacer la selección se utilizó el método de *Forward Selection and Backward Elimination* (ver sección Método de selección de variables *Forward Selection and Backward Elimination*). La herramienta utilizada fue la función *stepwise logistic* de Stata (`stepwise pr(.20) pe(.10): logistic`)⁴. En esta función `pr` corresponde al nivel de significancia por el cual se elimina una variable, `pe` el nivel de significancia para agregar una variable al modelo y `logistic` el modelo utilizado. Este método seleccionó a las siguientes variables:

⁴<https://www.stata.com/manuals13/rstepwise.pdf>

TABLA 6.14: Variables seleccionadas por método *Forward and Backward Elimination*

Variable	Tipo
1. PSU Matemáticas Final	Real
2. PSU Lenguaje Final	Real
3. Tramo	Categórica
4. Región	Categórica
5. Ive (2)	Categórica
6. Carrera	Categórica
7. Carrera Anterior No Terminada	Binaria
8. CAE-FSCU	Binaria
9. Gratuidad	Binaria
10. B. Vocación Profesor	Binaria

Es importante en este punto rescatar que variables que por intuición podría decirse que influyen en la deserción universitaria, como nivel de hacinamiento, nivel de educación de los padres o la presencia de una enfermedad crónica o permanente en la familia, no fueron seleccionados por este modelo. Esto se corroboró con análisis estadísticos aislados (chi-cuadrado) obteniendo que estas no son estadísticamente significativas.

6.4.2. Análisis de Cluster

El análisis de cluster se realizó con el método de X means y DBSCAN, considerando sólo las variables seleccionadas por el método de *Forward Selection and Backward Elimination*. Cómo estos métodos se basan en la medida de distancia para aglomerar las entidades, se separó el estudio en dos grupos:

1. **Estudio de variables reales:** Estudios de variables IVE(1), PSU Matemáticas Final y Lenguaje Final utilizando la distancia euclidiana.
2. **Estudio de variables binarias y categóricas:** Estudio de las demás variables seleccionadas

utilizando la distancia nominal de Rapidminer, la cual toma valor 0 si dos strings son iguales, y 1 si no.

6.4.2.1. Resultados clusters con variables Reales

Los cluster generados con la variable X-Mean no fueron significativos, debido a que se detectaron dos clusters con tasas de deserción institucional equivalentes a la muestra (13%).

En tanto que el método DBSCAN generó cuatro clusters, los cuales se observan en la figura 6.7.

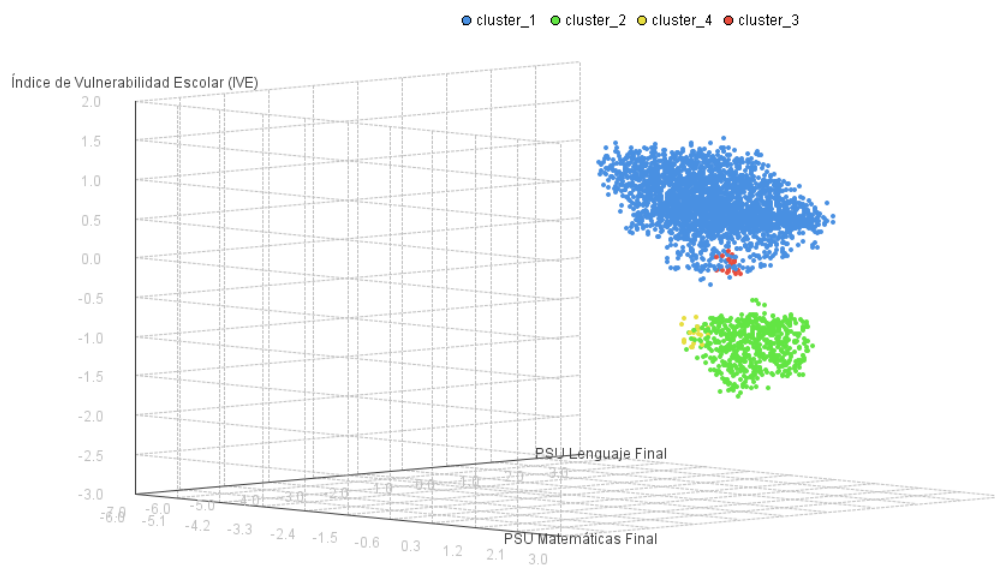


FIGURA 6.7: DBSCAN. Fuente: Elaboración propia.

En cuanto a los desertores de cada cluster, se puede apreciar en la imagen 6.8 que hay una mayor concentración de alumnos que deserta en el cluster 1.

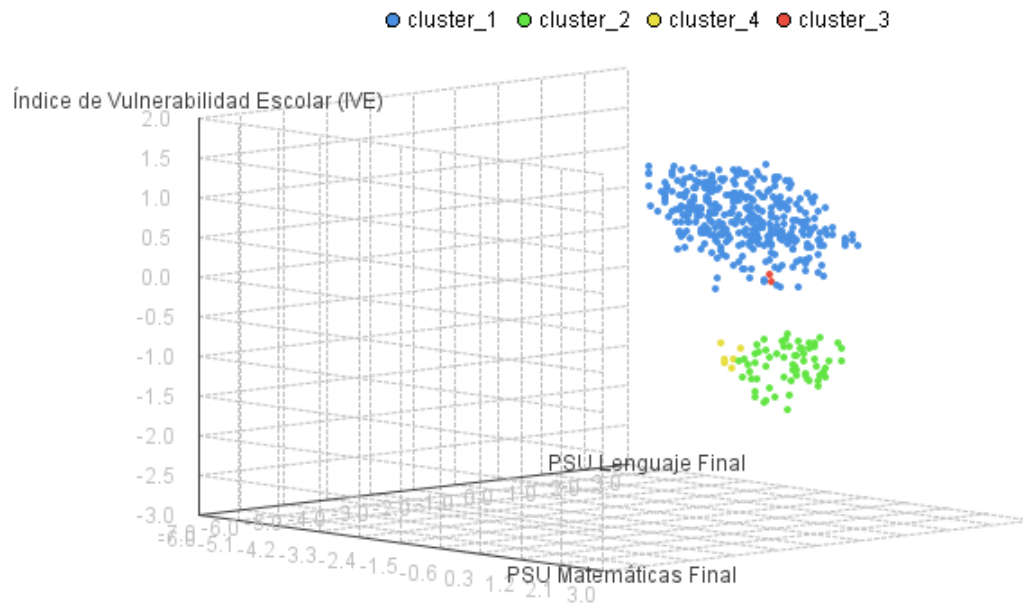


FIGURA 6.8: DBSCAN - Desertores. Fuente: Elaboración propia.

Empíricamente, el cluster 1 (ubicado en la zona superior de la figura 6.8) deserta un 13,4% (357 alumnos) y en tanto que en el cluster 2 (ubicado en la zona inferior de la misma figura) un 11,8% (65 alumnos), comprobando que la tasa de deserción de alumnos de cluster 1 es mayor que el cluster 2. Es por ello que las etiquetas generadas por estos cluster fueron incorporadas como variables. Posteriormente se incluyeron dentro del set de variables seleccionadas por el método de *Forward Selection and Backward Elimination*, pero esta variable no fue seleccionada por el algoritmo, siendo descartada para predecir la deserción.

Para ver más detalles del proceso de clusterización con variables reales ver anexo Proceso clusterización con variables reales.

6.4.2.2. Resultados clusters con variables Categóricas

Utilizando la distancia nominal entre variables categóricas con el algoritmo X-means se obtuvieron 4 clusters, pero los resultados fueron descartados del estudio al contar con clusters correlaciones un 100% con tener Beca Vocación Profesor, Gratuidad y CAE, no agregando valor a las variables ya existentes.

Mientras con el algoritmo DBSCAN se encontraron 2 clusters:

- Cluster 1: 1.996 registros. Corresponde a alumnos que poseen ingresos económicos bajos (tramo 40 y 50) y Gratuidad. Más del 40 % de los alumnos de las carreras de Medicina, Ingeniería Comercial e Ingeniería y Ciencias Plan Común pertenecen a este cluster. No posee diferencias relevantes con respecto a región.
- Cluster 2: 1.857 registros. Corresponde a alumnos que poseen ingresos económicos altos (tramo 90, 100 y sin calificación económica principalmente) y CAE. No poseen gratuidad. Más del 40 % de los alumnos de las carreras de Administración Pública, Bioquímica, Fonoaudiología, Geografía, Ingeniería Forestal, Kinesiología, Medicina Veterinaria, Nutrición y Dietética, Obstetricia y Puericultura, Pedagogía en Educación Parvularia, Química y Farmacia, Terapia Ocupacional y Trabajo Social pertenecen a este cluster. No posee diferencias relevantes con respecto a región.
- El resto de los registros considerados como ruido (2.401 registros= 38 % de la muestra).

De la misma forma que anterior análisis, estos cluster pueden ser nombrados como “alumnos vulnerables”(cluster 1) mientras que el cluster 2 como “alumnos no vulnerables”. De la misma manera que los clusters de tipo real, se creó una variable que identifica a los alumnos que pertenecen a cada uno de los cluster. Aplicando el algoritmos de *Forward Selection and Backward Elimination*, y al igual que el cluster generado con variables reales, esta fue descartada como una variable significativa para predecir la deserción.

6.4.3. Procedimiento Análisis Predictivos

Una vez realizado el análisis de cluster, se procede a realizar los análisis predictivos, el cual incorpora diferentes aspectos de minería de datos. El detalle de cada uno de estos aspectos se detalla a continuación:

- **Variables**

Se consideraron todas las variables seleccionadas con el método de *Forward Selection and Backward Elimination* (ver tabla 6.14).

■ Cross Validation

Para evitar el sobreajuste de los modelos, se utilizó una partición de datos según el método de Validación Cruzada con $k=10$, es decir, se considerará como set de entrenamiento el 90 % de los datos y el 10 % como set de evaluación.

■ Balanceo de Datos

Los algoritmos de Árboles de Decisión y Random Forest son sensibles a datos desbalanceados. En este caso, como desertores corresponde sólo al 13 % de la muestra, se requiere un balanceo de los datos para que el modelo no sobreestime la clase dominante. De esta forma, con la herramienta "*sample*" de rapidminer, se iguala la proporción de registros desertores y no desertores (828 desertores y 828 no desertores). Este operador debe ser incluido en el set de entrenamiento de Validación Cruzada.

■ Tuning / Ajuste de Parámetros

Como los modelos de Árboles de Decisión y Random Forest poseen muchos parámetros que deben ser ajustados, se utilizó la herramienta de *Optimize Parameters* de Rapidminer, el cual permite obtener una grilla con los resultados del modelo utilizando las diferentes combinaciones de los parámetros. Por ejemplo, si para un tipo de árbol de decisión se definen 3 tamaños diferentes de "*minimal size for split*" y 4 tamaños diferentes de "*minimal leaf size*", la herramienta arrojará una grilla con $3 * 4 = 12$ resultados diferentes. Se fueron ajustando los parámetros según las combinaciones de parámetros que poseían un mejor performance hasta llegar a un nivel donde el modelo sólo mejora marginalmente los resultados (menores a 1 %).

■ AUC

Una vez escogido la mejor combinación de parámetros, se escogió el modelo con los parámetros que poseían un mayor AUC.

■ Matriz de confusión/umbrales

Una vez ajustado los parámetros a los modelos que poseen un mayor AUC, se calcularon los valores de matriz de confusión (VP, VN, FP, FN) para cada modelo según 11 umbrales diferentes: 0; 0, 1; 0, 2; ...; 0, 9; 1. Un umbral permite definir cuando una predicción será considerada positiva o negativa, utilizando la siguiente regla:

$$\text{valor predicción} < \text{umbral} \implies \text{valor predicción} = 0$$

$$\text{valor predicción} \geq \text{umbral} \implies \text{valor predicción} = 1$$

A partir de la matriz de confusión, se calcularon las métricas *Accuracy*, *Precision* y *Recall*.

■ **Costos de clasificar de forma incorrecta a desertores (*misclassification costs*)**

Los errores de clasificación de los modelos corresponde a error tipo I y tipo II. Cada uno de estos posee asociados los siguientes costos:

- **Costos de Error de Tipo I**

Este error corresponde a predecir que un alumno desertará cuando en realidad no lo hará. El costo promedio anual de asignar un beneficio a un alumno por parte de la Universidad de Chile corresponde a \$700.000 (considerando sólo becas de mantención y no becas de arancel). Considerando un promedio de 6 años que se demora el alumno en terminar su carrera y una tasa de descuento de 10 %, se obtiene un costo total por la asignación de beneficios (en valor presente) de \$ 3.048.682.

- **Costos Error de Tipo II**

Este error corresponde a predecir que un alumno no desertará cuando en realidad si lo hace. Este costo puede ser calculado como los aranceles que se dejarán de percibir por la deserción del alumno. Considerando el costo promedio de arancel de los alumnos de la muestra (\$3.914.092), un periodo promedio de 6 años que el alumno permanece en la universidad y una tasa de anual del 10 %, se obtiene un costo total por ingresos no percibidos de \$17.046.891.

■ **Casos extremos**

Para justificar los modelos de minería de datos se requiere conocer los beneficios de estos en términos tangibles. Una forma sencilla es realizar un contraste entre los modelos y dos situaciones extremas: ayudar a todos los estudiantes con becas o no ayudar a nadie.

- Ayudar a todos los estudiantes provoca asignar ayudas de mantención por \$700.000 anuales por 6 años a 5.426 alumnos que no desertarán (error de tipo I), lo cual corresponde a \$16.542 millones.
- No ayudar a nadie provoca perder los ingresos futuros por concepto de aranceles por \$3.914.092 por 6 años de 828 alumnos que eran reales desertores que no fueron ayudados (error de tipo II), lo cual corresponde a \$14.114 millones.

Considerando estos dos casos, el caso económicamente óptimo sería los costos generados por error de tipo II por \$14.114 millones.

■ Elección de modelo

En cada modelo se escogió el umbral que minimiza el costo, y en caso de haber dos umbrales que minimizan el costo, el que poseen un mayor *Recall*.

6.4.4. Resultados y Análisis de Modelos Predictivos

En la tabla 6.15 se detallan los resultados de los modelos seleccionados de los 3 tipos de árboles de decisión (ID3, CART y CHAID), Random Forest y Regresión Logística.

TABLA 6.15: Resultados modelos seleccionados

Modelo	ID3	CART	CHAID	Random Forest	Regresión Logística
AUC	65,1 %	57 %	64,7 %	67,8 %	69 %
Tiempo de Procesamiento	5 min 1 s	57 s	24 min 16 s	1h 28 min	6 s
Cantidad de modelos	324	384	256	1728	1
Tiempo por modelo (segundos)	0,9	0,1	5,7	3,1	6
Umbral	1	1	0,9	0,8	0,2
VP	18	192	0	0	337

VN	5283	3903	5426	5426	4488
FP (Error Tipo I)	143	1523	0	0	938
FN (Error Tipo II)	810	636	828	828	491
Accuracy	85 %	65 %	87 %	87 %	77 %
Precision	11 %	11 %	0 %	0 %	26 %
Recall	2 %	23 %	0 %	0 %	41 %
Fmeasure	4 %	15 %	0 %	0 %	32 %
Costo (millones)	\$14.244	\$15.485	\$14.115	\$14.115	\$11.230

- **AUC:** Regresión Logística con un mayor AUC (69%). Por otro lado, Random Forest posee un AUC superior a los otros 3 modelos de árboles de decisión, lo cual demuestra que Random Forest es más estable y robusto que árboles de decisión aislados. Esto se debe a que Random Forest genera muchos árboles de decisión que poseen distintos set de registros y atributos.

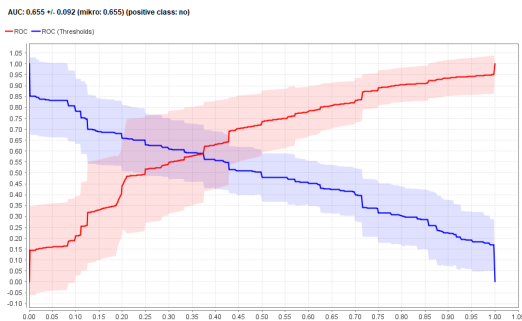


FIGURA 6.9: ROC y AUC ID3. Fuente: Elaboración propia.

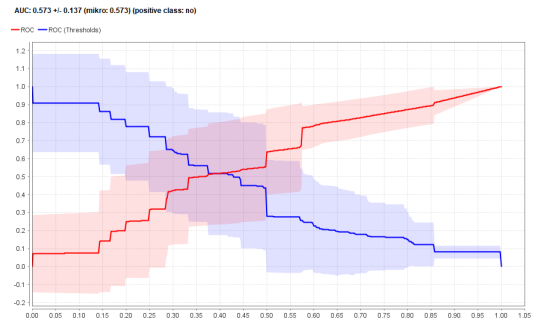


FIGURA 6.10: ROC y AUC CART. Fuente: Elaboración propia.

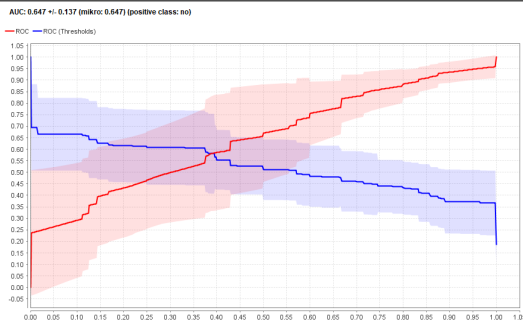


FIGURA 6.11: ROC y AUC CHAID. Fuente: Elaboración propia.

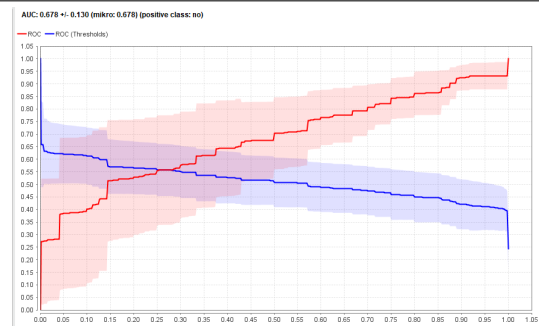


FIGURA 6.12: ROC y AUC Random Forest. Fuente: Elaboración propia.

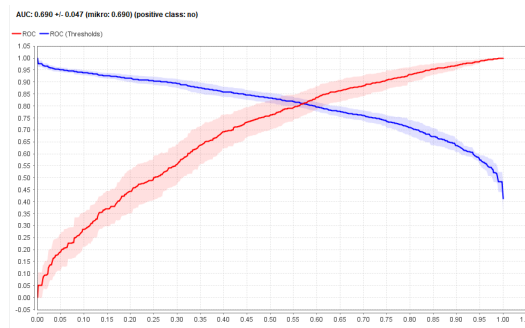


FIGURA 6.13: ROC y AUC Regresión Logística. Fuente: Elaboración propia.

- **Tiempo procesamiento:** El tiempo de procesamiento es mayor en Random Forest. Esto se debe a que posee una mayor cantidad de parámetros que se deben ajustar y también que a medida que aumenta el número de árboles más lento se vuelve el procesamiento.
- **Umbral:** Regresión logística indica que si existe una probabilidad superior a 20% de ser un desertor, este debe ser calificado como tal. Random Forest y Chaid se encuentran en el 80% y 90% respectivamente. ID3 y CART sólo califican como desertores aquellos que posean una probabilidad 1 de serlo, siendo estos últimos resultados inconsistentes con la realidad.
- **Falsos Negativos (FN):** Para efectos de este estudio, el error de tipo II (FN) es más relevante que el error de tipo I. Este es menor en la regresión logística.
- **Recall:** Un alto *recall* permite mitigar los efectos del Error de tipo II (FN). En este caso, el *recall* más alto corresponde a regresión logística.
- **F-measure:** La relación *recall-precisión* es superior en regresión logística (32%).
- **Costo:** Regresión logística corresponde al modelo más económico (costo de \$11.230 millones)

6.4.5. Modelo seleccionado

El mejor modelo de la sección anterior corresponde al de Regresión Logística con umbral=0,2, el cual posee el mayor AUC (69 %), *recall* (41 %), *precision* (26 %) y *fmeasure* (32 %).

El modelo detecta la deserción de $828 - 491 = 337$ alumnos. Si a partir de la generación de acciones sobre estos alumnos se hubiese evitado la deserción de todos ellos (337 alumnos), hubiese disminuido la tasa de deserción en un 41 %, o de forma equivalente, haber evitado la deserción de un 5,4 % del total de alumnos. Esto hubiese generado ahorros por un total de \$2.884 millones con respecto a la situación sin modelo.

Una alternativa es escoger este mismo modelo pero con un umbral 0,1 , el cual posee un *recall* de 72 % y comete 234 error de tipo II, en contraste de los 491 cometidos por el modelo más económico.

Es decir, la segunda alternativa previe hubiese detectado la deserción de $828 - 234 = 594$ alumnos De la misma manera que en el caso anterior, si a partir de la generación de acciones sobre estos alumnos se hubiese evitado la deserción de los 594 alumnos, se hubise disminuido la tasa de deserción en un 35 %, o de forma equivalente, haber evitado la deserción de un 9 % del total de alumnos. Si bien al elección de un umbral menor detecta a más alumnos desertores, este es \$472 millones más caro que el modelo óptimo, debido a que también aumenta los errores en la predicción del modelo (error de tipo I y de tipo II).

Capítulo 7

Propuesta de apoyo tecnológico

En este capítulo se detallan las capacidades tecnológicas que se requieren para poder desarrollar la nueva capacidad para retener alumnos.

Las componentes tecnológicas requeridas son dos:

1. Software de Minería de Datos

Contar con un software especializado para utilizar modelos de minería de datos que permitan predecir la deserción de alumnos.

2. Software de Visualización de Datos

Contar con un software especializado en visualización de datos para poder realizar análisis exploratorios de los datos y presentar de forma atractiva los hallazgos encontrados por minería de datos.

3. Desarrollo de un Data Mart de Alumnos

Desarrollar un Data Mart de forma que se pueda acceder a datos de múltiples fuentes de información bajo un único estándar de calidad. La implementación de esta componente también posee los siguientes beneficios:

- Permiten mantener información actualizada de datos.
- Generan ahorro de tiempos de analistas al no tener que adquirir y procesar los datos de múltiples fuentes de información, focalizando sus tareas hacia el análisis.

-
- Disminuye el riesgo de cometer errores en los análisis, al contar con datos con un formato y estándar de calidad definido.
 - Es más fácil para realizar mantenciones debido a que trabaja con un set acotado de datos (ej. contar sólo con datos de alumnos, a diferencia de un Data Warehouse, que incluye datos de diversos departamentos).

7.1. Especificación de requerimientos

En esta sección se detallan los requerimientos funcionales y no funcionales que permitan soportar la operación del sistema de minería de datos.

7.1.1. Requerimientos funcionales

Se requiere que el sistema posea las siguientes funcionalidades:

1. Extraer datos de alumnos de sistemas internos y externos y cargarlos en DSA.
2. Proceso ETL (*Extract, Transform and Load*) desde DSA a Data Mart alumnos.
3. Utilizar datos de Data Mart Alumnos para realizar análisis exploratorios, de clusterización y correr modelos que permitan predecir la deserción.
4. Puesta en producción de algoritmos para predecir la deserción.

Los inputs de estos sistemas corresponden a datos de alumnos provenientes de sistemas internos (Guía, U-Cursos y U-Campus) y externos (MINEDUC, IVE, JUNAEB, DEMRE, entre otros).

Como output se generan análisis sobre deserción estudiantil y una listado de los alumnos con riesgo de desertar.

7.1.2. Requerimientos no funcionales

Se detallan en la tabla 7.1 los aspectos del sistema que se deben tener para poder llevar a cabo los requerimientos funcionales. Se utiliza para ello la nemotecnia FURPS: *Functionality* (Funcionalidad), *Usability* (Usabilidad), *Reliability* (Fiabilidad), *Performance* (Desempeño) y *Support* (Soporte).

TABLA 7.1: Requerimientos no funcionales

Funcionalidad	Cumplir con los requerimientos funcionales solicitados.
Usabilidad	Facilidad de uso por Data Scientist y DBA.
Fiabilidad	Respaldo de datos de forma continua. Recuperación ante caídas.
Desempeño	Disponibilidad continua de sistemas. Sizing acorde a datos que se estima procesar. Sistema escalable en el tiempo.
Soporte	Soporte realizado a través de la Dirección de Tecnologías de la Universidad

7.2. Arquitectura Tecnológica

La arquitectura TI corresponde a una arquitectura de dos capas, las cuales corresponden a la capa de datos, y a la capa de lógica de negocios y visualización. A la primera capa corresponden las fuentes de datos internas y externas de alumnos, un repositorio de datos transitorio (DSA) y un Data Mart de Alumnos. En tanto que la capa de lógica de negocios y visualización se conjuga dentro del software de Minería de Datos y una herramienta de visualización.

7.3. Diseño de aplicación

7.3.1. Casos de Uso

Se requiere que los sistemas sean capaces de realizar las siguientes acciones ¹.

1. Cargar datos de sistemas internos y externos en DSA

Se requiere que el DSA tenga la capacidad para integrarse a los sistemas internos (Guía, U-Cursos y U-Campus) y que permita además la carga de información de fuentes externas (becas y créditos de MINEDUC, becas e Índice de Vulnerabilidad Escolar (IVE) de JUNAEB, DEMRE, entre otros).

2. Cargar datos en Data Mart Alumnos

Se requiere que el Data Mart permita la carga de datos desde DSA, permitiendo en este proceso seleccionar variables, transformar datos según requerimientos del negocio, y aplicar filtros y reglas de limpieza. El acceso a información confidencial debe ser restringido en esta etapa, no cargando datos personales de los alumnos (dirección, nombres, teléfono de contacto, entre otros) y enmascarando datos sensibles (ej. RUT).

3. Procesar datos en ambiente de pruebas

El software de minería de datos debe ser capaz de leer los datos almacenados en el Data Mart, con el objetivo de probar diferentes algoritmos. Además, este software debe permitir seleccionar variables, aplicar filtros, reemplazar datos faltantes, crear gráficos y estadísticas simples para explorar la data.

4. Almacenar análisis

El software de minería de datos debe permitir guardar los análisis realizados. De esta manera se construyen los análisis sobre los ya existentes, evitando destinar tiempo y esfuerzos en trabajos ya realizados.

¹Elaboración de casos de uso a partir de publicación en LinkedIn “Data Science data architecture” del Dr. Olav Laudy, Chief Data Scientist de Causality Link <https://www.linkedin.com/pulse/data-science-architecture-dr-olav-laudy>

5. Procesar datos en ambiente de producción

El software de minería de datos debe ser capaz de leer los datos almacenados en el ambiente de producción del Data Mart para la puesta en marcha de los modelos de minería de datos (proceso conocido como *scoring*). Como output de este proceso se genera un listado de alumnos en riesgo de deserción.

6. Cargar datos en ambiente de producción del Data Mart Alumnos

El Data Mart debe permitir cargar datos desde el DSA al ambiente de producción, permitiendo seleccionar y adecuar los datos según lo requerido por el negocio.

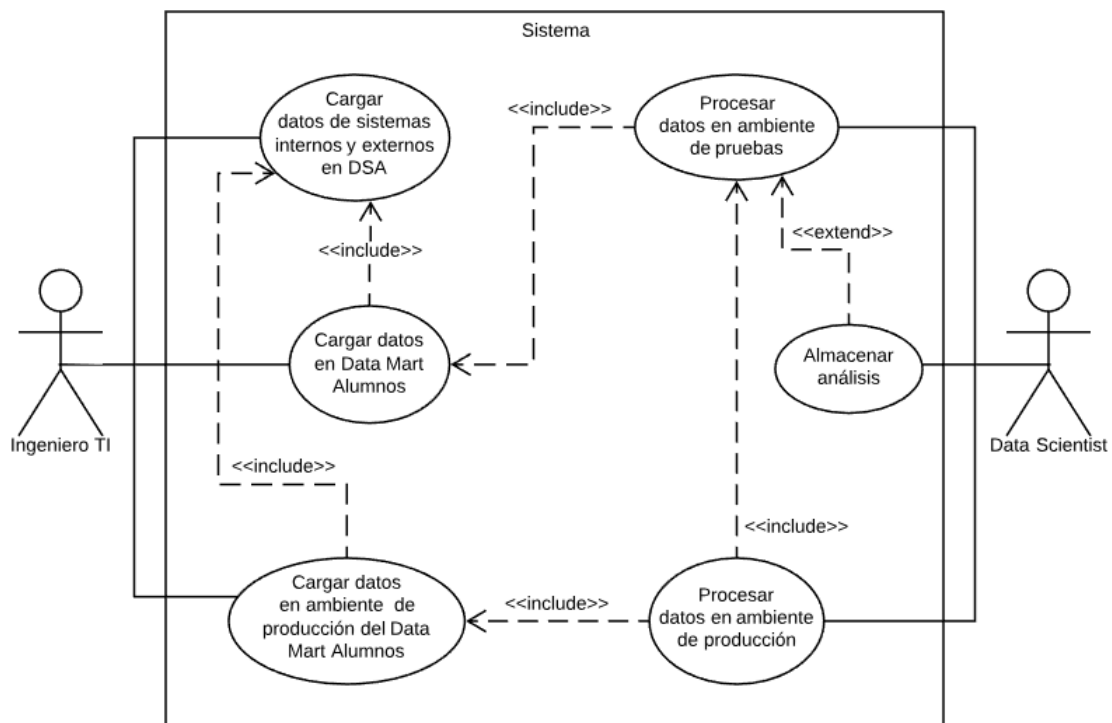


FIGURA 7.1: Casos de Uso. Fuente: Elaboración propia.

7.3.2. Diagrama de arquitectura del sistema

Como se mencionó anteriormente, el sistema de minería de datos está compuesto por las fuentes de datos internas y externas de alumnos, un repositorio de datos transitorio denominado *Data*

Staging Area (DSA), un Data Mart de Alumnos, software de Minería de Datos y una herramienta de visualización.

De las fuentes de datos tanto internas como externas se extraen los datos para luego ser cargados en DSA, previa aplicación de códigos de limpieza. Luego estas son cargadas a través de un proceso de *Extract, Transform and Load* (ETL) en el Data Mart de Alumnos, cargando de forma exclusiva aquellas variables que son relevantes de estudio y protegiendo datos sensibles de alumnos. Una vez poblado el Data Mart, se extraen los datos trabajados para ser utilizados por los software de Minería de Datos y Herramientas de Visualización.

El DSA y el Data Mart es administrado por un Ingeniero en Tecnologías de Información, el software de Minería de Datos por un *Data Scientist* y la herramienta de visualización por el *Analytics Manager*.

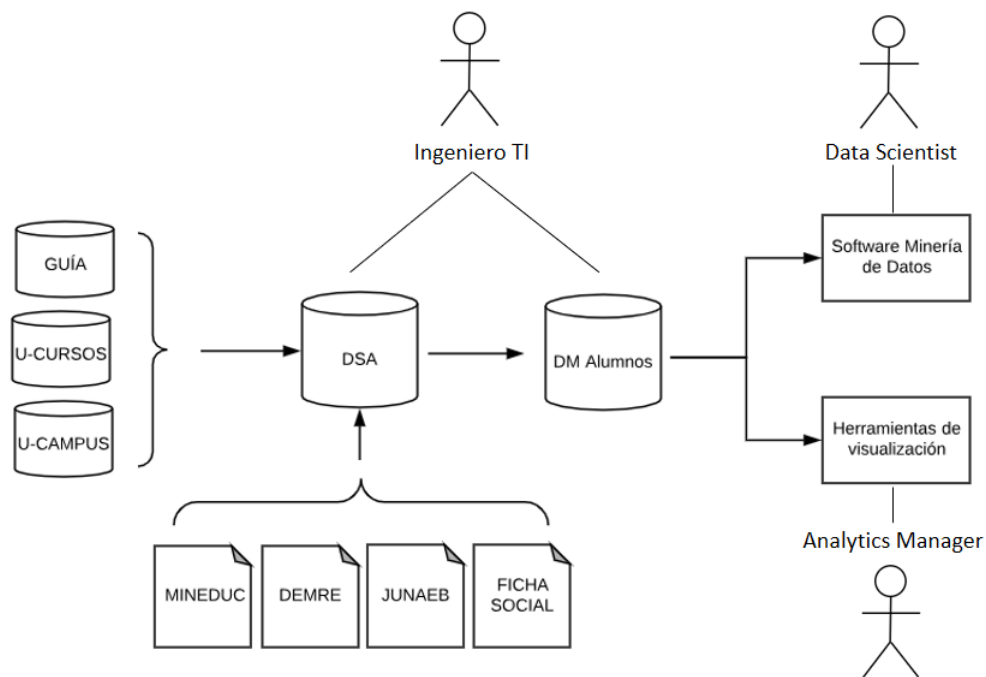


FIGURA 7.2: Arquitectura Tecnológica. Fuente: Elaboración propia.

7.3.3. Diagrama de despliegue

En la figura 7.3 se observa el diagrama de despliegue, el cual detalla la estructura física de la arquitectura propuesta. En el sector izquierdo se observan los sistemas guía, u-campus y u-cursos con sus respectivas bases de datos las cuales operan bajo Oracle versión 11g. Estas bases de datos se encontrarán integradas al DSA mediante protocolo TCP/IP en un red LAN para una mayor velocidad en la transmisión de datos. En la parte inferior se observa la información proveniente de fuentes externas (MINEDUC, DEMRE, JUNAEB) e internas (Ficha Social) en formato excel, lo cual será cargado de forma manual en el DSA. El DSA se encuentran integrado con el Data Mart alumnos también bajo el mismo protocolo y tipo de Red. Finalmente, el Data Mart de Alumnos es consumido por Software de Minería de Datos y Herramientas de Visualización a través de consultas al DM.

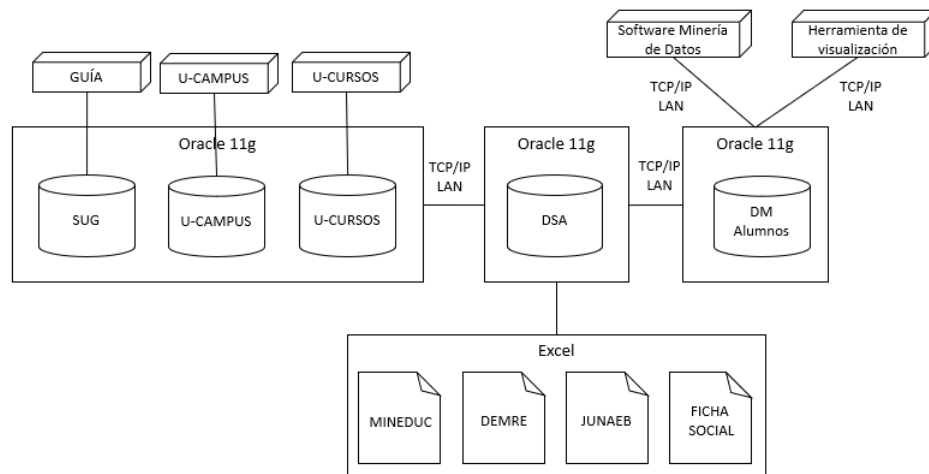


FIGURA 7.3: Diagrama de despliegue. Fuente: Elaboración propia.

Capítulo 8

Gestión del Cambio

En este capítulo se abordan los aspectos de gestión del cambio que deben ser considerados para una consecución exitosa del proyecto. Para ello se realiza un análisis en base al Modelo Integral de Liderazgo y Gestión del Cambio [14], para luego definir que aspectos son críticos dentro de este proceso y un plan para abordar la gestión del cambio.

8.1. Modelo Integral de Liderazgo y Gestión del Cambio

En este modelo se establece que existen diez dominios centrales de acción para liderar y gestionar proyectos de cambio tecnológico. El detalle de cada uno de estos dominios a continuación:

1. Liderazgo y gestión del proyecto de cambio

El liderazgo se hace cargo de generar las interpretaciones, contextos y estados de ánimo que permitan que el proyecto avance de forma adecuada, de darle un sentido, determinando el espacio de acciones y prácticas que se deben llevar a cabo durante todo el proyecto [14].

En este proyecto, el liderazgo debe ser llevado a cabo por el Director del Proyecto de Minería de Datos, el cual estará a cargo de posicionar a nivel directivo la relevancia de desarrollar esta nueva capacidad en la universidad y ser responsable de gestionar el proyecto (planificaciones y decisiones de todo el proceso). Este dirigirá al equipo de proyecto (Data Scientist, Analytics

Manager e Ingeniero TI) y poseerá el apoyo de expertos de negocio (asistentes sociales y profesionales en contacto con alumnos).

2. Estrategia y sentido del proceso de cambio

Este dominio indica que se debe establecer una comunicación clara de la dirección, sentido y forma que se llevará a cabo el proceso de cambio . Para ello es importante contar con relatos y narrativas que sean comprensibles y claros para todos los actores, para facilitar la comprensión del proceso e identificar los beneficios que se esperar lograr para la organización y las personas [14].

Los actores relevantes dentro de este proceso de cambio corresponden a los asistentes sociales y a directivos, ya que los primeros operacionalizan la nueva capacidad y los segundos toman las decisiones del proyecto.

A los asistentes sociales se le debe comunicar de forma detallada que aspectos se conservan y cuales son cambiados, lo cual es detallado en el dominio 3 de esta metodología, denominado “Cambio y Conservación”.

Mientras que a los directivos, se les debe transmitir de forma clara y comprensible los beneficios de la toma de decisiones en base a analítica, y dejar en claro que la instalación de esta nueva capacidad no sólo se traduce en una disminución de la deserción, sino en una transformación organizacional, al comprender que la mayoría de las funciones de negocio pueden ser abordadas con analítica, generando una apuesta en práctica global en el uso de datos en la toma de decisiones [11].

3. Cambios y conservación

Es importante en proyectos tecnológicos detallar que aspectos serán conservados y cuales cambiados. Estas declaraciones son relevantes ya que permiten bajar las resistencias, obstáculos, miedos y temores (sean estos fundados o no) y para dejar en claro que lo que se busca es seguir mejorando el desempeño de la organización [14].

■ Espacio de conservación

Se debe dar énfasis en que se conservan los programas vigentes de beneficios estudiantiles y los métodos actuales de evaluación de alumnos a través de la ficha social se conservarán, así como los procesos de generación de nuevos programas y planificación de asignación de

beneficios, con la salvedad que se agregarán los resultados de los análisis para contribuir al diseño de estos.

■ **Espacio de cambio**

Se cambian los procesos de análisis de comportamiento y predicción de la deserción, incorporando proceso de minería de datos, los cuales consisten en procesos ordenados de obtención y depuración de información, para luego aplicar diferentes algoritmos que permitan detectar patrones en los datos. Los resultados de los modelos validados en este proceso, serán utilizados para complementar la decisión de asignación de becas y poder contactar a aquellos alumnos que posean alto riesgo de desertar, para que así comprender sus necesidades en detalle, de forma de poder entregarle recomendaciones y facilitar el acceso a beneficios estudiantiles y otros programas (psicología personal, familiar, asistencia médica, tutorías u otro) que sean útiles a sus necesidades.

4. Organización y estructura del proyecto de cambio

Los principales actores y entidades que forman parte del proyecto de cambio, corresponden a la Directora del Departamento de Bienestar Estudiantil, el Director de Tecnologías de Información y el Director del Proyecto de Minería de Datos.

La Directora de Bienestar Estudiantil, debe aportar en este proceso generando el apoyo requerido al proceso, socializando el valor del proyecto y generando acuerdos con altos directivos, otros departamentos relacionados al proyecto, autoridades de facultades, asistentes sociales, entre otros.

El Director de Tecnologías debe prestar el apoyo político para que facultades accedan a dar su apoyo al proyecto en términos tecnológicos, logrando el acceso a integración con sistemas y a datos de alumnos. Además, generar acuerdos con facultades en temáticas de gobiernos de datos y administración de sistemas, especialmente en lo que respecta a datos y sistemas de alumnos.

El Director de Proyecto debe estar a cargo de liderar el proceso de cambio, siendo este el interlocutor con asesores externos de gestión del cambio en caso de solicitar servicios de consultoría, preocupándose de que hitos y planes se cumplan dentro de los plazos propuestos.

5. Gestión emocional

Se deben diseñar instancias formales para entregar información sobre el proyecto y los impactos que tendrá en la operación diaria. En estas instancias se resolverán todas las dudas que pudiesen surgir con respecto al proyecto, buscando generar estados de ánimo de confianza, apropiación y compromiso con el proyecto.

Durante el proceso de puesta en marcha, se deberán comunicar de forma estratégica los éxitos alcanzados en materias de retención de alumnos y el cumplimiento de hitos relevantes, de forma de que se genere una percepción de avance y logro que motive a seguir realizando esfuerzos para alcanzar los objetivos propuestos.

6. Comunicaciones

Se deben comunicar de forma estratégica:

- a) Los hallazgos encontrados en materia de minería de datos.
- b) Los nuevos programas que se han desarrollado a partir de los análisis realizados.
- c) La experiencia de alumnos que han sido beneficiados gracias a haber sido contactados de forma proactiva.
- d) Los aumentos de tasas de retención.

Esto será comunicado a través de reuniones, informativos y página web (de proyecto o institucional), donde se recogerán apreciaciones y resolverán dudas.

7. Desarrollo de habilidades

Se deben diseñar mecanismos de entrenamiento y formación de habilidades en técnicas de minería de datos y trabajo de equipo.

Aprender técnicas de minería de datos es un proceso intenso, debido a que se debe aprender a manejar múltiples herramientas tecnológicas y algoritmos matemáticos, además de tener los conocimientos de las necesidades de negocio de la DBE. Este proceso debe ser diseñado a través de una ruta de aprendizaje y evaluado de forma periódica. La ventaja es que actualmente esta temática es abordada en cursos en línea tanto gratuitos como pagados de calidad (Ej. cursos dictados por Harvard y Stanford) y múltiples páginas para compartir material sobre minería de datos y resolver dudas (ej. KDnuggets) facilitando de esta manera el proceso de aprendizaje.

En cuanto a trabajo de equipo, como el desarrollo de nuevos programas de ayuda estudiantil requiere de la interacción de múltiples disciplinas (tecnologías de información, ingeniería, trabajo social, sociología, entre otros) se deben establecer actividades y cursos que permitan integrar las visiones de estas disciplinas en un trabajo conjunto.

8. **Gestión de poder**

Un poder relevante en este proyecto corresponde a aquellos que son dueños de las bases de datos y herramientas que generan datos de alumnos. Es por ello, que el Director de Minería de Datos debe ser capaz de involucrar a los dueños de estos sistemas para poder crear un repositorio de datos común de alumnos. Para lograr estos resultados, es necesario contar con el apoyo del Director de Tecnologías de Información de la Universidad (DSTI), para así poseer el respaldo necesario para acceder a datos e información de forma continua.

Otro poder relevante corresponde al apoyo del Vicerrector de Asuntos Comunitarios y Estudiantiles (VAEC), para que respalde en todo momento la realización del proyecto.

También se debe gestionar los acuerdos y formas de trabajar con el Departamento de Pregrado, para así potenciar las expertices de ambas áreas (Pregrado y DBE).

Por último se deben mantener resultados demostrables, para asegurar el financiamiento. Ello debe ser gestionado con el Vicerrector de Asuntos Económicos y de Gestión Institucional (VAEGI).

9. **Monitoreo y evaluación del proceso**

Se debe monitorear los procesos de gestión del cambio, evaluando la adopción de herramientas de minería de datos, si la toma de decisiones se realiza en base a datos, si las asitentes sociales han contactado y entregado recomendaciones de becas a alumnos en riesgo de desertar, si hay una comunicación continua entre los dueños de sistemas y administradores de Data Mart Alumnos, y si los altos cargos transmiten a terceros (a través de prensa, informativos internos u otros) el valor de la nueva capacidad.

10. **Inicio, hitos, ritos y cierre**

Para posicionar la nueva capacidad en la universidad, se debe socializar el inicio del proyecto, el cumplimiento de hitos y avances, y dar a conocer a la comunidad universitaria los alumnos que han sido beneficiados de esta nueva capacidad. La finalización del proyecto de gestión

del cambio se realizará al término del tercer año cuando se cuente con la nueva capacidad de retener a alumnos con uso de minería de datos estabilizada.

8.2. Caracterización del cambio

El cambio de este proyecto consiste en generar una nueva capacidad que permita formalizar y retener a los alumnos de contextos más vulnerables, para lo cual se generan procesos que permiten capturar información de alumnos para luego procesar dicha información con modelos de minería de datos, para luego en base a los análisis realizados generar ayudas de forma proactiva.

Dentro de la implementación de esta nueva capacidad hay procesos que son modificados. Los procesos de analítica son modificados pasando de ser procesos basados en análisis estadísticos a procesos de minería de datos. El procesamiento de postulantes a beneficios se ve modificado al otorgar recomendaciones de becas al momento de postular según el riesgo de deserción, y de la misma manera, la selección de postulantes complementará el puntaje obtenido por la ficha social con el riesgo de deserción. Por último, la generación de nuevos programas de beneficios estudiantiles y la planificación de asignación de beneficios estudiantiles serán complementados con los hallazgos de los modelos de minería de datos. Como todos estos procesos involucran personas que cambian su forma de operar, se debe gestionar con cuidado la adopción de las nuevas formas de trabajar.

8.3. Factores críticos de éxito

Los factores críticos que deben ser considerados dentro del proyecto de cambio son:

1. Conseguir apoyo de directivos, ya que estos otorgan respaldo político, financiero y tecnológico. Para ello se debe comunicar de forma correcta los beneficios de implementar el uso de minería de datos dentro de la Universidad de Chile.
2. Conseguir que asistentes sociales puedan comprender y adoptar los nuevos procesos. Para ello se les debe comunicar de forma correcta los beneficios, capacitar en los nuevos procesos y entregar mensajes claros sobre qué aspectos se conservarán y cuáles serán modificados.

8.4. Plan de gestión del cambio

Las principales acciones a realizar en este proceso de gestión del cambio son:

- Reuniones periódicas con directivos para alinear los esfuerzos y coordinar las actividades necesarias durante el desarrollo del proyecto.
- Reuniones periódicas con asistentes sociales para recibir retroalimentación de los hallazgos de los modelos de minería de datos.
- Comunicar de forma estratégica los éxitos alcanzados en materia de retención de alumnos y cumplimiento de hitos relevantes, de forma de generar una percepción de avance y logro que motive seguir avanzando con el proyecto.
- Capacitar al personal de bienestar estudiantil en los nuevos procesos.
- Crear actividades y cursos que permitan integrar los conocimientos de las distintas disciplinas involucradas en la generación de programas (tecnologías de información, ingeniería, trabajo social, sociología, entre otras).
- Monitorear el proceso de gestión del cambio con diferentes indicadores.
- Socializar el proyecto mediante un kickoff con autoridades de la Universidad y comunicar el cumplimiento de hitos relevantes.

Capítulo 9

Evaluación del Proyecto

En esta sección se presenta la factibilidad del proyecto presentado en esta tesis, para lo cual se realiza un análisis retrospectivo que permita validar que los modelos desarrollados son generalizables para alumnos que no pertenecen al cohorte en estudio (alumnos de primer año de 2017), para luego detallar los beneficios y costos de la situación con y sin proyecto, el flujo de caja de cada uno de los escenarios y un análisis de sensibilidad.

9.1. Análisis Retrospectivo

En minería de datos es relevante que los modelos desarrollados sean generalizables, es decir, que puedan aplicarse a un set de datos que no se encuentran en el set utilizado para entrenar y testear los modelos. En el caso de la deserción universitaria, corresponde a que los modelos desarrollados puedan ser aplicados a alumnos de primer año de otras generaciones. Es por ello que se utilizó un análisis retrospectivo, el cual consiste en probar los modelos desarrollados con datos históricos de alumnos.

9.1.1. Presentación de Datos

Para realizar el análisis retrospectivo se utilizaron los resultados del modelo seleccionado, el cual corresponde a una Regresión Logística entrenada con datos de alumnos del cohorte 2017 (ver Modelo

seleccionado). Los resultados de este modelo (ver Coeficientes Regresión Logística) fueron aplicados a los datos de alumnos del cohorte 2016 que rindieron la PSU, que corresponden a un total de 6.254 alumnos.

Cabe mencionar que el periodo de deserción institucional considerada para alumnos del cohorte 2016 corresponde a un total de dos años (enero de 2016 a diciembre 2017), que a diferencia del cohorte 2017, fue de 1 año y 4 meses (enero 2017, mayo 2018). El motivo de considerar un periodo más extenso para evaluar si desertaron los alumnos en comparación al cohorte 2017, es que la mayoría de las eliminaciones académicas se formalizan durante el segundo semestre del segundo año de estudios y a que la fecha de extracción de datos para efectos de este proyecto de tesis fue en mayo 2018.

9.1.2. Resultados obtenidos

Los resultados obtenidos para la predicción de la deserción del cohorte 2017 y la aplicación de los resultados del modelo seleccionado sobre los datos del cohorte 2016 se observan en la tabla 9.1.

TABLA 9.1: Resultados análisis Retrospectivo

Cohorte	2016	2017
Alumnos nuevos que rindieron PSU	6.078	6.254
Deserción Institucional	1.161 alumnos (19,1 %)	828 alumnos (13,2 %)
Recall	35 %	41 %
Precision	32 %	26 %
F-Measure	33 %	32 %
Desertores detectados	407 alumnos (6,7 % del total de alumnos)	337 alumnos (5,4 % del total de alumnos)

A partir de los resultados, se puede afirmar que de haber aplicado este modelo sobre alumnos del cohorte 2016 se hubiese detectado la deserción de 407 alumnos y que el modelo, al igual que en el cohorte 2017, mantiene un buen comportamiento frente a errores de tipo I y de tipo II (valores recall, precision y fmeasure similares al cohorte 2017).

En consecuencia, se valida que los modelos utilizados son generalizables a la población de alumnos de primer año que ingresa a la Universidad de Chile. Además, se rescata el alto nivel de predicción de los modelos (más del 1 % del total de alumnos), demostrando la gran utilidad de utilizar modelos de minería de datos para predecir de la deserción.

9.1.3. Validación Análisis Retrospectivo

Una vez obtenido los resultados del análisis retrospectivo, la Dirección de Bienestar Estudiantil validó los resultados. La retroalimentación brindada se detallada en esta sección.

Con respecto a los alumnos que poseen Crédito con Aval del Estado (CAE), el modelo predice menos alumnos desertores con respecto a la realidad (6 %, ver tabla 9.2). Según expertos de la Dirección de Bienestar Estudiantil, este fenómeno se debe a que las personas que poseen CAE generalmente corresponde a personas que no calificaron para obtener una beca, los cuales en su mayoría corresponde a personas de deciles más altos y con un Índice de Vulnerabilidad Escolar (IVE) menor. Dado que dentro del modelo estos perfiles poseen un menor riesgo de desertar, baja el riesgo de deserción asociado, y por ende, se detectan menos desertores.

TABLA 9.2: Análisis Retrospectivo - CAE. Fuente: Elaboración propia.

Crédito con Aval del Estado (CAE)	Total alumnos	Desertores Reales	% Deserción Real	Predicción Deserción	% Predicción Deserción	Predicción - Reales desertores	Variación
Sin CAE	5.465	1.063	19%	1.216	22%	153	3%
Con CAE	613	98	16%	60	10%	-38	-6%
Total	6.078	1.161	19%	1.276	21%	115	2%

En cuanto a IVE, la categoría IVE Rango 1, que corresponde a alumnos que provienen de colegios donde el 76,7 % y 100 % de sus alumnos provienen de situación socioeconómica de pobreza o extrema pobreza, posee la más alta variación entre la predicción y la realidad de desertores (10 % más, ver tabla 9.3). Si bien la realidad muestra que la deserción de alumnos provenientes de IVE Rango 1 es mayor (31 % en alumnos nuevos de 2016, ver tabla 9.3), el modelo genera un sobreestimación de la deserción de estos alumnos. Ahora bien, la variación del 10 % corresponde sólo a 36 alumnos, un número menor en comparación al total de alumnos del cohorte 2016 (6.078 alumnos).

TABLA 9.3: Análisis Retrospectivo - Índice de Vulnerabilidad Escolar (IVE). Fuente: Elaboración propia.

IVE	Total alumnos	Desertores Reales	% Deserción Real	Predicción Deserción	% Predicción Deserción	Predicción - Reales desertores	Variación
IVE Rango 1 (entre 76,7% y 100% de vulnerabilidad)	378	117	31%	153	40%	36	10%
IVE Rango 2 (entre 53,4% y 76,6% de vulnerabilidad)	1.420	287	20%	393	28%	106	7%
IVE Rango 3 (entre 30% y 53,3% de vulnerabilidad)	1.987	359	18%	298	15%	-61	-3%
SIN Rango IVE (sin calificación de vulnerabilidad)	2.293	398	17%	432	19%	34	1%
Total	6.078	1.161	19%	1.276	21%	115	2%

En cuanto a la carrera de los alumnos, en Programa Académico de Bachillerato, Diseño y Filosofía el modelo estima más desertores de lo que en realidad son (213, 88 y 42 más respectivamente, ver figura 9.1), mientras que en Ingeniería Comercial e Ingeniería y Ciencias Plan Común el modelo estima menos desertores de lo que en realidad son (54 y 73 alumnos menos respectivamente, ver figura 9.1) ¹. Estos resultados indican que la deserción opera de diferentes maneras según las diferentes carreras. Es por ello, que en el caso de Bachillerato, Diseño y Filosofía, se podrían incorporar nuevas variables para mejorar la precisión de los modelos, como por ejemplo, variables relacionadas a vocación profesional. En cuanto Ingeniería comercial e Ingeniería y Ciencias Plan Común, al poseer una menor cantidad de alumnos definidos como prioritarios por la universidad, se podrían incorporar otros datos que pudieran reflejar la deserción, como rendimiento académico o vocación profesional.

¹Para ver más detalles, ver anexo Análisis Retrospectivo - Carrera

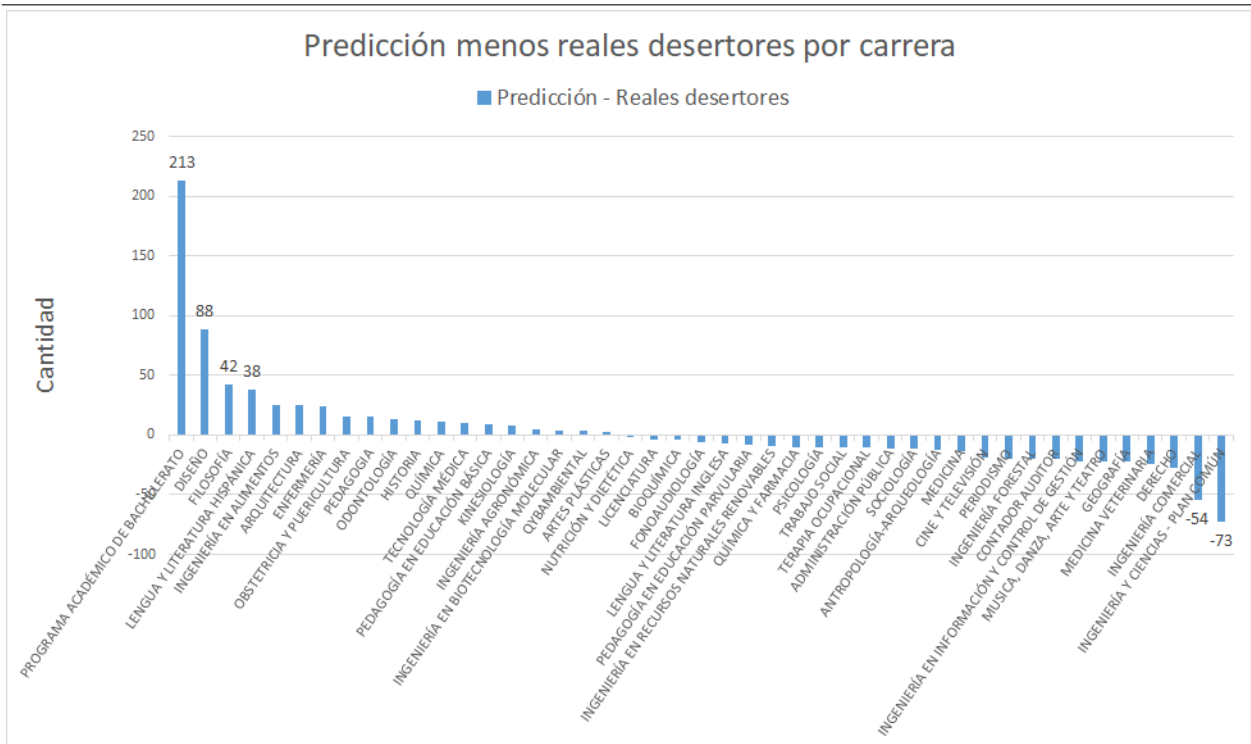


FIGURA 9.1: Desvío absoluto predicción vs realidad. Fuente: Elaboración propia.

Con respecto a los alumnos que poseen gratuidad, el modelo refleja un buen comportamiento, ya que se mantienen sólo con una pequeña variación con respecto a la realidad (1%, ver anexo C.1). De la misma manera ocurre con los alumnos alumnos que poseen Beca Vocación Profesor (variación 1%, ver anexo C.2) y con la variable Tramo Socioeconómico, la cual posee una variación promedio de 2% con respecto a la realidad (ver anexo C.3)

La Dirección de Bienestar concluye que los resultados son adecuados a la realidad de la Universidad de Chile, con la salvedad de que se debe considerar que existen particularidades dentro de las distintas carreras, como por ejemplo, que algunas carreras concentran determinados perfiles socioeconómicos, lo cual es importante considerar al momento aplicar acciones para poder retener a los alumnos.

9.2. Definición de Beneficios y Costos

Para evaluar el proyecto se consideraron dos escenarios: escenario sin proyecto (situación actual optimizada) y situación con proyecto (desarrollo de nueva capacidad).

1. Escenario sin proyecto

Este escenario corresponde a la situación actual optimizada, y consiste en utilizar nuevos modelos estadísticos y variables para comprobar hipótesis de la deserción de alumnos, utilizando los mismos procesos de extracción y generación de información.

La estructura de beneficios del escenario sin proyecto, corresponde a disminuir la tasa de retención institucional en un 8% o de forma equivalente a evitar la deserción de un 1% del total de alumnos, pasando de una tasa de deserción de 13,2% a 12,2%. Alcanzando estos objetivos, se generan ahorros económicos equivalentes a \$245 millones anuales (ver sección Cuantificación del Problema u Oportunidad).

Bajo este escenario se incurre en costos de un *Analytics manager*, el cual se encuentra actualmente a cargo de realizar análisis estadísticos en la DBE, y un encargado TI, a cargo de extraer los datos de alumnos de las bases de datos que actualmente administra la DBE. Esto tiene costos de \$1,3 millones por persona.

2. Escenario con proyecto

Corresponde al escenario con la implementación de la nueva capacidad para retener a los alumnos en la Universidad de Chile.

La estructura de beneficios de este escenario se calcula en base a los resultados obtenidos con el cohorte 2017 con el modelo de regresión logística (ver sección Modelo seleccionado), el cual detecta a 337 alumnos desertores.

Los beneficios proyectados se calculan como la disminución de la deserción gradual, alcanzando al tercer año la disminución efectiva de la deserción de los 337 alumnos, generando ingresos por \$1.325 millones (ver sección 5.5).

Dado el supuesto de la gradualidad en la obtención de beneficios, se obtendría una reducción de la deserción en el primer año de 1/3 del total de alumnos detectados, 2/3 para el segundo año y el total detectado por los modelos al tercer año (evitar la deserción de 337 alumnos, o de forma equivalente, disminuir la deserción universitaria en un 41%).

En tanto que a costos se considera:

- Director de Minería de Datos.

Persona capaz de posicionar a nivel directivo la relevancia de desarrollar esta nueva capacidad en la universidad y poseer a la vez amplios conocimientos en herramientas de minería de datos, sus aplicaciones y de integración de sistemas.

- *Data Scientist.*

Persona orientada a resolver problemas de negocio con datos, con amplios conocimientos en estadísticas y en construir algoritmos de minería de datos, capaces de encontrar patrones ocultos en modelos tanto supervisados como no supervisados ².

- *Analytics Manager*

Cargo que se encuentra entre le *Data Scientist* y el Encargado TI. Debe poseer amplios conocimientos de negocio, proveyendo de esta forma dirección a los análisis, además de aportar con sus propios análisis predictivos y de clusterización. También es el enlace con el Encargado TI, entregando soporte a la administración del Data Mart de Alumnos y procesos ETL ³.

- Data Engineer

Costos asociados a persona a cargo de mantener Data Mart, para lo cual deberá habilitar el flujo de datos desde sistemas internos y fuentes externas a través de procesos ETL. Este además deberá ir incorporando nuevas datos relacionados al comportamiento de alumnos según los requerimientos del *Data Scientist* y del *Analytics Manager*. ⁴.

Considerando los ingresos obtenidos por personas a cargo de proyectos de minería de datos y los salarios de la universidad, se considera un costo de \$3 millones de pesos para un director de minería de datos, \$2 millones para el Data Scientist y \$1,6 millones para *Data Scientist* y *Analytics Manager*.

9.3. Flujo de Caja

Se detalla en esta sección los consideraciones utilizadas para realizar el cálculo de flujo de caja y se analizan 3 indicadores de evaluación de proyectos: Valor Actual Neto (VAN), Tasa Interna de Retorno (TIR) y Periodo de Recuperación del Capital (PRC).

²<https://towardsdatascience.com/how-to-structure-a-high-performance-analytics-team-f564c92a1aaa>

³<https://towardsdatascience.com/how-to-structure-a-high-performance-analytics-team-f564c92a1aaa>

⁴<https://towardsdatascience.com/how-to-structure-a-high-performance-analytics-team-f564c92a1aaa>

■ Tipo de evaluación y Tasa de Descuento

El tipo de evaluación corresponde a una evaluación privada, debido a que no se consideran los beneficios sociales de reducir la deserción debido a la complejidad de su cálculo. De todas maneras, como se trata de una institución estatal y los beneficios de este proyecto repercuten en la sociedad, se utilizará la tasa de descuento del Ministerio de Desarrollo Social, la cual corresponde para el año 2018 al 6 % real anual ⁵.

■ Horizonte de Evaluación

Debido al acelerado avance de las tecnologías y reelección de rector, se utilizará una evaluación de proyecto de 3 años.

■ Inversión

Se invierte en dos notebooks Dell Intel Core i7 16 GB de ram para trabajar los algoritmos de minería de datos, por un costo de \$ 907.000 por cada uno.

Por concepto de remuneraciones, este proyecto de tesis no tuvo costos asociados, ya que se trabajo con personal ya contratado por la universidad para realizar las actividades y reuniones.

Se utilizarán herramienta de minería de datos de código abierto durante los primeros 3 años de proyecto (costo 0).

Se invertirá en un Ingeniero Civil en Computación para armar el Data Mart de Alumnos por 6 meses por un costo de 3 millones mensuales. Este además deberá documentar todo el proceso de extracción, transformación y carga de datos (ETL) para poder mantener el Data Mart a posterioridad.

También se invertirá en una consultora de gestión del cambio. Se consideran dos etapas:

1. La primera etapa estará enfocada en preparar a los directivos de la universidad para la adopción de una cultura orientada al manejo de datos y unificación de esfuerzos entre el Departamento de Pregrado, Dirección de Tecnologías de Información y Dirección de Bienestar Estudiantil, además de contar con el apoyo financiero de la Vicerrectoría de Asuntos Económicos. Contratación entre los meses 3 y 6 del proyecto por un costo de 3 millones mensuales.

⁵ <http://sni.ministeriodesarrollosocial.gob.cl/download/precios-sociales-vigentes-2017/?wpdmdl=2392>

2. La segunda etapa consistirá en planificar las capacitaciones y mensajes que serán dados a asistentes sociales, de forma que el valor de la solución sea comprendido y adoptado para su permanencia. Contratación entre los meses 6 y 9 con por un costo de 3 millones mensuales.

■ **Reajuste**

Se considera un reajuste anual de 2,5% ⁶ en el pago de remuneraciones.

■ **Impuestos**

No se considera impuestos, debido a que al tratarse de un institución pública, los ingresos por concepto de arancel no poseen impuestos asociados.

■ **Depreciación**

No aplica la depreciación del software, ya que se utiliza software gratuito durante los tres años del proyecto.

Flujo de caja escenario sin proyecto

El flujo de caja sin proyecto con situación actual optimizada genera ingresos de \$245 millones por cada año al disminuir en 8% la tasa de deserción (evitar que un 1% del total de alumnos deserte) y no posee costos adicionales. De esta manera se pasaría de una tasa de deserción de 13,2% a 12,2%. En este escenario se obtiene un VAN de \$569 millones, mejorando de esta manera la situación actual.

TABLA 9.4: Flujo de Caja - Escenario sin Proyecto

Escenario sin proyecto	Año 0	Año 1	Año 2	Año 3
Ingresos		\$ 245.000.000	\$ 245.000.000	\$ 245.000.000
Egresos		\$ 31.200.000	\$ 31.980.000	\$ 32.779.500
Utilidad		\$ 213.800.000	\$ 213.020.000	\$ 212.220.500
Inversión	\$ -			
Flujos finales		\$ 201.698.113	\$ 189.587.042	\$ 178.184.424

⁶www.anaff.cl/documentos/category/17-2017?download=32:tabla-reajuste-ano-2018

Flujo de caja escenario con proyecto

El flujo de caja con proyecto genera ingresos de \$441, \$883 y \$1.325 millones los años 1, 2 y 3 respectivamente, debido a que se utiliza el supuesto de que la disminución de la deserción será gradual alcanzando la disminución pronosticada por los modelos (41%) al tercer año. Además, posee costos por concepto de remuneraciones por \$5 millones mensuales más un reajuste de sector público (\$98,4 millones anuales más reajustes) y una inversión de \$37,8 millones (2 notebook, un ingeniero en computación por 6 meses y asesoría gestión del cambio en dos etapas de 3 meses cada una). A partir de este se obtiene un VAN de \$2.008 millones, TIR de 955% y un PRC de sólo un año.

TABLA 9.5: Flujo de Caja - Escenario con Proyecto

Escenario con proyecto	Año 0	Año 1	Año 2	Año 3
Ingresos		\$ 441.666.667	\$ 883.333.333	\$1.325.000.000
Egresos		\$ 98.400.000	\$ 100.860.000	\$ 103.381.500
Utilidad		\$ 343.266.667	\$ 782.473.333	\$1.221.618.500
Inversión	\$ 37.815.845			
Flujos finales	\$ -37.815.845	\$ 323.836.478	\$ 696.398.481	\$1.025.694.449

Para comparar ambos escenarios, resulta útil utilizar el ΔVAN , el cual permite comparar el VAN situación actual optimizada con el VAN del nuevo proyecto. De este indicador se obtiene un $\Delta VAN = \$1.520$ millones, es decir, que la situación con proyecto es mejor por \$1.520 millones con respecto a la situación actual optimizada, sin considerar los beneficios sociales, lo cual generaría retornos aún mayores. Además, la inversión se recupera el primer año de puesta en marcha del proyecto (PRC=1). Dado estos antecedentes, se puede afirmar que la situación con proyecto es la que debe ser realizada.

9.4. Análisis de Sensibilidad

Un aspecto clave para que el proyecto sea rentable es que este permita disminuir la deserción universitaria. Es por ello que se escogió el porcentaje de disminución de la tasa de deserción como la variable relevante a estudiar en el análisis de sensibilidad. Los resultados obtenidos pueden ser observado en el anexo Análisis de Sensibilidad.

De estos resultados se desprende que si la tasa de deserción disminuye en un 5% el proyecto es rentable ($VAN = \$52$ millones). Además, la recuperación del capital es obtenida durante el primer año para todos los escenarios sobre una disminución de la tasa de deserción por sobre 13%. Considerando el escenario actual optimizado, el proyecto se justifica cuando se logra disminuir la tasa de deserción en un 15% o de forma equivalente evite la deserción de un 2,0% del total de alumnos, es decir, que la tasa de deserción pase de un 13,2% a un 11,2%, ya que en este punto el ΔVAN comienza a ser positivo. Considerando que el modelo de minería de datos seleccionado indica que la tasa de deserción se disminuye en un 41%, el proyecto debe ser llevado a cabo.

Capítulo 10

Conclusiones

La Universidad de Chile se rige bajo la Política de Equidad e Inclusión promulgada en 2014, la cual garantiza el acompañamiento de estudiantes en el acceso, permanencia y egreso oportuno de sus alumnos. Parte importante de esta política corresponde a que la Universidad de Chile debe realizar esfuerzos por retener a sus alumnos, respondiendo a un cuerpo estudiantil cada vez más diverso en términos económicos, sociales y culturales. Pese a los esfuerzos, aún continúa existiendo una marcada heterogeneidad en la retención de alumnos de primer año en sus distintas facultades, siendo la más alta en la Facultad de Ciencias Físicas y Matemáticas, con un 95 % de retención, y la más baja en Bachillerato, con sólo un 53 % de retención.

Dado este problema y al compromiso institucional que posee la Universidad de Chile con respecto al acompañamiento de sus alumnos durante su trayectoria universitaria, se propone dentro de este proyecto complementar la forma actual de asignación de beneficios estudiantiles con el uso de modelos de minería de datos, con el objetivo de detectar alumnos en riesgo de deserción para así ofrecer beneficios estudiantiles de forma proactiva.

Para poder implementar esta nueva capacidad se realizó un estudio de la estrategia de la Universidad de Chile, su arquitectura de procesos, procesos de negocio y capacidades tecnológicas actuales, para luego definir los elementos que deben ser modificados para materializar la nueva capacidad.

En cuanto a Arquitectura de Procesos, la Universidad de Chile posee un esquema mixto para generar nuevas capacidades, ya que esto le permite generar nuevas capacidades que benefician de forma

transversal a la universidad, lo cual es impulsado por las diferentes vicerreorías, y también generar las propias capacidades según las necesidades de negocio específicas de cada facultad. En este sentido, se indica dentro de este proyecto de tesis que la estructura actual de la Dirección de Bienestar Estudiantil (DBE) debe ser mantenida ya que permite concentrar el conocimiento en temáticas de bienestar de alumnos y administrar de forma eficiente los recursos universitarios. Esto es positivo ya que permite otorgar programas de ayuda a todos los alumnos de la universidad, de forma indistinta a la capacidades técnicas y financieras de las distintas facultades.

Lo que se propone dentro de esta estructura es contar con una política universitaria de gobierno de datos y administración de sistemas, en especial de aquellos sistemas que son relevantes para el quehacer de la Universidad, como lo son sistemas relacionados a alumnos e investigación, para así facilitar los procesos de analítica dentro de la universidad. Esto es respaldado por la literatura, donde se indica que para tener éxito en la implementación de analítica en las organizaciones es necesario contar con un “liderazgo, herramientas y tecnologías en común” [11]. De esta manera se evita la creación de múltiples sistemas, datos e indicadores, que generalmente son utilizados sólo por el área o personal que los administra.

Los procesos de analítica son modificados pasando de ser procesos basados en análisis estadísticos a procesos de minería de datos, poniendo en práctica la metodología de minería de datos CRISP-DM. Esta metodología contiene 6 fases, las que corresponden a comprensión de negocio, comprensión de datos, preparación de datos, modelado, evaluación y puesta en marcha. Aplicando este proceso sobre el cohorte de alumnos nuevos del año 2017 que rindieron la PSU, se obtuvo que de haber sido aplicados los modelos durante dicho año, se hubiese detectado la deserción de 337 alumnos, es decir, el 5,4 % del total de alumnos del cohorte 2017, demostrando el potencial del uso de minería de datos para predecir la deserción. Además se demostró que el modelo seleccionado es generalizable, es decir, que de ser aplicados los resultados de los modelos de minería de datos a nuevas generaciones de alumnos que ingresan a la universidad, los modelos tendrían un rendimiento similar en la detección de la deserción, lo cual fue validado por la Dirección de Bienestar Estudiantil. De esta manera se afirma que el criterio de éxito de minería de datos definido al inicio del proyecto es alcanzado, superando las expectativas iniciales (haber detectado la deserción de 1 % del total de alumnos del cohorte de alumnos nuevos del 2017, mientras que lo alcanzado fue un 4,4 %).

Además de lo anterior, se modificaron cuatro procesos relevantes de la Dirección de Bienestar Estudiantil. En “Generación de nuevos programas de beneficios estudiantiles” y “Planificación de asignación de beneficios”, se incorporaron los resultados de las predicciones y segmentación de alumnos para crear nuevos programas y planificar la entrega de beneficios. En “Procesamiento de postulantes” se incorporó la entrega de recomendaciones de becas y programas de ayuda a estudiantes de forma proactiva. Y finalmente, “Selección de postulantes”, se complementa la decisión de asignar beneficios incorporando el riesgo de desertar de los alumnos.

Desde el punto de vista tecnológico, se creó una arquitectura que consta de un Software de Minería de Datos, Herramientas de Visualización y un Data Mart de Alumnos. El primero, es utilizado principalmente para desarrollar los modelos de minería de datos, el segundo, para realizar análisis exploratorios de los datos, y el tercero, para mantener los datos de alumnos bajo un único estándar de calidad. Los motivos de implementar un Data Mart corresponden a que en la universidad existen diferentes sistemas de alumnos no integrados y a que contar con datos bajo un único estándar de calidad permitiría que los analistas puedan destinar su tiempo a descubrir patrones en los datos en vez de dedicar tiempo a realizar actividades de recopilación, unificación y limpieza de bases de datos.

Para que las personas involucradas en el proyecto puedan adoptar los nuevos procesos es necesario desarrollar plan detallado de gestión del cambio, a modo de evitar resistencias de directivos y asistentes sociales a las nuevas estrategias y procesos generadas para disminuir la deserción estudiantil.

El disminuir la tasa de deserción genera importantes ingresos a la universidad, los cuales justifican la realización del proyecto, con un ΔVAN de \$ 1.520 millones con respecto a la situación actual optimizada y con un periodo de recuperación del capital invertido de un año. Si se considerase los costos sociales de un alumno desertor, el valor del proyecto serían aún mayor.

10.1. Lecciones Aprendidas

A través de la realización de este proyecto de tesis se aprendió que, para desarrollar una nueva capacidad en una organización esta debe responder en primer lugar a necesidades estratégicas. Una vez definida la capacidad que se desea desarrollar, debe ser emplazada en una estructura organizacional adecuada para su funcionamiento, lo cual involucra el desarrollo de nuevos procesos o la mejora de ellos; generar modelos que permitan tomar decisiones en base a datos y desarrollar las componentes

tecnológicas que permitan soportar la nueva capacidad. Todo esto debe ir acompañado de un plan de gestión del cambio para que la nueva capacidad sea adoptada por los diferentes actores involucrados, además de ser económica o socialmente rentable.

Existen patrones de implementación de nuevas capacidades en las organizaciones, lo cual es muy práctico para diseñar innovaciones tecnológicas en instituciones o empresas, ya que otorgan una pauta de como diseñar la estructura organizacional, qué nuevos procesos implementar o mejorar, y qué tecnologías permitirán alcanzar los resultados deseados.

La metodología de minería de datos CRISP-DM facilita mantener una estructura para analizar los datos, involucrando dentro de sus fases todo lo relevante para resolver problemas de la organización, desde identificar el problema de negocio que se desea resolver, extraer y tratar los datos, correr los modelos, evaluar los resultados y la puesta en producción de los modelos aprobados.

El contar con múltiples fuentes de datos y en planillas excel hace que sea imposible o muy complejo desarrollar modelos de minería de datos, ya que se tendría que estar constantemente ajustando y depurando los datos, solicitando información, además de incrementar la posibilidad de contar con datos errados y desactualizados. El Data Mart es una herramienta de almacenamiento y tratamiento de datos muy útil en este sentido, ya que permite almacenar información de múltiples fuentes de datos bajo un estándar de calidad y formato en común. Para que esto se mantenga en el largo plazo, se requiere contar con un gobierno de datos y de sistemas que permita regular la modificación e implementación de nuevos sistemas y definir estándares para el tratamiento y disponibilidad de datos.

La creación de proyectos de minería de datos debe ser una capacidad que quede instalada en la universidad, para estar continuamente agregando nuevas variables, realizando nuevos análisis e ir probando nuevos modelos para seguir obteniendo mejores resultados.

10.2. Trabajos Futuros

Como trabajo futuro se encuentra la puesta en marcha del proyecto, lo cual queda a cargo de la Dirección de Bienestar Estudiantil.

Además, se debe evaluar implementar la nueva capacidad con proveedores externos, para lo cual se

debe realizar un *benchmark* de proveedores de minería de datos, experiencia de otras universidades, factibilidad técnica, evaluación económica, entre otros.

Se propone para trabajos futuros desarrollar soluciones para mejorar el rendimiento de los modelos en las diferentes carreras, especialmente en aquellas con una mayor variación en la predicción (Bachillerato, Diseño, Filosofía, Ingeniería Comercial e Ingeniería y Ciencias Plan Común), para que así puedan responder a las necesidades específicas de sus alumnos. También, se propone incorporar nueva información para mejorar el rendimiento de los modelos.

Se propone utilizar además de los modelos considerados en este estudio, otros modelos de minería de datos para predecir la deserción y segmentar alumnos.

También, se deja como propuesto extender el uso de minería de datos para predecir resultados académicos de los alumnos, permitiendo de esta manera focalizar tutorías y programas de nivelación.

Bibliografía

- [1] Oscar Barros. *Business engineering and service design*. Service Systems and Innovations in Business and Society Collection. New York: Business Expert Press, 2nd edition edition, 2017.
- [2] A. C. Hax and D. L. Wilde. *El Proyecto Delta*. Editorial Norma, 2003.
- [3] Daniel T. Larose. *Data Mining Methods and Models*. John Wiley & Sons, 2006.
- [4] D. Pelleg and A. Moore. X-means: Extending k-means with efficient estimation of the number of clusters. 2000.
- [5] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
- [6] Erika Himmel. Modelo de análisis de la deserción estudiantil en la educación superior. *Calidad en la Educación*, 2002.
- [7] S. Celis, L. Moreno, P. Poblete, J. Villanueva, and R. Weber. Un modelo analítico para la predicción del rendimiento académico de estudiantes de ingeniería. 2015.
- [8] R. Díaz. El impacto del tipo de financiamiento sobre la probabilidad de retención de primer año en la educación superior: el caso de la universidad de chile. 2017.
- [9] M. Alarcón. Impacto de los distintos tipos de ayuda financiera gubernamental e institucional en la persistencia de los estudiantes de la universidad de chile. santiago. 2015.
- [10] V. Santelices, X. Catalán, C. Horn, and D. Kruger. Determinantes de deserción en la educación superior chilena, con énfasis en efecto de becas y créditos. 2013.
- [11] Thomas H. Davenport. *Competing on analytics*. 2006.

-
- [12] Rodolfo Schmal, Reinaldo Ruiz, Sebastián Donoso, and Martin Schaffernicht. Factores que inciden en el financiamiento de los estudios universitarios en Chile. 2007.
- [13] Oscar Barros. *Ingeniería de Negocios. Diseño Integrado de Negocios, Aplicaciones y Procesos TI*. 2009.
- [14] E. Macaya, B. Crawford, and R. Soto. Gestión del cambio para proyectos tecnológicos: Usando un modelo integral de gestión del cambio. 2016.

Anexo A

Marco Teórico

A.1. Parámetros árboles de decisión

TABLA A.1: Parámetros árboles de decisión

Parámetro	Descripción	ID3	CART	CHAID
Minimal size for split	Sólo pueden dividirse los nodos donde el data set es mayor o igual al parámetro minimal size for split.	1	1	1
Minimal leaf size	El árbol posee hojas poseen un cantidad de registros mayor o igual al parámetro minimal leaf size.	1	1	1
Minimal gain	Corresponde a la ganancia mínima que debe aportar un split de un nodo. A menor valor de minimal gain, más splits se realizan. A mayor valor, menos splits se realizarán, hasta quedar todos los registros en un único nodo.	1	1	1
Maximal Depth	Este parámetro restringe la cantidad de ramificaciones del árbol.	0	1	1
Confidence	Nivel de significancia mínimo para el split (p-valor mínimo)	0	0	1

A.2. Parámetros Random Forest

TABLA A.2: Parámetros Random Forest

Parámetros	Definición
Number of trees	Especifica la cantidad de árboles que son generados de forma aleatoria. Un mayor número árboles incrementa el performance del modelo y genera predicciones más estables, al precio de una mayor exigencia computacional.
Criterion	Information Gain Gini Index Gain Ratio Accuracy Least square
Maximal Depth	Este parámetro restringe la cantidad de ramificaciones del árbol.
Minimal size for split	Sólo pueden dividirse los nodos donde el data set es mayor o igual al parámetro minimal size for split.
Minimal leaf size	El árbol posee hojas poseen un cantidad de registros mayor o igual al parámetro minimal leaf size.
Minimal gain	Corresponde a la ganancia mínima que debe aportar un split de un nodo. A menor valor de minimal gain, más splits se realizan. A mayor valor, menos splits se realizarán, hasta quedar todos los registros en un único nodo.
Confidence	Nivel de significancia mínimo para el split (p-valor mínimo)
Voting Strategy	Confidence Vote: Selecciona la clase que tiene el mayor nivel de confianza acumulada. Majority Vote: Selecciona la clase que fue predecida por la mayoría de los modelos de árboles de decisión
Guess subset ratio	Si se activa se utilizan $\text{int}(\log(m) + 1)$ atributos. De lo contrario debe definirse un valor entre 0 y 1 (subset ratio).

Anexo B

Lógica de Negocio

B.1. Detalle atributos Guía Matrícula e Índice de Vulnerabilidad Escolar (IVE)

TABLA B.1: Atributos Guía Matrícula e Índice de Vulnerabilidad Escolar (IVE)

Inicio de Tabla B.1				
Nombre	Tipo	Valores	Missing Values	Fuente
Situación Académica (Años 2017 y 2018)	Categorica	1. Matriculado 2. Renuncia a la Carrera 3. Eliminación Académica 4. Postergación 2do semestre 6. Postergación Anual 7. Postergación por Salud 8. Postergación 1er semestre	0	Guía Matrícula

Continuación de Tabla B.1				
Nombre	Tipo	Valores	Missing Values	Fuente
Región	Categorica	1. RM (5.012) 2. O'Higgins (313) 3. Valparaíso (192) 4. Maule (147) 5. Biobío(139) 6. Coquimbo (105) 7. Los Lagos (80) 8. Antofagasta (60) 9. Araucanía (57) 10. Atacama (30) 11. Tarapacá (30) 12. Aysén (27) 13. Magallanes (22) 14. Arica y Parinacota (20) 15. Los Ríos (20)	0	Guía M.
Sexo	Binaria	1 = Mujeres (3.249) 0 = Hombres (3.005)	0	Guía M.
Nacionalidad	Binaria	1 = Chilena (6.188) 0 = Extranjera (66)	0	Guía M.

Continuación de Tabla B.1				
Nombre	Tipo	Valores	Missing Values	Fuente
Carrera	Catagórica	<p>63 carreras de pregrado.</p> <p>- Carreras con más alumnos:</p> <ol style="list-style-type: none"> 1. Ingeniería Plan Común (795) 2. Ingeniería Comercial (439) 3. Derecho (389) 4. Bachillerato (373) 5. Arquitectura (246) <p>- Carreras con menos alumnos:</p> <ol style="list-style-type: none"> 1. Interpretación musical (16) 2. Danza (6) 3. Etapa Básica Interpretación musical (5) 4. Composición (4) 5. Etapa Básica Composición (2) 	0	Guía M.
Procedencia Educativa	Catagórica	<ol style="list-style-type: none"> 1. Municipal (1.775) 2. Particular (1.912) 3. Subvencionado (2.582) 	39	Guía M.

Continuación de Tabla B.1				
Nombre	Tipo	Valores	Missing Values	Fuente
Vía Ingreso	Categórica	1. PSU (5.447) 2. SIPEE (313) 3. BEA (209) 4. Deportista Destacado (69) 5. Otros Ingresos Especiales (60) 6. PACE (45) 7. Equidad de Género (40) 8. Ciclo Básico Artes (36) 9. Escuela de Desarrollo de Talentos (15) Otras	0	Guía M.
Tramo Socioeconómico	Categórica	1. Tramo 40 (1.886) 2. Tramo 50 (716) 3. Tramo 60 (329) 4. Tramo 70 (371) 5. Tramo 80 (263) 6. Tramo 90 (813) 7. Tramo 100 (496) 8. Sin calificación socioeconómica (1.380)	0	Guía M.
Índice de Vulnerabilidad Escolar (IVE)	Real	Entre 13 % y 95 %	2.005	JUNAEB
Enfermedad Catastrófica	Binaria	1= Si (189) 2= No (1.941)	4.124	Ficha Social

Continuación de Tabla B.1				
Nombre	Tipo	Valores	Missing Values	Fuente
Enfermedad Permanente	Categorica	Asma Bronquial (34) Artritis Crónica y Artrosis (2) Bronquitis crónica (1) Cardiopatía Coronaria (1) Diabetes (5)	6.211	Ficha Social
Enfermedad Crónica	Binaria	1= Si (569) 2= No (1561)	4.124	Ficha Social
Número de dormitorios	Numérica	1 (39) 2 (459) 3 (947) 4 (248) 5 (49) 6 (8) 7 (2) 9 (1)	4.501	Ficha Social
Número de integrantes del hogar	Numérica	1 (26) 2 (321) 3 (704) 4 (639) 5 (8.302) 6 (96) 7 (26) 8 (5) 9 (7) 10 (2) 11 (1) 14 (1)	4.124	Ficha Social

Continuación de Tabla B.1				
Nombre	Tipo	Valores	Missing Values	Fuente
Nivel de educación jefe de hogar	categórica	19 categorías	4.210	Ficha Social
Nivel de educación madre	categórica	20 categorías	4.411	Ficha Social

B.2. Créditos Universitarios

TABLA B.2: Créditos Universitarios

Nombre	Tipo	Missing Values	Fuente	Beneficiados	Monto total (mm CLP)
Crédito con Aval del Estado (CAE)	Binaria	0	MINEDUC	924	\$ 2.478
Fondo Solidario de Crédito Universitario (FS-CU)	Binaria	0	MINEDUC	287	\$ 375

B.3. Becas de Arancel Internas y Externas

TABLA B.3: Becas de Arancel Internas y Externas

Nombre	Tipo	Missing Values	Fuente	Beneficiados	Monto total (mm CLP)
B. U. de Chile	Binaria	0	U. de Chile	90	\$ 295
B. Andrés Bello	Binaria	0	U. de Chile	10	\$ 47
B. Bicentenario	Binaria	0	MINEDUC	674	\$ 2.097
B. para Estudiantes Hijos de Profesionales de la Educación	Binaria	0	MINEDUC	38	\$ 19
B. Vocación del Profesor	Binaria	0	MINEDUC	107	MV
B. Puntaje PSU	Binaria	0	MINEDUC	1	\$ 1
B. Excelencia Académica (BEA)	Binaria	0	MINEDUC	91	\$ 104
B. de Reparación Valech	Binaria	0	MINEDUC	22	\$ 68
Gratuidad	Binaria	0	MINEDUC	2.555	\$ 9.210
B. Las Condes	Binaria	0	Munic.de Las Condes	8	\$ 5

B.4. Becas de Mantención Internas y Externas

TABLA B.4: Becas de Mantención Internas y Externas

Nombre	Tipo	Valores Faltantes	Fuente	Beneficiados	Monto total (mm CLP)
B. de Apoyo Preescolar	Binaria	0	U. de Chile	8	\$ 3,2
B. Emergencia	Binaria	0	U. de Chile	39	\$ 2,5
B. Compensación Laboral	Binaria	0	U. de Chile	1	\$ 0,092
B. Residencia Interna	Binaria	0	U. de Chile	8	\$ 3
B. de Residencia SIPEE	Binaria	0	U. de Chile	15	\$ 23
Hogares Universitarios	Binaria	0	U. de Chile	26	\$ 45
B. de Atención Económica (BAE)	Binaria	0	U. de Chile	2.576	\$ 429
B. de Alimentación para la Educación Superior (BAES)	Binaria	0	JUNAEB	2.951	\$ 853
B. de Mantención Educación Superior (BMES)	Binaria	0	JUNAEB	83	\$ 12,3
B. de Mantención Vocación del Profesor	Binaria	0	JUNAEB	14	\$ 11
B. Presidente de la República (BPR)	Binaria	0	JUNAEB	63	\$ 305
B. Indígena	Binaria	0	JUNAEB	42	\$ 19
B. Residencia Indígena	Binaria	0	JUNAEB	3	\$ 3
B. de Integración Territorial	Binaria	0	JUNAEB	11	\$19
B. Patagonia Aysén	Binaria	0	JUNAEB	26	\$ 45
B. Irma Salas	Binaria	0	JUNAEB	5	\$ 1

B.5. Resultados proceso PSU

TABLA B.5: Resultados proceso PSU

Nombre	Tipo	Valores	Missing Values	Fuente
Puntaje Matemáticas Actual	real	264 a 850	441	DEMRE
Puntaje Matemáticas Anterior	real	294 a 835	4.824	DEMRE
Puntaje Lenguaje Actual	real	201 a 850	441	DEMRE
Puntaje Lenguaje Anterior	real	409 a 824	4.824	DEMRE
Puntaje Ciencias Actual	real	262 a 850	2.193	DEMRE
Puntaje Ciencias Anterior	real	407 a 835	5,217	DEMRE
Puntaje Historia Actual	real	292 a 850	3.009	DEMRE
Puntaje Historia Anterior	real	213 a 836	5.515	DEMRE
Puntaje Ranking	real	399 a 850	0	DEMRE
Puntaje NEM	real	0 a 826	0	DEMRE

B.6. Creación de Nuevas Variables

■ Deserción

Todos aquellos alumnos que se matriculan en 2017 y no se matriculan en 2018 antes del cierre de recepción de matrículas (29 de abril) corresponde a desertores institucionales.

Aquellos alumnos que durante el año 2018 postergan sus estudios en cualquiera de sus causales (postergación anual, postergación primer semestre, postergación por salud y renuncia a la carrera) corresponden a desertores institucionales.

■ Enfermedad Permanente o Crónica

La ficha social de la universidad otorga puntaje en la ficha social a personas con enfermedades crónica o permanentes. Corresponden a estas enfermedades aquellas enfermedades que no

pueden curarse o eliminarse totalmente con un tratamiento y que comprometan los ingresos familiares, pero que no impliquen un riesgo vital.

Esta variable se construyó a partir de las variables enfermedad Permanente y Enfermedad Crónica. El método es simple: Si tiene enfermedad, valor=1. De lo contrario=0.

TABLA B.6: Nueva variable “Enfermedad Permanente o Crónica”

Nombre	Tipo	Valores	Missing Values	Fuente
Enfermedad Permanente o Crónica	Binaria	1 = Posee enfermedad (581) 0 = No posee (5.673)	0	Ficha Social

■ Hacinamiento

La ficha social de la universidad otorga puntaje a aquellas personas donde la cantidad de personas que habitan una vivienda se encuentra por encima de su capacidad. Para ello definen un *índice de hacinamiento*, el cual si es mayor o igual a 3 se considera que el estudiante vive en condiciones de hacinamiento. Este se calcula de la siguiente manera:

$$\text{Índice de Hacinamiento} = \frac{\text{Número Dormitorios}}{\text{Número Integrantes Hogar}}$$

Como la cantidad de alumnos que poseen un índice de hacinamiento mayor que 3 es baja (alumnos), también se consideró un índice de hacinamiento mayor=2 y menor que 3, la cual fue construida a partir de las variables Número de Dormitorios y Número de Integrantes Hogar. Los missing values fueron reemplazados con valor sin hacinamiento ("SIN"), debido a que se trata de personas que no completaron la ficha social, y se posee el supuesto que personas que no la completan no poseen condiciones de hacinamiento.

TABLA B.7: Nueva variable “Hacinamiento”

Nombre	Tipo	Valores	Missing Values	Fuente
Hacinamiento	Catógórica	SIN = índice de hacinamiento < 2 (6.022) H2 = índice de hacinamiento >=2 y menor a 3 (205) H3)= índice de hacinamiento >=3 (27)	0	Ficha Social

■ Nivel de Educación Familiar

Se creó la variable Nivel de Educación Familiar con el objetivo de identificar si influye en la deserción de alumnos el nivel académico de la persona que posee más alta remuneración en el hogar (jefe de hogar) y la madre. Para esto se considerará el nivel de estudios máximo alcanzado por el jefe de hogar y la madre. Esta variable se construyó a partir de las variables nivel de educación jefe de hogar y nivel de educación madre. Para reemplazar los missing values, se utiliza el supuesto de que familias que no postularon a la ficha social (jefe de hogar o la madre) poseen estudios profesionales.

TABLA B.8: Nueva variable “Nivel de Educación Familiar”

Nombre	Tipo	Valores	Missing Values	Fuente
Nivel de educación familiar	Catógórica	Básica (289) Media Completa (867) Técnico Completa (434) Universitaria Completa (4.411)	0	Ficha Social

■ Región

Se decidió agrupar la variable región en 3 categorías según la distancia que se encuentran de la Universidad de Chile, específicamente la Región Metropolitana.

1. Región Metropolitana
2. Región de Valparaíso y Libertador General Bernardo O'higgins.
3. Otras regiones

TABLA B.9: Nueva variable "Región"

Nombre	Tipo	Valores	Missing Values	Fuente
Región	Categórica	1 = RM (5.012) 2 = V y VI región (505) 3 = Otras regiones (737)	0	Guía Matricula

■ Carrera Año Anterior

Con el objetivo de identificar a aquellos alumnos que han cursado una carrera distinta dentro de la universidad sin haberla completado, se creó la variable denominada "Carrera Año Anterior". Esta fue construida a partir de la data histórica de alumnos de los años 2011 a 2016, la cual se contrastó con los registros de alumnos de primer año del cohorte 2017. No se consideraron dentro de esta categoría a alumnos que ya egresaron de otra carrera (4 alumnos).

TABLA B.10: Nueva variable "Carrera anterior no terminada"

Nombre	Tipo	Valores	Missing Values	Fuente
Carrera Año Anterior	Binaria	564 alumnos con carrera anterior no terminada	0	Guía Matricula

■ Créditos

El crédito CAE es complementario con el FSCU, ya que los beneficiarios del CAE pueden financiar la diferencia entre el arancel referencial y real con el FSCU. En efecto, puede existir un efecto de interacción entre los alumnos que poseen ambos créditos que haga que estos alumnos

deserten en menor porcentaje que aquellos que poseen estas becas de forma independiente. Es por ello que se creó una variable binaria de interacción entre ambas becas, la cual consiste en la multiplicación de ambas.

TABLA B.11: Nueva variable “CAE - FSCU”

Nombre	Tipo	Cantidad de alumnos con ambos créditos	Missing Values	Fuente
CAE - FSCU	Binaria	115	0	MINEDUC

■ **Becas de Arancel:**

Las becas de arancel se utilizan para financiar el arancel real, regulado o referencial. El arancel real corresponde al arancel que es determinado por la propia institución en base a sus propias estimaciones de gastos e inversiones. El arancel regulado es determinado por la subsecretaría de educación y el ministerio de hacienda con la colaboración y participación de las universidades adscritas a gratuidad. En tanto que el arancel referencial es fijado anualmente por el MINEDUC en base a estadísticas de calidad docente, tasas de titulación, retención, entre otras. Por lo general, el arancel regulado y referencial son menores al arancel real. Es decir, que los aranceles fijados por el Estado son en general menores a los fijados por las universidades. En anexo se describe a cada una de las beca, indicando su objetivo y nivel de cobertura (arancel real, regulado o referencial).

Las becas de arancel pueden ser tanto internas (Universidad de Chile) como externas (MINEDUC y Municipalidad de Las Condes). Para su asignación se exigen requisitos socioeconómicos y académicos, solicitando generalmente pertenecer a la población de menores ingresos y obtener resultados PSU sobre 500 o 700 puntos dependiendo del tipo de beca. Además, poseen requisitos de renovación, como haber aprobado al menos el 60 % de los ramos realizados durante el año.

TABLA B.12: Análisis Becas de Arancel

N	Nombre Beca	Requisitos			
		Socioeconómicos	Prom. PSU	NEM	Renovación
1	Beca Universidad de Chile (BUCH)	En tramo 40 % y 90 %	≥ 700	$\geq 6,0$	No reprobado más de dos asignaturas en primeros cuatro semestres
2	Andrés Bello	-	Ptje. Nacional	$\geq 6,0$	No reprobado más de dos asignaturas en primeros cuatro semestres
3	Beca Bicentenario	70 % menores ingresos	> 500	-	60 % avance académico
4	Beca para Estudiantes Hijos de Profesionales de la Educación	80 % menores ingresos	≥ 500	$\geq 5,5$	60 % avance académico
5	Beca Vocación del Profesor	-	≥ 600	-	60 % avance académico
6	Beca Puntaje PSU	80 % menores ingresos. y provenir de colegio mun. o subv.	Ptje. Nacional	-	-
7	Beca Excelencia Académica	80 % menores ingresos. y provenir de colegio mun. o subv.	10 % más alto del establecimiento	-	-
8	Beca de Reparación Valech	-	-	-	-
9	Gratuidad	60 % menores ingresos	-	-	-
10	Beca Las Condes	Establecido por Muni. Las Condes	161-	Establecido por Muni. Las Condes	Establecido por Muni. Las Condes

En lo que respecta a la muestra de estudio, se identifica que hay becas que no fueron asignadas al cohorte de alumnos nuevos del año 2017 que rindieron la PSU (6.254 alumnos), por lo cual estas fueron descartadas del estudio de forma directa. Estas corresponden a Beca Juan Gómez Millas, Juan G. Millas Extranjeros, primera generación en pedagogía, convenio internacional Chile- Colombia, entre otras.

En cuanto a las becas de arancel que sí fueron asignadas a alumnos pertenecientes a la muestra, se identifican patrones de asignación que son comunes a las becas, como por ejemplo puntajes mínimos de PSU superiores a 700 puntos. Estos requisitos provocan en si mismo un sesgo en los resultados de deserción de alumnos que poseen las becas, no siendo atribuible directamente la no deserción al obtener una beca, si no que al hecho de discriminar por rendimiento académico. Este efecto es conocido como endogeneidad, es decir, que la ayuda financiera está destinada a ciertos grupos, que cuentan con características que a su vez, están relacionadas con la variable de resultado (Verónica Santelices Estudio Pontificia Universidad Católica). Es por ello que se crean dos nuevas variables que se encuentren asociadas por requisitos mínimos de rendimiento académico:

1. Becas arancel tipo I

Corresponde a las becas Universidad de Chile (BUCH), Andrés Bello, Puntaje PSU y Beca Excelencia Académica. Estas poseen requisitos académicos de promedio de PSU de Lenguaje y Matemáticas sobre 700 puntos, NEM sobre 6,0 o pertenecer al 10% con resultados PSU más alto de su establecimiento.

2. Becas de arancel tipo II

Corresponde a las becas Bicentenario, BHPE, Beca Las Condes y Reparación Valech. Estas exigen un promedio PSU de Matemáticas y Lenguaje de al menos 500 puntos y promedio de enseñanza media superior a 5,5. Si bien la Beca de Reparación Valech no posee exigencias académicas, esta se considera dentro de este grupo debido a que el promedio mínimo PSU del cohorte 2017 de la Universidad de Chile fue de 558 puntos¹.

¹Anuario 2017 Universidad de Chile p. 21 (sin considerar los criterios de ponderación particulares de cada carrera)

TABLA B.13: Nuevas variables “Becas de Arancel tipo I y II”

Nombre	Tipo	Missing Values	Fuente	Beneficiados	Monto total (mm CLP)
Becas de arancel tipo I	Binaria	0	U. de Chile	192	\$ 447
Becas de arancel tipo II	Binaria	0	U. de Chile	742	\$ 2.189

Beca Vocación de Profesor

Esta beca posee requisitos académicos de promedio PSU de Lenguaje y Matemáticas superior a 600 puntos. Esta va acompañada de la Beca de Mantención del Profesor, la cual consiste en la entrega de \$80.000 mensuales para gastos de libre disposición. Se posee la hipótesis que: dado los requisitos académicos mínimos para postular a este beneficio, la flexibilidad de dedicar tiempo a las labores universitarias al contar con dineros mensuales para uso de libre disposición y al supuesto de que los alumnos que poseen esta beca tienen vocación de estudiar pedagogía, los alumnos con esta beca tendrán tasas de deserción menor en comparación al resto de los estudiantes.

Gratuidad

Este beneficio posee exigencias netamente socioeconómicas y es poseída por el 41 % de la muestra (2.555 alumnos). Además, los alumnos que poseen esta beca, se encuentran beneficiados con la beca de mantención de la U. de Chile “Beca de Atención Económica” a través de la política interna “Piso Básico Garantizado” que corresponde a una asignación mensual de \$23.700 para todos aquellos estudiantes que sean beneficiarios de gratuidad. La gratuidad financia el 100 % del arancel regulado, no requiriendo el alumno pagar el diferencial con el arancel real. Dada la cantidad de alumnos que poseen este beneficio y los montos destinados al financiamiento de esta beca (\$9.210 millones, equivalente a 3,2 veces el financiamiento del CAE y FSCU en esta muestra), se estudiarán los efectos de esta de forma independiente.

■ **Porcentaje de cobertura de créditos y becas**

Con el objetivo de identificar si el porcentaje de cobertura de arancel juega un rol relevante en la retención del alumno, se crearon dos variables:

-
1. **Porcentaje de cobertura de créditos:** Corresponde al porcentaje de cobertura de la carrera con créditos universitarios (CAE o FSCU).

$$\text{Porcentaje cobertura créditos} = \frac{\text{Monto CAE} + \text{Monto FSCU}}{\text{Arancel real de la carrera}} * 100\%.$$

2. **Porcentaje de cobertura de becas y gratuidad:** Corresponde al porcentaje de cobertura de la carrera con becas de arancel y gratuidad.

Si posee gratuidad:

$$\text{Porcentaje cobertura arancel} = 100\%$$

Si no:

$$\text{Porcentaje cobertura arancel} = \max\left(\frac{\sum \text{Monto Beca}}{\text{Arancel real de la carrera}} * 100\%, 100\%\right)$$

El motivo de establecer una cota máxima en la beca de arancel se debe a que hay casos en donde los montos de asignaciones superan el 100% del arancel real, debido a que se les asignan beneficios que terminan por cubrir más del costo real de la carrera. Esto es regularizado posteriormente por la unidad de aranceles.

■ **Becas de Mantención**

El estudio de las becas de mantención se realizará según el objetivo que buscan alcanzar. El detalle a continuación:

1. **Beca de Apoyo Preescolar:** Apoyo a madres y padres del cuidado de sus hijos o hijas. Total de 8 beneficiados.
2. **Beca de Emergencia:** Obtener ayuda económica en caso de tener una emergencia que dificulte el normal desarrollo de estudios del alumno. Total de 39 beneficiados.
3. **Beca Compensación Laboral:** Obtener una retribución económica por trabajos realizados en la universidad. 1 beneficiado.

-
4. **Becas de Residencia:** Apoyo para alumnos de regiones diferentes a la metropolitana que no poseen situación económica ni redes de contacto para encontrar alojamiento en la capital. Total de 49 beneficiados.
 - Becas de Residencia Interna.
 - Beca de Residencia Sistema Prioritario de Equidad Educativa SIPEE.
 - Hogares Universitarios.
 5. **Becas de libre disposición y de alimentación:** Becas destinadas a cubrir gastos de transporte, alimentación u otros que pudiera requerir el alumno para el normal desarrollo de sus estudios. Total de 5.687 beneficiados.
 - Beca de Atención Económica (BAE)
 - Becas de Alimentación para la educación Superior (BAES)
 - Beca de Mantención Educación Superior (BMES)
 - Beca de Mantención Vocación del Profesor
 - Beca Presidente de la República
 6. **Becas Pueblos Originarios:** Becas destinadas a la inclusión de estudiantes de ascendencia indígena. Total de 45 beneficiados.
 - Beca Indígena
 - Beca Residencia Indígena
 7. **Becas de Integración Territorial:** Becas destinadas a cubrir gastos de viajes u otros que pudieran requerir alumnos de zonas extremas del país. Total de 37 beneficiados.
 - Beca de Integración Territorial
 - Beca Patagonia Aysén
 8. **Beca máximos puntajes ponderados mujeres:** Reconocimiento a mujeres destacadas de la universidad. Total de 5 beneficiadas.
 - Beca Irma Salas

Las becas de Apoyo Prescolar, Compensación Laboral y máximo puntajes ponderados mujeres fueron descartas del estudio debido a la baja cantidad de beneficiados (menos de 10 cada una) en contraste al total de la muestra (6.254 alumnos). La baja cantidad de beneficiados provoca que los modelos no tengan insuficiente evidencia para determinar la significancia de estas variables, por lo que es natural descartarlas del estudio.

B.7. Distribución de Variables

1. PSU Matemática Final y Lenguaje Final poseen una distribución normal a simple vista (ver figura B.1 y B.2), lo cual fue corroborado para cada una de estas variables con el test de stata de skewness y kurtosis (sktest), el cual indicó que distribuyen como una normal con más de un 99% de confianza ($p=0.000$)

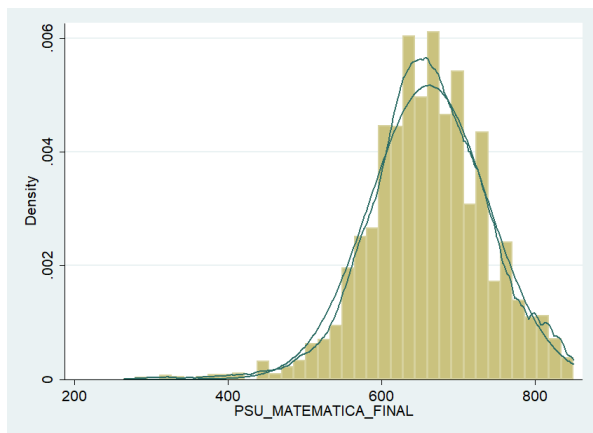


FIGURA B.1: Histograma PSU Matemáticas.
Fuente: Elaboración propia.

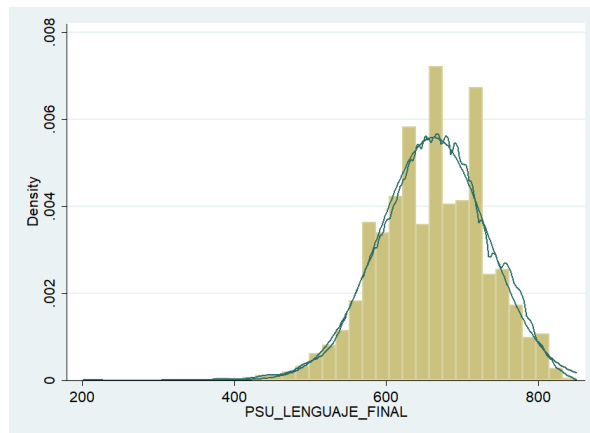


FIGURA B.2: Histograma PSU Lenguaje.
Fuente: Elaboración propia.

2. La variable PSU Ranking posee una cola hacia la izquierda (left skewness) y un peak al lado derecho (ver imagen B.3). Se aplicó el test de skewness y kurtosis de stata, obteniendo como resultado que la variable distribuye como una normal ($p=0.000$). Si bien el resultado de normalidad fue positivo, se aplicaron transformaciones cuadráticas, raíz cúbica y logaritmo para reducir el peak del lado derecho. Pese a ello, el peak se mantuvo. La razón de este peak es el método de cálculo del Puntaje Ranking, el cual asigna 850 puntos a todos aquellos alumnos que hayan tenido un promedio igual o superior al máximo promedio de los últimos 3 años de su establecimiento educacional (para más detalles ver en anexos método de cálculo de Ranking). Se posee la hipótesis de que alumnos que hayan tenido un puntaje ranking equivalente a 850 posean una retención más alta, ya que se trata de alumnos que destacan académicamente dentro de sus propios contextos educativos. Es por ello que se creó la variable binaria Flag_PSU_Ranking_Maximo, la cual toma valor 1 si el estudiante obtuvo un ranking de 850 puntos y 0 si no.

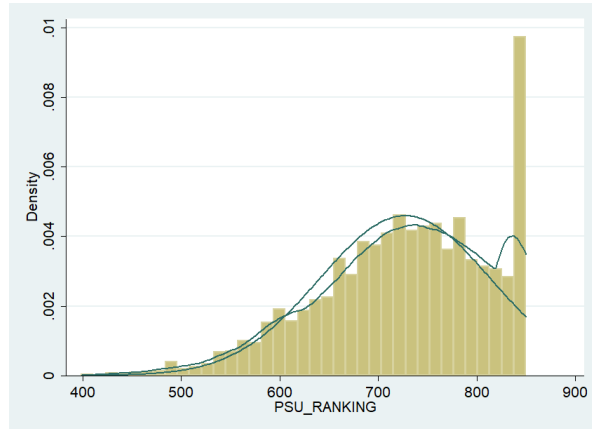


FIGURA B.3: Histograma PSU Ranking. Fuente: Elaboración propia.

3. La variable “IVE (1)” posee un peak en el sector izquierdo y una distribución normal en el sector derecho (ver imagen B.4). Ello se debe a que en este estudio se utilizó el supuesto de que alumnos de colegios particulares poseen un índice de vulnerabilidad escolar menor al de colegios municipales y subvencionados, por lo cual se les asignó valores aleatorios entre 0 y 13%. Realizando el test de skewness y kurtosis de stata, se obtuvo que esta variable no distribuye como una normal. Para lograr que esta variable distribuya como una normal, se aplicaron transformaciones raíz cuadrada y raíz cúbica para reducir el peak izquierdo (no se utilizó logaritmo porque este se indefine para registros que toman valor 0). La mejor transformación aplicada corresponde a raíz cúbica (ver imagen B.5), donde el test de stata arrojó que corresponde a una normal con un nivel de significancia $p=0.000$. De todas maneras, una solución más robusta para este caso es utilizar la variable categórica “IVE (2)”.

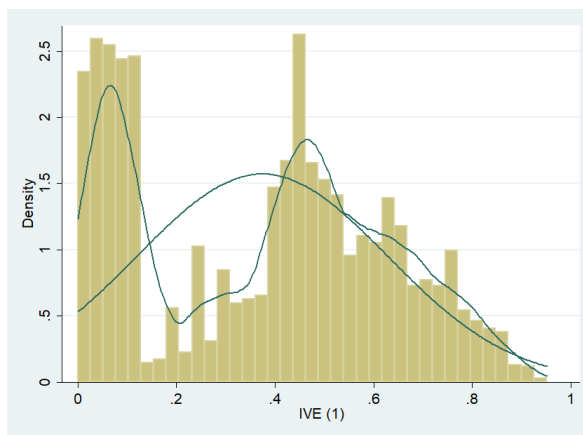


FIGURA B.4: Histograma IVE (1). Fuente: Elaboración propia.

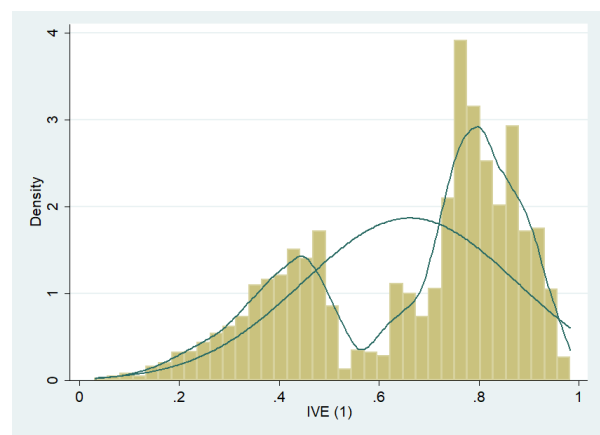


FIGURA B.5: Histograma raíz cúbica IVE (1). Fuente: Elaboración propia.

4. En tanto que la variable “Porcentaje Cobertura Arancel” posee dos peaks: uno en el sector izquierdo y otro en el sector derecho. El primero corresponde a alumnos que no poseen becas de arancel (2.709 alumnos de 6.254) mientras que el derecho corresponde principalmente a alumnos que poseen gratuidad (2.555 alumnos). La mejor opción en este caso es sólo considerar las variables binarias que conforman esta variable (gratuidad, arancel tipo I, arancel tipo II y Vocación profesor).

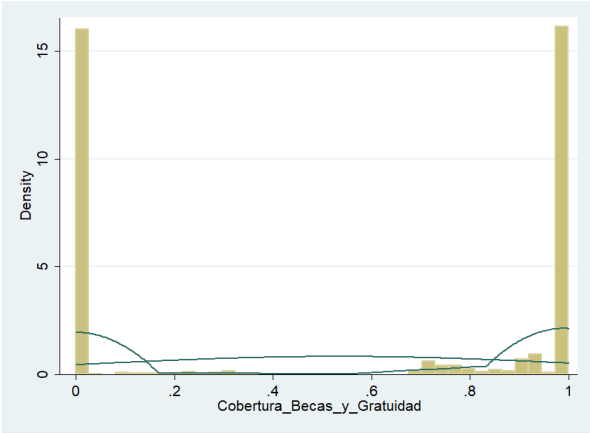


FIGURA B.6: Histograma Porcentaje Cobertura Arancel. Fuente: Elaboración propia.

5. En el caso de “Porcentaje Cobertura Crédito” ocurre algo similar con Porcentaje Cobertura de Arancel, ya que el peak de la izquierda se debe a que 5.158 alumnos no poseen crédito (82% del total de la muestra). De la misma manera, la mejor opción para esta variable es considerar sólo las variables binarias que la conforman (CAE y FSCU).

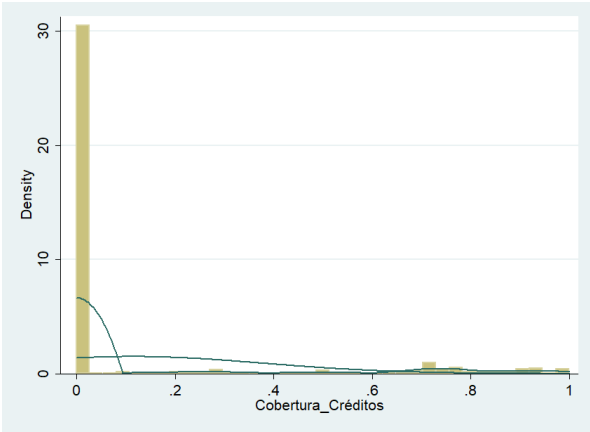


FIGURA B.7: Histograma Porcentaje Cobertura Crédito. Fuente: Elaboración propia.

B.8. Proceso clusterización con variables reales

B.8.1. X Means

El algoritmo de X- Means entregó como resultado óptimo la generación de dos clusters (ver imagen B.8).

El primer cluster (color azul) se encuentra compuesto por 3.551 alumnos, y este corresponde a alumnos con un alto Índice de Vulnerabilidad Escolar (IVE) y que obtuvieron un menor puntaje en las pruebas de PSU de Matemáticas y Lenguaje.

En tanto que el segundo cluster (color rojo) está compuesto por 2.903 alumnos, y se trata de alumnos con un bajo IVE y que obtuvieron un mayor puntaje en las pruebas de PSU de Matemáticas y Lenguaje.

Dado estos resultados, el primer cluster puede ser nombrado como "alumnos vulnerables" mientras que el segundo "alumnos no vulnerables".

En cuanto a los desertores de cada cluster (ver imagen B.9) no es claro visualmente si uno de los dos clusters posee una mayor concentración de alumnos desertores.

Contrastando estos gráficos con el Índice de Vulnerabilidad de Escolar (ver imagen B.10), se valida visualmente que el cluster superior corresponde mayoritariamente a alumnos que provienen de colegios vulnerables.

Notar que el algoritmo rescata los efectos de cada una de las variables, generando un corte diagonal a la muestra (ver imagen B.8), no realizando una separación dicotómica entre alumnos con un alto y bajo IVE.

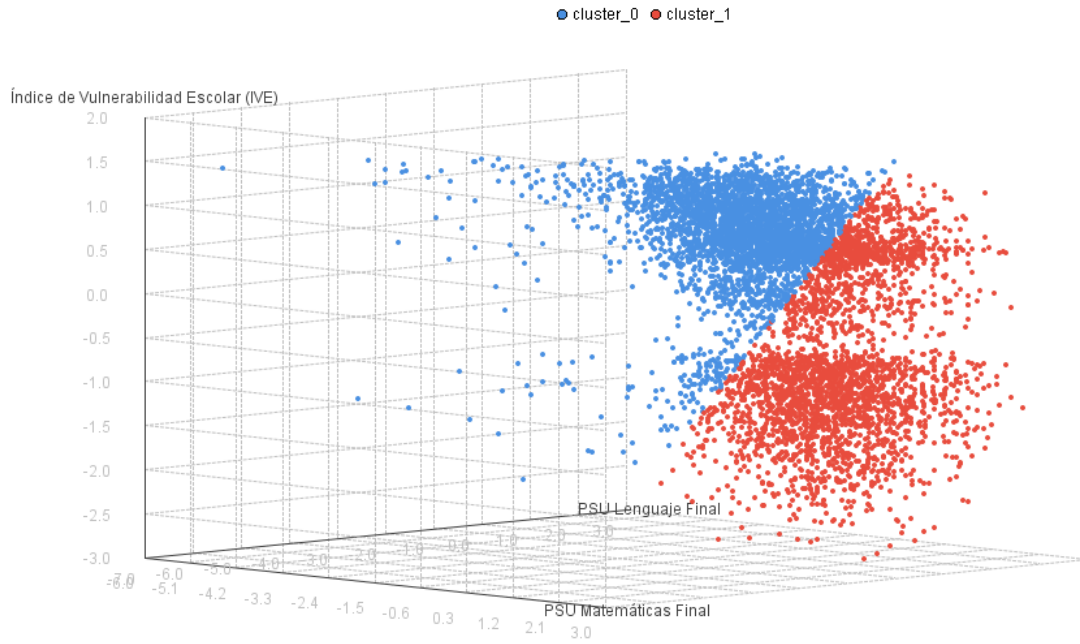


FIGURA B.8: Cluster X Mean

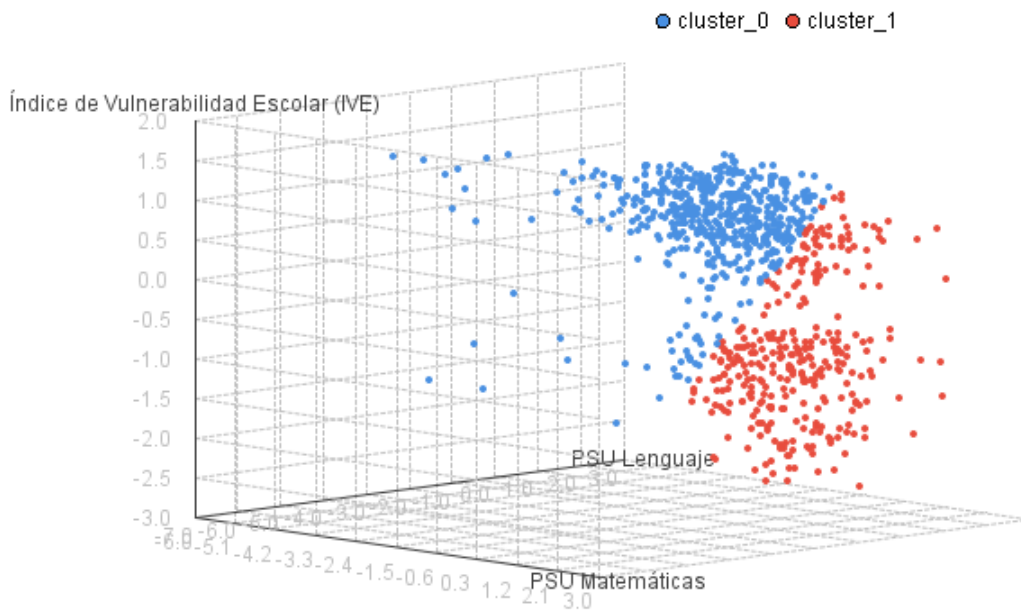


FIGURA B.9: Cluster X Mean - Desertores. Fuente: Elaboración propia.

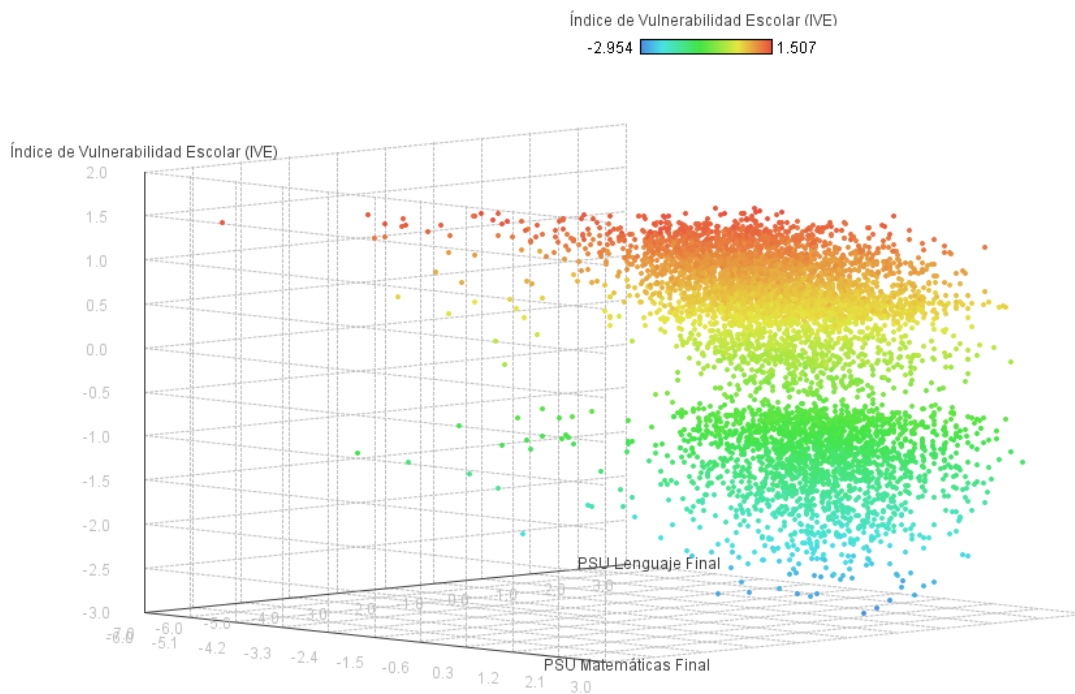


FIGURA B.10: PSU Matemática, Lenguaje Final e IVE. Fuente: Elaboración propia.

B.8.2. DBSCAN

Con respecto al algoritmo DBSCAN, se generaron diferentes instancias con distintos valores de EPS y min points (ver anexo B.9).

El cluster escogido fue el con 0.25 epsilon y 20 min points. El motivo de escoger esta configuración de epsilon y min points corresponde a que esta asociación arrojó cuatro cluster, donde dos de ellos poseen más de 30 registros, abordan en conjunto a más del 51 % de la muestra y a que sus resultados son interpretables. Si bien estos resultados se asemejan a la configuración N=9 y N=12 (ver anexo B.9), se optó por una configuración que posea un menor epsilon (0.25) y mayor min points (20) para así contar con alumnos que posean características más similares dentro de cada cluster.

Los cluster con la configuración escogida pueden observarse en la imagen B.11, la cual cuenta con cuatro clusters:

- Cluster 1: 2667 registros. Corresponde a alumnos de un alto índice de vulnerabilidad escolar y puntajes PSU medios y altos (color azul).

-
- Cluster 2: 549 registros. Corresponde a alumnos de un bajo índice de vulnerabilidad escolar y puntajes PSU altos (color verde).
 - Cluster 3: 20 registros. Corresponde a alumnos de un alto índice de vulnerabilidad escolar y puntajes PSU medios (color rojo).
 - Cluster 4: 20 registros. Corresponde a alumnos de un bajo índice de vulnerabilidad escolar y puntaje PSU medios (color amarillo).
 - El resto de los registros son considerados como ruido (2.998 registros = 48 % de la muestra).

De la misma forma que los resultados obtenidos por el método de X-Means, el cluster 1 y 3 pueden nombrarse como "alumnos vulnerables" mientras que el cluster 2 y 3 como "alumnos no vulnerables".

En cuanto a los desertores de cada cluster, se puede apreciar en la imagen B.12 que hay una mayor concentración de alumnos que deserta en el cluster 1. Empíricamente, el cluster 1 (color azul) deserta un 13,4% (357 alumnos) y en tanto que en el cluster 2 (color verde) un 11,8% (65 alumnos), comprobando que la tasa de deserción de alumnos de cluster 1 es mayor que el cluster 2.

Notar que a diferencia de el algoritmo de x-means, el cual realiza un corte diagonal a la muestra (ver imagen B.13), el algoritmo de DBSCAN hace una separación categórica entre alumnos que provienen de colegios vulnerables y no vulnerables (ver imagen B.11).

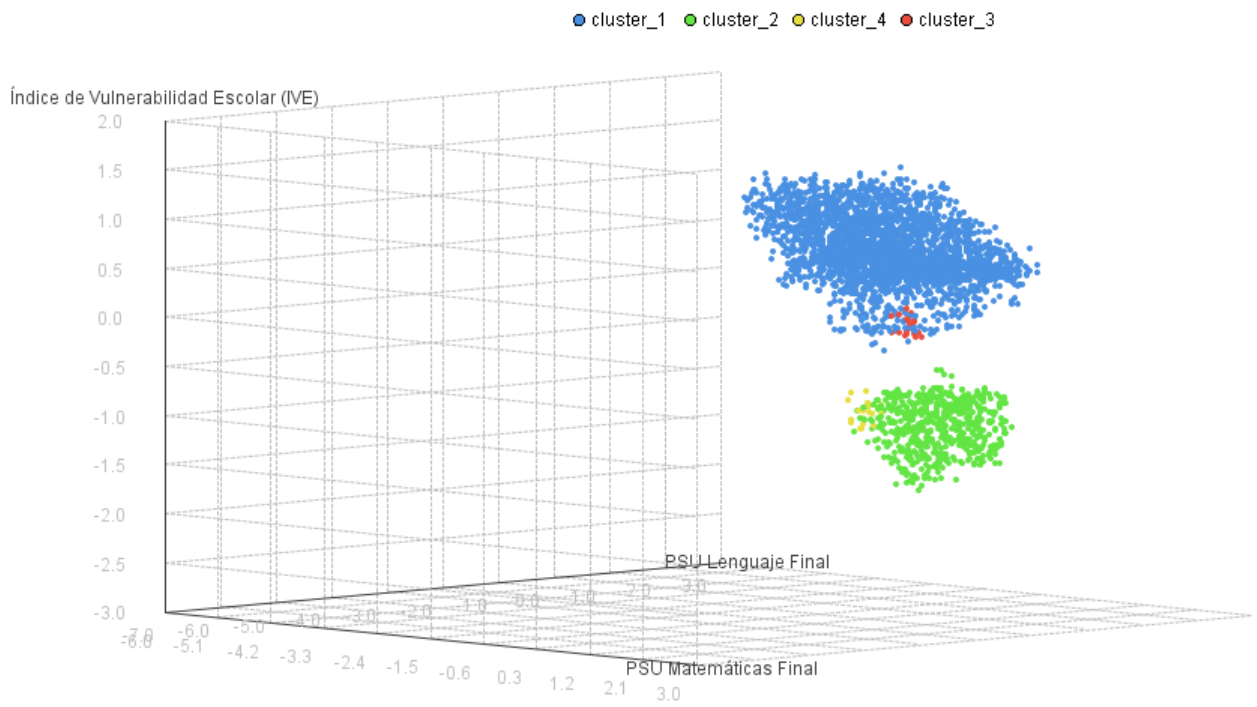


FIGURA B.11: DBSCAN. Fuente: Elaboración propia.

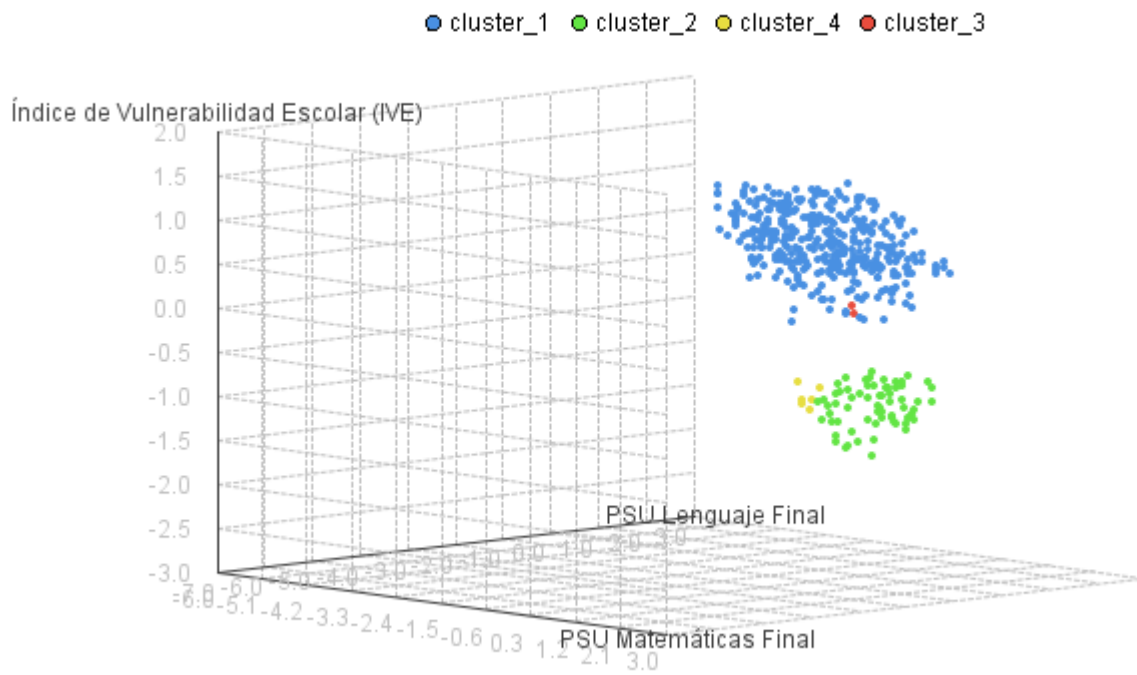


FIGURA B.12: DBSCAN - Desertores. Fuente: Elaboración propia.

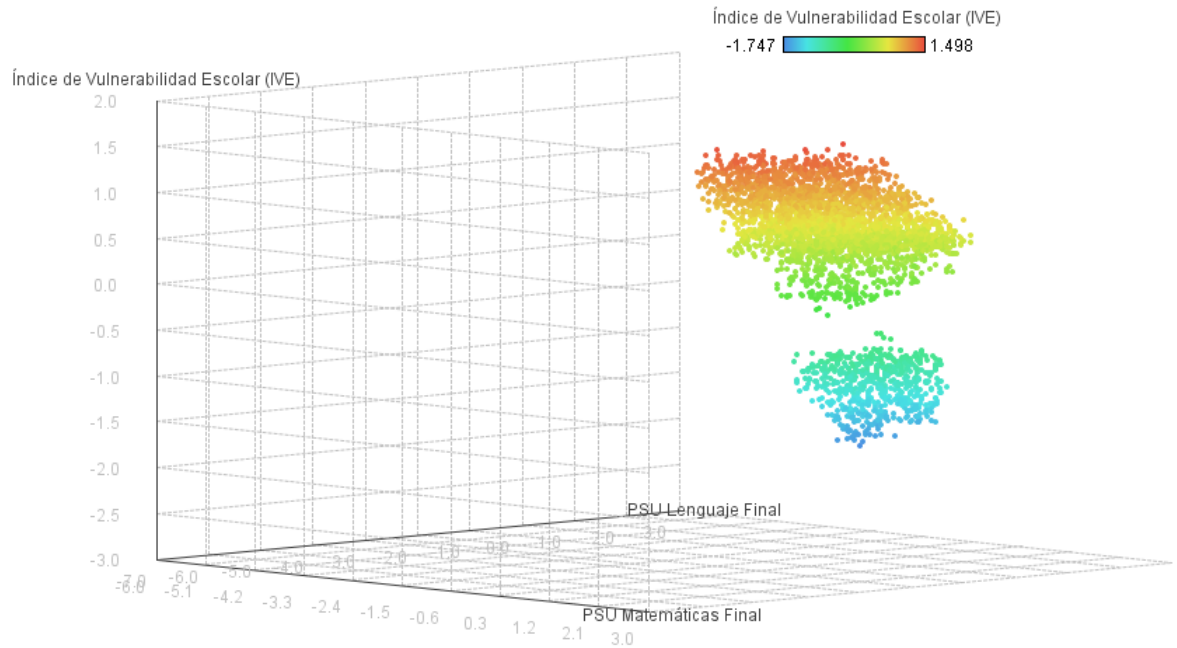


FIGURA B.13: DBSCAN - PSU Matemática, PSU Lenguaje e IVE. Fuente: Elaboración propia.

Dado los resultados de los distintos mecanismos de cluster, se adicionarán los resultados óptimos obtenidos con DBSCAN (epsilon=0.25 y 20 min points) como variable explicativa en los modelos de predicción. El motivo de escoger estos resultados, es que como se explicó anteriormente, existen diferencias las tasas de deserción de los cluster más relevantes (cluster 1 y 2), mientras que en los resultados obtenidos con el método de X-Means no son claras las diferencias de deserción entre clusters.

Como los cluster 3 y 4 poseen pocos registros (menos de 30, lo cual provocará que estos sean desestimados de los algoritmos de predicción) y además se encuentran próximos a los cluster 1 y 2 (ver imagen B.11) se optó por realizar una unión de cluster 1 (azul) con el cluster 3 (rojo) y del cluster 2 (verde) con el cluster 4 (amarillo), quedando como resultado dos clusters en vez de cuatro.

B.9. Selección epsilon y min points DBSCAN

TABLA B.14: Selección epsilon y min points DBSCAN

N	epsilon	min point	Clusters	Clusters con más de 30 registros	Registros considerados como ruido	Porcentaje de ruido
1	1.0	5	1	1	7	0 %
2	0.5	5	2	1	91	1 %
3	0.1	5	137	0	5189	83 %
4	0.20	10	22	4	2951	47 %
5	0.25	10	13	2	1452	23 %
6	0.30	10	3	1	766	12 %
7	0.20	15	17	3	4032	64 %
8	0.25	15	5	2	2383	38 %
9	0.30	15	4	2	1265	20 %
10	0.20	20	5	1	4797	77 %
11	0.25	20	4	2	2998	48 %
12	0.30	20	3	2	1830	29 %

B.10. Proceso clusterización con variables categóricas

B.10.1. X Means

Utilizando la distancia nominal entre variables categóricas con el algoritmo de X-Means, se obtuvieron 4 clusters (ver tabla B.15):

- Cluster 1: 107 registros. Corresponde a alumnos que poseen Beca Vocación de Profesor.
- Cluster 2: 2.555 registros. Corresponde a alumnos que poseen Gratuidad.

- Cluster 3: 924 registros. Corresponde a alumnos que poseen CAE o lo combinación de créditos CAE-FSCU.
- Cluster 4: 2.688 registros. Corresponde a alumnos que no poseen Beca Vocación de Profesor, Gratuidad ni CAE.

TABLA B.15: Cluster X- Mean: Proporción de true values de variables binarias

Variable	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Total
Carrera Anterior No Terminada	2 %	48 %	15 %	35 %	100 %
CAE-FSU			100 %		100 %
CAE			100 %		100 %
Gratuidad		100 %			100 %
Vocación Profesor	100 %				100 %

Como los resultados poseen una correlación de 100 % con tener o no una de las becas anteriormente mencionadas, se descartan estos resultados para ser utilizados en los algoritmos predictivos, ya que no agregan valor a las variables existentes (los cluster generados son reemplazables por las variables binarias Beca Vocación Profesor, Gratuidad y CAE).

B.10.2. DBSCAN

La configuración óptima escogida de DBSCAN correspondió a un epsilon de 1.1 y 300 min points. Esta combinación posee 2 clusters (ver tabla B.16):

- Cluster 1: 1.996 registros. Corresponde a alumnos que poseen ingresos económicos bajos (tramo 40 y 50) y Gratuidad. Más del 40 % de los alumnos de las carreras de Medicina, Ingeniería Comercial e Ingeniería y Ciencias Plan Común pertenecen a este cluster. No posee diferencias relevantes con respecto a región.

- Cluster 2: 1.857 registros. Corresponde a alumnos que poseen ingresos económicos altos (tramo 90, 100 y sin calificación económica principalmente) y CAE. No poseen gratuidad. Más del 40 % de los alumnos de las carreras de Administración Pública, Bioquímica, Fonoaudiología, Geografía, Ingeniería Forestal, Kinesiología, Medicina Veterinaria, Nutrición y Dietética, Obstetricia y Puericultura, Pedagogía en Educación Parvularia, Química y Farmacia, Terapia Ocupacional y Trabajo Social pertenecen a este cluster. No posee diferencias relevantes con respecto a región.
- El resto de los registros considerados como ruido (2.401 registros= 38 % de la muestra).

De la misma forma que anterior análisis, estos cluster pueden ser nombrados como “alumnos vulnerables”(cluster 1) mientras que el cluster 2 como “alumnos no vulnerables”.

TABLA B.16: DBSCAN: Proporción de true values de una selección de variables categóricas

Variable	Cluster 1	Cluster 2	Ruido	Total
Carrera Anterior No Terminada	11 %	11 %	77 %	100 %
CAE-FSCU			100 %	100 %
CAE		14 %	86 %	100 %
Gratuidad	78 %		22 %	100 %
Vocación Profesor			100 %	100 %
tramo 40	89 %		11 %	100 %
tramo 50	45 %		55 %	100 %
tramo 60		17 %	83 %	100 %
tramo 70		18 %	82 %	100 %
tramo 80		16 %	84 %	100 %
tramo 90		30 %	70 %	100 %
tramo 100		42 %	58 %	100 %
Sin Calificación Socioeconómica		90 %	10 %	100 %

En cuanto a deserción, el cluster 1 y el cluster 2 poseen tasas casi idénticas de deserción (12,78% y 12,87% respectivamente). Se utilizará de todas maneras como variable explicativa de los modelos predictivos.

B.11. Puntaje Ranking

El Puntaje Ranking es uno de los factores de selección del proceso de admisión de alumnos en universidades de Chile, el cual tiene por objetivo ayudar a seleccionar a los mejores estudiantes durante su trayectoria escolar y favorecer la equidad en el acceso al sistema universitario.

El método de cálculo se realiza considerando el promedio acumulado del estudiante (promedio de nota de primero a cuarto medio) y una población de referencia que considera las 3 generaciones que realizaron sus estudios en el contexto educativo del alumno. Este consiste en los siguientes pasos:

Paso 1: Cálculo de promedio máximo y promedio de notas del contexto educativo

- Cálculo prom. máximo histórico de notas contexto educativo

$$\text{Promedio máximo histórico de notas contexto educativo} = (\text{MAXG1} + \text{MAXG2} + \text{MAXG3}) / 3$$

Donde:

MAXG1 = Promedio máximo acumulado de la primera generación anterior del estudiante.

MAXG2 = Promedio máximo acumulado de la segunda generación anterior del estudiante.

MAXG3 = Promedio máximo acumulado de la tercera generación anterior del estudiante.

- Promedio histórico de notas contexto educativo

$$\text{Promedio histórico de notas contexto educativo} = (\text{PAG1} + \text{PAG2} + \text{PAG3}) / 3$$

Donde:

PAG1 = Promedio acumulado de la primera generación anterior del estudiante.

PAG2 = Promedio acumulado de la segunda generación anterior del estudiante.

PAG3 = Promedio acumulado de la tercera generación anterior del estudiante.

Paso 2: Cálculo Puntaje Ranking

-
- Si el promedio acumulado del estudiante es igual o superior al promedio máximo histórico de notas del contexto educativo este obtiene el máximo puntaje ranking (850 puntos).
 - Si el promedio acumulado del estudiante es superior al promedio de notas del contexto educativo pero menor al promedio máximo histórico, obtiene un puntaje equivalente al NEM más una bonificación que crece linealmente desde el NEM obtenido por el promedio histórico de notas del contexto educativo hasta los 850 puntos.
 - Si el promedio acumulado del estudiante es menor al promedio de notas del contexto educativo este obtiene un puntaje equivalente a su puntaje NEM (no recibe bonificación del ranking).
 - En el caso de que el alumno haya estado en más de un establecimiento educacional, el cálculo se realiza de la misma forma para cada establecimiento y se pondera por la cantidad de años que estuvo cada uno. A modo de ejemplo, si el alumno estuvo en primero y segundo medio en el colegio 1 y tercero y cuarto medio en el colegio 2, se calculará el promedio máximo histórico y el promedio del colegio 1 de primero y segundo medio, y se obtendrá el puntaje ranking del colegio 1. De la misma forma para el colegio 2. Finalmente, se calculará el puntaje ranking como $\text{puntaje ranking} = ((\text{puntaje ranking colegio 1}) * (\text{años cursados en colegio 1}) + (\text{puntaje ranking colegio 2}) * (\text{años cursados en colegio 2})) / 4 = ((\text{puntaje ranking colegio 1}) * 2 + (\text{puntaje ranking colegio 2}) * 2) / 4$

B.12. Índice de Vulnerabilidad Escolar (IVE)

Corresponde a la proporción de alumnos de un establecimiento educacional (municipal o subvencionado) que poseen una condición socioeconómica de pobreza o extrema pobreza, lo cual es determinado por la JUNAEB a través de la metodología de Medición de la Condición de Vulnerabilidad (SINAE), la cual determina la dimensión socioeconómica del alumno en base a encuestas de vulnerabilidad e información proveniente de la afiliación de salud del estudiante (FONASA o ISAPRE), SENAME, Registro Civil, MINEDUC, entre otras, considerando variables como nivel socioeconómico de la familia, notas, asistencia, repitencia, puntaje SIMCE, entre otras.

Método de cálculo:

$$A = \frac{1^{ra} P + 2^{da} P + 3^{ra} P}{Matricula} * 100 \% \quad (B.1)$$

Donde:

$1^{ra} P$ = Alumnos catalogados como primera prioridad por encuestas de vulnerabilidad de JUNAEB (condición socioeconómica de extrema pobreza).

$2^{da} P$ = Alumnos catalogados como segunda prioridad por encuestas de vulnerabilidad de JUNAEB (condición socioeconómica de pobreza).

$3^{ra} P$ = Alumnos catalogados como tercera prioridad por encuestas de vulnerabilidad de JUNAEB (condición socioeconómica de pobreza).

B.13. Coeficientes Regresión Logística

TABLA B.17: Coeficientes Regresión Logística

Variable	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
BA_Gratuidad	1,41	0,69	0,40	3,53	0,00
BA_Vocacion_Profesor	1,08	0,14	0,43	2,53	0,01
CAE	0,64	0,23	0,15	4,29	0,00
CARR_ADMINISTRACIÓN PÚBLICA	0,08	0,08	0,66	0,12	0,90
CARR_ANTROPOLOGÍA-ARQUEOLOGÍA	-0,39	-0,39	0,63	-0,61	0,54
CARR_ARQUITECTURA	-1,75	-1,75	0,50	-3,53	0,00
CARR_ARTES PLÁSTICAS	-1,57	-1,57	0,56	-2,83	0,00
CARR_BIOQUÍMICA	-1,87	-1,87	0,59	-3,14	0,00
CARR_CINE Y TELEVISIÓN	-0,20	-0,20	0,70	-0,29	0,77
CARR_CONTADOR AUDITOR	-1,40	-1,40	0,57	-2,46	0,01
CARR_DERECHO	-1,05	-1,05	0,50	-2,12	0,03
CARR_DISEÑO	-2,19	-2,19	0,51	-4,29	0,00
CARR_ENFERMERÍA	-1,97	-1,97	0,53	-3,72	0,00
CARR_FILOSOFÍA	-2,02	-2,02	0,54	-3,73	0,00
CARR_FONOAUDIOLOGÍA	-1,02	-1,02	0,64	-1,60	0,11
CARR_GEOGRAFÍA	-0,90	-0,90	0,67	-1,35	0,18
CARR_HISTORIA	-1,64	-1,64	0,54	-3,01	0,00
CARR_INGENIERÍA AGRONÓMICA	-1,68	-1,68	0,52	-3,23	0,00
CARR_INGENIERÍA COMERCIAL	-0,95	-0,95	0,49	-1,93	0,05
CARR_INGENIERÍA EN ALIMENTOS	-2,61	-2,61	0,57	-4,59	0,00
CARR_INGENIERÍA EN BIOTECNOLOGÍA MOLECULAR	-2,02	-2,02	0,60	-3,36	0,00
CARR_INGENIERÍA EN INFORMACIÓN Y CONTROL DE GESTIÓN	-1,17	-1,17	0,54	-2,17	0,03
CARR_INGENIERÍA EN RECURSOS NATURALES	-0,65	-0,65	0,69	-0,94	0,35
CARR_INGENIERÍA FORESTAL	-0,80	-0,80	0,62	-1,29	0,20
CARR_INGENIERÍA Y CIENCIAS - PLAN COMÚN	-0,80	-0,80	0,48	-1,66	0,10
CARR_KINESIOLOGÍA	-1,66	-1,66	0,59	-2,79	0,01
CARR_LINGÜÍSTICA Y LITERATURA	-1,86	-1,86	0,53	-3,48	0,00
CARR_LINGÜÍSTICA Y LITERATURA INGLESAS	-1,23	-1,23	0,58	-2,13	0,03
CARR_LICENCIATURA	-1,78	-1,78	0,52	-3,39	0,00
CARR_MEDICINA VETERINARIA	-0,97	-0,97	0,53	-1,82	0,07
CARR_MUSICA, DANZA, ARTE Y TEATRO	-1,06	-1,06	0,54	-1,96	0,05

TABLA B.18: Coeficientes Regresión Logística (continuación)

Variable	Coefficient	Std. Coefficient	Std. Error	z-Value	p-Value
CARR._NUTRICIÓN Y DIETÉTICA	-1,36	-1,36	0,62	-2,19	0,03
CARR._OBSTETRICIA Y PUERICULTURA	-1,81	-1,81	0,54	-3,35	0,00
CARR._ODONTOLOGÍA	-1,83	-1,83	0,53	-3,47	0,00
CARR._PEDAGOGIA	-2,37	-2,37	0,56	-4,21	0,00
CARR._PEDAGOGÍA EN EDUCACIÓN PARVULARIA	-0,72	-0,72	0,75	-0,97	0,33
CARR._PEDAGOGÍA EN EDUCACIÓN BÁSICA	-1,74	-1,74	0,68	-2,57	0,01
CARR._PERIODISMO	-0,26	-0,26	0,61	-0,42	0,67
CARR._PROGRAMA ACADÉMICO DE BACHILLERATO	-2,80	-2,80	0,48	-5,86	0,00
CARR._PSICOLOGÍA	-0,73	-0,73	0,60	-1,22	0,22
CARR._QUÍMICA	-2,22	-2,22	0,57	-3,87	0,00
CARR._QUÍMICA Y FARMACIA	-1,56	-1,56	0,52	-2,99	0,00
CARR._QYBAMBIENTAL	-1,63	-1,63	0,56	-2,92	0,00
CARR._SOCIOLOGÍA	-0,81	-0,81	0,60	-1,36	0,17
CARR._TECNOLOGÍA MÉDICA	-1,83	-1,83	0,54	-3,39	0,00
CARR._TERAPIA OCUPACIONAL	-1,08	-1,08	0,64	-1,70	0,09
CARR._TRABAJO SOCIAL	-0,44	-0,44	0,67	-0,65	0,51
Carrera_Anterior_No_Termina	0,49	0,14	0,16	3,02	0,00
IVE(2)_IVE RANGO 1	-0,44	-0,44	0,19	-2,34	0,02
IVE(2)_IVE RANGO 2	-0,26	-0,26	0,13	-2,10	0,04
IVE(2)_IVE RANGO 3	0,00	0,00	0,11	0,00	1,00
PSU_Lenguaje_FINAL	0,02	0,02	0,05	0,42	0,67
PSU_MATEMATICA_FINAL	0,30	0,30	0,06	4,83	0,00
REGION.Otras Regiones	-0,30	-0,30	0,12	-2,53	0,01
REGION.V y VI región	-0,33	-0,33	0,14	-2,37	0,02
TRAM._Sin Calificación Socioeconómica	0,88	0,88	0,42	2,11	0,03
TRAM._tramo 40	0,20	0,20	0,14	1,44	0,15
TRAM._tramo 100	1,01	1,01	0,43	2,33	0,02
TRAM._tramo 60	1,40	1,40	0,44	3,21	0,00
TRAM._tramo 70	1,79	1,79	0,44	4,03	0,00
TRAM._tramo 80	1,31	1,31	0,45	2,91	0,00
TRAM._tramo 90	1,05	1,05	0,42	2,49	0,01
Intercept	2,13	2,86	0,39	5,43	0,00

Anexo C

Evaluación Económica

C.1. Análisis Retrospectivo - Gratuidad

TABLA C.1: Análisis Retrospectivo - Gratuidad. Fuente: Elaboración propia.

Gratuidad	Total alumnos	Desertores Reales	% Deserción Real	Predicción Deserción	% Predicción Deserción	Predicción - Reales desertores	Variación
Sin gratuidad	3.578	667	19%	762	21%	95	3%
Con gratuidad	2.500	494	20%	514	21%	20	1%
Total	6.078	1.161	19%	1.276	21%	115	2%

C.2. Análisis Retrospectivo - Beca Vocación Profesor

TABLA C.2: Análisis Retrospectivo - Beca Vocación Profesor. Fuente: Elaboración propia.

Beca Vocación Profesor	Total alumnos	Desertores Reales	% Deserción Real	Predicción Deserción	% Predicción Deserción	Predicción - Reales desertores	Variación
Sin Vocación Profesor	5.988	1.143	19%	1.257	21%	114	2%
Con Vocación Profesor	90	18	20%	19	21%	1	1%
Total	6.078	1.161	19%	1.276	21%	115	2%

C.3. Análisis Retrospectivo - Tramo Socioeconómico

TABLA C.3: Análisis Retrospectivo - Tramo Socioeconómico. Fuente: Elaboración propia.

TRAMO	Total alumnos	Desertores Reales	% Deserción Real	Predicción Deserción	% Predicción Deserción	Reales-Predicción	Variación
tramo 40	2.094	431	21%	463	22%	32	2%
tramo 50	529	98	19%	133	25%	35	7%
tramo 60	391	62	16%	79	20%	17	4%
tramo 70	441	90	20%	43	10%	-47	-11%
tramo 80	380	81	21%	80	21%	-1	0%
tramo 90	725	136	19%	141	19%	5	1%
tramo 100	446	61	14%	89	20%	28	6%
Sin Calificación Socioeconómica	1.072	202	19%	248	23%	46	4%
Total	6.078	1.161	19%	1.276	21%	115	2%

C.4. Análisis Retrospectivo - Carrera

TABLA C.4: Análisis Retrospectivo - Carrera. Fuente: Elaboración propia.

Carrera	Total alumnos	Desertores Reales	% Deserción Real	Predicción Deserción	% Predicción Deserción	Predicción - Reales desertores	Variación
PROGRAMA ACADÉMICO DE BACHILLERATO	372	140	38%	353	95%	213	57%
DISEÑO	155	39	25%	127	82%	88	57%
FILOSOFÍA	87	26	30%	68	78%	42	48%
LENGUA Y LITERATURA HISPÁNICA	113	25	22%	63	56%	38	34%
INGENIERÍA EN ALIMENTOS	48	19	40%	44	92%	25	52%
ARQUITECTURA	245	67	27%	92	38%	25	10%
ENFERMERÍA	101	21	21%	45	45%	24	24%
OBSTETRICIA Y PUERICULTURA	97	16	16%	32	33%	16	16%
PEDAGOGIA	111	34	31%	50	45%	16	14%
ODONTOLOGÍA	107	17	16%	30	28%	13	12%
HISTORIA	109	25	23%	37	34%	12	11%
QUÍMICA	53	20	38%	31	58%	11	21%
TECNOLOGÍA MÉDICA	96	17	18%	27	28%	10	10%
PEDAGOGÍA EN EDUCACIÓN BÁSICA	38	4	11%	13	34%	9	24%
KINESIOLOGÍA	51	8	16%	16	31%	8	16%
INGENIERÍA AGRONÓMICA	166	28	17%	33	20%	5	3%
INGENIERÍA EN BIOTECNOLOGÍA MOLECULAR	42	12	29%	16	38%	4	10%
QYBAMBIENTAL	73	18	25%	22	30%	4	5%
ARTES PLÁSTICAS	100	37	37%	40	40%	3	3%
NUTRICIÓN Y DIETÉTICA	59	10	17%	8	14%	-2	-3%
LICENCIATURA	124	40	32%	36	29%	-4	-3%
BIOQUÍMICA	55	18	33%	14	25%	-4	-7%
FONOAUDILOGÍA	58	9	16%	3	5%	-6	-10%
LENGUA Y LITERATURA INGLESA	84	20	24%	13	15%	-7	-8%
PEDAGOGÍA EN EDUCACIÓN PARVULARIA	20	9	45%	1	5%	-8	-40%
INGENIERÍA EN RECURSOS NATURALES RENOVABLES	69	9	13%	0	0%	-9	-13%
QUÍMICA Y FARMACIA	159	25	16%	15	9%	-10	-6%
PSICOLOGÍA	96	12	13%	2	2%	-10	-10%
TRABAJO SOCIAL	65	10	15%		0%	-10	-15%
TERAPIA OCUPACIONAL	59	11	19%	1	2%	-10	-17%
ADMINISTRACIÓN PÚBLICA	114	11	10%	0	0%	-11	-10%
SOCIOLOGÍA	96	11	11%	0	0%	-11	-11%
ANTROPOLOGÍA-ARQUEOLOGÍA	107	12	11%	0	0%	-12	-11%
MEDICINA	184	13	7%	0	0%	-13	-7%
CINE Y TELEVISIÓN	72	19	26%	0	0%	-19	-26%
PERIODISMO	119	21	18%	1	1%	-20	-17%
INGENIERÍA FORESTAL	83	21	25%	1	1%	-20	-24%
CONTADOR AUDITOR	81	22	27%	2	2%	-20	-25%
INGENIERÍA EN INFORMACIÓN Y CONTROL DE GESTIÓN	145	23	16%	1	1%	-22	-15%
MUSICA, DANZA, ARTE Y TEATRO	140	41	29%	19	14%	-22	-16%
GEOGRAFÍA	80	24	30%	2	3%	-22	-28%
MEDICINA VETERINARIA	197	24	12%	0	0%	-24	-12%
DERECHO	398	42	11%	15	4%	-27	-7%
INGENIERÍA COMERCIAL	451	55	12%	1	0%	-54	-12%
INGENIERÍA Y CIENCIAS - PLAN COMÚN	798	75	9%	2	0%	-73	-9%

C.5. Análisis de Sensibilidad

TABLA C.5: Análisis de sensibilidad

Disminución Tasa de Deserción	F1	F2	F3	VAN	ΔVAN	TIR	PRC
1%	\$ -87.600.000	\$ -79.260.000	\$ -70.981.500	\$ -174.964.219	\$ -744.433.798	N.A.	No recupera el capital
2%	\$ -76.800.000	\$ -57.660.000	\$ -38.581.500	\$ -118.347.951	\$ -687.817.530	104%	No recupera el capital
3%	\$ -66.000.000	\$ -36.060.000	\$ -6.181.500	\$ -61.731.684	\$ -631.201.263	136%	No recupera el capital
4%	\$ -55.200.000	\$ -14.460.000	\$ 26.218.500	\$ -5.115.417	\$ -574.584.996	168%	No recupera el capital
5%	\$ -44.400.000	\$ 7.140.000	\$ 58.618.500	\$ 51.500.850	\$ -517.968.729	200%	3
6%	\$ -33.600.000	\$ 28.740.000	\$ 91.018.500	\$ 108.117.117	\$ -461.352.462	232%	3
7%	\$ -22.800.000	\$ 50.340.000	\$ 123.418.500	\$ 164.733.384	\$ -404.736.195	263%	3
8%	\$ -12.000.000	\$ 71.940.000	\$ 155.818.500	\$ 221.349.651	\$ -348.119.928	294%	2
9%	\$ -1.200.000	\$ 93.540.000	\$ 188.218.500	\$ 277.965.918	\$ -291.503.661	325%	2
10%	\$ 9.600.000	\$ 115.140.000	\$ 220.618.500	\$ 334.582.185	\$ -234.887.393	356%	2
11%	\$ 20.400.000	\$ 136.740.000	\$ 253.018.500	\$ 391.198.453	\$ -178.271.126	386%	2
12%	\$ 31.200.000	\$ 158.340.000	\$ 285.418.500	\$ 447.814.720	\$ -121.654.859	416%	2
13%	\$ 42.000.000	\$ 179.940.000	\$ 317.818.500	\$ 504.430.987	\$ -65.038.592	446%	1
14%	\$ 52.800.000	\$ 201.540.000	\$ 350.218.500	\$ 561.047.254	\$ -8.422.325	476%	1
15%	\$ 63.600.000	\$ 223.140.000	\$ 382.618.500	\$ 617.663.521	\$ 48.193.942	506%	1
16%	\$ 74.400.000	\$ 244.740.000	\$ 415.018.500	\$ 674.279.788	\$ 104.810.209	536%	1
17%	\$ 85.200.000	\$ 266.340.000	\$ 447.418.500	\$ 730.896.055	\$ 161.426.476	565%	1
18%	\$ 96.000.000	\$ 287.940.000	\$ 479.818.500	\$ 787.512.322	\$ 218.042.744	595%	1
19%	\$ 106.800.000	\$ 309.540.000	\$ 512.218.500	\$ 844.128.590	\$ 274.659.011	624%	1
20%	\$ 117.600.000	\$ 331.140.000	\$ 544.618.500	\$ 900.744.857	\$ 331.275.278	654%	1
21%	\$ 128.400.000	\$ 352.740.000	\$ 577.018.500	\$ 957.361.124	\$ 387.891.545	683%	1
22%	\$ 139.200.000	\$ 374.340.000	\$ 609.418.500	\$ 1.013.977.391	\$ 444.507.812	712%	1
23%	\$ 150.000.000	\$ 395.940.000	\$ 641.818.500	\$ 1.070.593.658	\$ 501.124.079	741%	1
24%	\$ 160.800.000	\$ 417.540.000	\$ 674.218.500	\$ 1.127.209.925	\$ 557.740.346	770%	1
25%	\$ 171.600.000	\$ 439.140.000	\$ 706.618.500	\$ 1.183.826.192	\$ 614.356.613	800%	1
26%	\$ 182.400.000	\$ 460.740.000	\$ 739.018.500	\$ 1.240.442.459	\$ 670.972.880	829%	1
27%	\$ 193.200.000	\$ 482.340.000	\$ 771.418.500	\$ 1.297.058.727	\$ 727.589.148	858%	1
28%	\$ 204.000.000	\$ 503.940.000	\$ 803.818.500	\$ 1.353.674.994	\$ 784.205.415	887%	1
29%	\$ 214.800.000	\$ 525.540.000	\$ 836.218.500	\$ 1.410.291.261	\$ 840.821.682	916%	1
30%	\$ 225.600.000	\$ 547.140.000	\$ 868.618.500	\$ 1.466.907.528	\$ 897.437.949	945%	1
31%	\$ 236.400.000	\$ 568.740.000	\$ 901.018.500	\$ 1.523.523.795	\$ 954.054.216	974%	1
32%	\$ 247.200.000	\$ 590.340.000	\$ 933.418.500	\$ 1.580.140.062	\$ 1.010.670.483	1002%	1
33%	\$ 258.000.000	\$ 611.940.000	\$ 965.818.500	\$ 1.636.756.329	\$ 1.067.286.750	1031%	1
34%	\$ 268.800.000	\$ 633.540.000	\$ 998.218.500	\$ 1.693.372.596	\$ 1.123.903.017	1060%	1
35%	\$ 279.600.000	\$ 655.140.000	\$ 1.030.618.500	\$ 1.749.988.863	\$ 1.180.519.285	1089%	1
36%	\$ 290.400.000	\$ 676.740.000	\$ 1.063.018.500	\$ 1.806.605.131	\$ 1.237.135.552	1118%	1
37%	\$ 301.200.000	\$ 698.340.000	\$ 1.095.418.500	\$ 1.863.221.398	\$ 1.293.751.819	1147%	1
38%	\$ 312.000.000	\$ 719.940.000	\$ 1.127.818.500	\$ 1.919.837.665	\$ 1.350.368.086	1176%	1
39%	\$ 322.800.000	\$ 741.540.000	\$ 1.160.218.500	\$ 1.976.453.932	\$ 1.406.984.353	1204%	1
40%	\$ 333.600.000	\$ 763.140.000	\$ 1.192.618.500	\$ 2.033.070.199	\$ 1.463.600.620	1233%	1
41%	\$ 344.400.000	\$ 784.740.000	\$ 1.225.018.500	\$ 2.089.686.466	\$ 1.520.216.887	1262%	1