



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**DISEÑO DE UNA ESTRATEGIA PARA DISMINUIR LAS DESERCIONES
TEMPRANAS DE LOS ESTUDIANTES DEL PREGRADO DE LA UNIVERSIDAD
DE CHILE**

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

MAURICIO GONZALO SERRANO VERGARA

PROFESOR GUÍA:
EZEQUIEL MUÑOZ KRSULOVIC

MIEMBROS DE LA COMISIÓN:
ANDRÉS VERGARA PRIETO
LUCIANO VILLARROEL PARRA
RICHARD WEBER HAAS

SANTIAGO DE CHILE
2017

RESUMEN EJECUTIVO

La deserción conlleva costos altísimos a los estudiantes, sobre todo aquellos que sus familias no disponen de los recursos y debe solicitar créditos, quedando endeudados y debiendo pagar por no lograr obtener un grado o título. Por otro lado, también genera costos a la institución, ya que se determinan cupos de ingreso a cada Programa de Formación, los cuales se utilizan en su totalidad y se espera que los estudiantes permanezcan hasta lograr su graduación y/o titulación.

El objetivo del proyecto es diseñar una estrategia para disminuir las deserciones tempranas de los estudiantes de pregrado de la Universidad de Chile, a través de la creación y aplicación de un modelo que permita detectar tempranamente a los posibles desertores de los estudiantes de pregrado.

El presente proyecto contempla estudiantes que hayan ingresado sólo a través de la Prueba de Selección Universitaria y las deserciones que ocurren en el primer año, realizándose un piloto con el Programa de Formación de Arquitectura.

Se utilizan las metodologías de Ingeniería de Negocios y Proceso de Extracción del Conocimiento (KDD) para obtener un modelo el cual debe ser ejecutado los primeros días del año, cuando se obtiene la matrícula de los estudiantes nuevos, luego se publica el listado de estudiantes que podrían desertar en el Sistema de Información para la Gestión Académica (SIGA), donde los Profesionales contratados para realizar el seguimiento y acompañamiento de los estudiantes se informan, para luego llevar a cabo sus actividades con los estudiantes.

En el presente proyecto se ha obtenido un modelo de redes neuronales, el cual considera las variables de Educación Secundaria: Grupo Dependencia¹ y Nota de Enseñanza Media; PSU: Puntaje de Prueba de Lenguaje y Comunicación, puntaje de Prueba de Matemáticas, Puntaje Ponderado y Preferencia; Socioeconómico: Ingreso Bruto y si tiene o no beca; Personal: Género; y Grupo Familiar: Número de integrantes del grupo familiar y si están vivos sus padres. El modelo tiene un resultado sobresaliente, ya que predice al 74,32% de los estudiantes que van a desertar en su primer año de carrera.

Se realizó un análisis de sensibilidad en la evaluación económica del proyecto, lo cual permite apreciar que los beneficios sobre los costos son notablemente mayores, haciendo que el proyecto sea muy rentable.

Se han obtenido beneficios que no estaban presupuestados como la creación del portal SIGA, los formularios de registro con posteriores reportes de análisis y un comparador de modelos, que permite realizar en menor tiempo todas las combinaciones de modelos evaluados.

¹ Grupo Dependencia: Tipo Establecimiento Educacional: Municipal, Subvencionado y Particular.

*A mi querida Mercedes
por su amor y apoyo incondicional
durante este largo proceso*

TABLA DE CONTENIDO

CAPÍTULO 1: INTRODUCCIÓN Y CONTEXTO	1
1.1 ANTECEDENTES GENERALES DEL SISTEMA DE EDUCACIÓN SUPERIOR	1
1.2 DESCRIPCIÓN GENERAL DE LA UNIVERSIDAD DE CHILE.....	1
1.2.1 <i>Misión:</i>	3
1.2.2 <i>Visión de futuro:</i>	3
1.3 PROBLEMA U OPORTUNIDAD IDENTIFICADA.....	4
1.4 OBJETIVOS Y RESULTADOS ESPERADOS DEL PROYECTO.....	5
1.4.1 <i>Objetivo General</i>	5
1.4.2 <i>Objetivos Específicos</i>	5
1.4.3 <i>Resultados Esperados</i>	5
1.5 ALCANCE.....	6
1.6 RIESGOS POTENCIALES.....	6
CAPÍTULO 2: MARCO TEÓRICO	8
2.1 METODOLOGÍA DE INGENIERÍA DE NEGOCIOS.....	8
2.1.1 <i>Planteamiento estratégico</i>	8
2.1.2 <i>Definición del Modelo de Negocios</i>	8
2.1.3 <i>Diseño de la Arquitectura de Procesos</i>	8
2.1.4 <i>Diseño de los Procesos</i>	8
2.1.5 <i>Diseño de la Aplicación de Apoyo</i>	9
2.1.6 <i>Construcción e Implementación</i>	9
2.2 LÓGICA DE NEGOCIOS.....	9
2.2.1 <i>Proceso de Extracción del Conocimiento</i>	9
2.2.2 <i>Modelos de Minería de Datos</i>	10
2.2.3 <i>Métodos No Supervisados</i>	11
2.2.4 <i>Métodos Supervisados</i>	11
2.2.5 <i>Árbol de decisiones</i>	12
2.2.6 <i>Regresión Logística</i>	13
2.3 DESERCIÓN UNIVERSITARIA.....	14
CAPÍTULO 3: PLANTEAMIENTO ESTRATÉGICO Y MODELO DE NEGOCIOS	15
3.1 POSICIONAMIENTO ESTRATÉGICO	15
3.2 MAPA ESTRATÉGICO	15
3.3 MODELO DE NEGOCIOS	16
3.3.1 <i>Propuesta de Valor al Cliente</i>	17
3.3.2 <i>Recursos claves</i>	17
3.3.3 <i>Procesos claves</i>	17
3.3.4 <i>Fórmula de utilidades</i>	18
3.3.5 <i>Amenazas de nuevos competidores</i>	18
3.3.6 <i>Amenazas de productos y servicios sustitutos</i>	18
3.3.7 <i>Poder de Negociación de clientes</i>	18
3.3.8 <i>Poder de Negociación de Proveedores</i>	19
3.3.9 <i>Rivalidad entre Competidores</i>	19
3.4 ANÁLISIS FODA.....	19
3.4.1 <i>Fortalezas</i>	19
3.4.2 <i>Debilidades</i>	19
3.4.3 <i>Oportunidades</i>	19
3.4.4 <i>Amenazas</i>	19
CAPÍTULO 4: ANÁLISIS SITUACIÓN ACTUAL	20
4.1 MACROPROCESO 1: CADENA DE VALOR.	20

4.2	MACROPROCESO 2: DESARROLLO DE NUEVAS CAPACIDADES	20
4.3	MACROPROCESO 3: PLANIFICACIÓN DEL NEGOCIO.....	20
4.4	MACROPROCESO 4: PROCESOS DE APOYO.....	20
4.5	SERVICIOS ACADÉMICOS DE PREGRADO.....	21
4.5.1	<i>Administración de relación con el estudiante</i>	22
4.5.2	<i>Administración de relación con el académico</i>	22
4.5.3	<i>Gestión de Servicios Académicos</i>	23
4.5.4	<i>Servicios Académicos</i>	23
4.6	GESTIÓN DE SERVICIOS ACADÉMICOS	23
4.6.1	<i>Implementación de nuevos programas</i>	24
4.6.2	<i>Planificación y control de servicios académicos</i>	24
4.6.3	<i>Decidir entrega de procesos académicos</i>	24
4.7	DIAGNÓSTICO DE LA SITUACIÓN ACTUAL.....	24
4.8	CUANTIFICACIÓN DEL PROBLEMA U OPORTUNIDAD.....	25
4.8.1	<i>Deserción en la Educación Superior</i>	25
4.8.2	<i>Deserción en la Universidad de Chile</i>	25
CAPÍTULO 5: PROPUESTA DE DISEÑO DE PROCESOS		27
5.1	DIRECCIONES DE CAMBIO Y ALCANCE	27
5.2	DISEÑO DETALLADO DE PROCESOS TO BE	29
5.3	DISEÑO DE LÓGICA DE NEGOCIOS	30
5.3.1	<i>Selección de datos:</i>	31
5.3.2	<i>Preprocesamiento de Datos:</i>	33
5.3.3	<i>Transformación:</i>	33
5.3.4	<i>Minería de Datos:</i>	34
5.3.5	<i>Interpretación y Evaluación:</i>	35
CAPÍTULO 6: PROPUESTA DE APOYO TECNOLÓGICO		38
6.1	ESPECIFICACIÓN DE REQUERIMIENTOS.....	38
6.1.1	<i>Requerimientos Funcionales</i>	38
6.1.2	<i>Requerimientos No Funcionales</i>	38
6.2	ARQUITECTURA TECNOLÓGICA.....	38
6.2.1	<i>SQL Navigator</i>	39
6.2.2	<i>Trello</i>	39
6.2.3	<i>Tableau Desktop</i>	40
6.2.4	<i>Google Sites</i>	40
6.2.5	<i>Integraciones con otros sistemas</i>	40
6.3	DISEÑO DE LA APLICACIÓN.....	41
6.3.1	<i>Casos de Uso</i>	41
CAPÍTULO 7: EVALUACIÓN DEL PROYECTO		43
7.1	BENEFICIOS COLATERALES DEL PROYECTO.....	43
7.2	DEFINICIÓN DE BENEFICIOS Y COSTOS.....	43
7.2.1	<i>Inversión Inicial del Proyecto</i>	44
7.2.2	<i>Costos de mantención</i>	44
7.2.3	<i>Costos Directos por Profesionales</i>	45
7.3	FLUJO DE CAJA.....	46
7.4	ANÁLISIS DE SENSIBILIDAD.....	48
7.4.1	<i>Análisis de Sensibilidad 1</i>	48
7.4.2	<i>Análisis de Sensibilidad 2</i>	49
CAPÍTULO 8: CONCLUSIONES.....		50
8.1	CONCLUSIONES SOBRE METODOLOGÍA	50
8.2	CONCLUSIONES SOBRE LA EJECUCIÓN DEL PROYECTO	50
8.3	CONCLUSIONES SOBRE RESULTADOS DEL PROYECTO.....	51

8.4	TRABAJO FUTUROS	52
CAPÍTULO 9: BIBLIOGRAFÍA.....		54
CAPÍTULO 10: ANEXOS		57
10.1	ANEXO 1: CASOS DE USO	57
10.2	ANEXO 2: CREACIÓN DE DATAMART	60
10.2.1	<i>Información de Becas</i>	<i>60</i>
10.2.2	<i>Información de Quintiles</i>	<i>61</i>
10.2.3	<i>Información de Matrícula</i>	<i>61</i>
10.2.4	<i>Información de Situaciones Académicas</i>	<i>62</i>
10.2.5	<i>Archivo B: Inscritos, datos socioeconómicos y direcciones.....</i>	<i>63</i>
10.2.6	<i>Archivo E: Postulaciones efectivas</i>	<i>65</i>
10.3	ANEXO 3: COMPARADOR DE MODELOS	66

ÍNDICE DE ILUSTRACIONES

Ilustración 1:	Clasificación Universidades Chilenas. Fuente: www.mifuturo.cl	1
Ilustración 2:	Proceso de Extracción del Conocimiento (KDD).	9
Ilustración 3:	Modelos de Data Mining.	10
Ilustración 4:	Detalles de una neurona.	11
Ilustración 5:	Hiperplano SVM.....	12
Ilustración 6:	Ejemplo Árbol de Decisión. Fuente: Elaboración propia.....	13
Ilustración 7:	Función Logística.....	13
Ilustración 8:	Función Regresión Logística.....	13
Ilustración 9:	Matriz topológica de la sistematización de la literatura relacionada con la deserción estudiantil. Fuente: Christian Díaz Peralta	14
Ilustración 10:	Posicionamiento Estratégico. Fuente: Elaboración Propia.	15
Ilustración 11:	Mapa Estratégico. Fuente: Elaboración Propia.	16
Ilustración 12:	Modelo de Negocios. Fuente: Elaboración Propia.....	16
Ilustración 13:	Macroprocesos Universidad de Chile. Fuente: Elaboración Propia.	21
Ilustración 14:	Servicio académicos de Pregrado. Fuente: Elaboración Propia.	22
Ilustración 15:	Gestión de servicios académicos. Fuente: Elaboración Propia.	23
Ilustración 16:	Tasa de Deserción. Fuente: Anuario U. de Chile. Síntesis 2006-2013.....	26
Ilustración 17:	Generación de listas de estudiantes.	29
Ilustración 18:	Proceso en RapidMiner. Fuente: Elaboración Propia.....	30
Ilustración 19:	Preprocesamiento de datos. Fuente: Elaboración Propia	33
Ilustración 20:	Transformación. Fuente: Elaboración Propia	33
Ilustración 21:	Filtros. Fuente: Elaboración Propia	33
Ilustración 22:	Validación. Fuente: Elaboración Propia	34
Ilustración 23:	Predicción. Fuente: Elaboración Propia	34
Ilustración 24:	VARIABLES utilizadas en el modelo. Fuente: Elaboración Propia	35
Ilustración 25:	Matriz de Confusión. Fuente: Elaboración Propia	35
Ilustración 26:	Casos de prueba, Propuesta 1. Fuente: Elaboración Propia.....	36
Ilustración 27:	Propuesta 1. Fuente: Elaboración Propia.....	36
Ilustración 28:	Casos de Prueba, Propuesta 2. Fuente: Elaboración Propia	37
Ilustración 29:	Propuesta 2. Fuente: Elaboración Propia.....	37
Ilustración 30:	Arquitectura Tecnológica. Fuente: Elaboración propia.	39
Ilustración 31:	Arquitectura Tecnológica con Casos de Usos. Fuente: Elaboración propia....	41
Ilustración 32:	Casos de Usos. Fuente: Elaboración propia.	41
Ilustración 33:	Estudiantes que desertan en Pregrado por año. Fuente: Elaboración propia. 43	43

Ilustración 34: Datamart. Fuente: Elaboración propia.....	60
Ilustración 35: Comparador de Modelos. Fuente: Elaboración propia.	66

ÍNDICE DE TABLAS

Tabla 1: Riesgos Potenciales.....	6
Tabla 2: Variables de Estructura de Empresa y Mercado	27
Tabla 3: Variables de Anticipación	28
Tabla 4: Variables de Coordinación	28
Tabla 5: Variables de Prácticas de Trabajo.....	28
Tabla 6: Variables de Integración de Procesos Conexos	29
Tabla 7: Variables de Mantenimiento Consolidada de Estado	29
Tabla 8: Fuentes de datos con sus respectivos campos	31
Tabla 9: Balanceo por año.	32
Tabla 10: Inversión Inicial del Proyecto.....	44
Tabla 11: Costos de mantención	44
Tabla 12: Costos Directos por Profesionales	45
Tabla 13: Tasa Social de Descuento. Fuente: Ministerio de Desarrollo Social	46
Tabla 14: Flujo de Caja del Proyecto	47
Tabla 15: Flujo de Caja del Análisis de Sensibilidad 1	48
Tabla 16: Flujo de Caja del Análisis de Sensibilidad 2	49
Tabla 17: Caso de Uso 1	57
Tabla 18: Caso de Uso 2	58
Tabla 19: Caso de Uso 3	58
Tabla 20: Caso de Uso 4	59
Tabla 21: Información de becas.....	60
Tabla 22: Información de Quintil	61
Tabla 23: Información de Matrícula.....	61
Tabla 24: Información de Situaciones Académicas.....	62
Tabla 25: Información de Archivo B	63
Tabla 26: Información de Archivo E	65

CAPÍTULO 1: INTRODUCCIÓN Y CONTEXTO

1.1 Antecedentes Generales del Sistema de Educación Superior

El sistema de Educación Superior está compuesto por 59 Universidades, 43 Institutos Profesionales y 54 Centros de Formación Técnica.

Las universidades se pueden clasificar como se muestra a continuación:

- 25 pertenecen al Consejo de Rectores de las Universidades Chilenas (CRUCH).
 - 16 Universidades Estatales, las cuales han sido creadas por ley, pertenecen al Estado de Chile.
 - 9 Universidades Particulares con Aporte del Estado, las cuales son universidades privadas creadas antes de 1980 o derivan de ellas.
- 34 Universidades Privadas, las cuales son todas las universidades creadas después de 1980, a partir de lo establecido en el DFL 1 de 1980 o la Ley 18.962 de 1990.

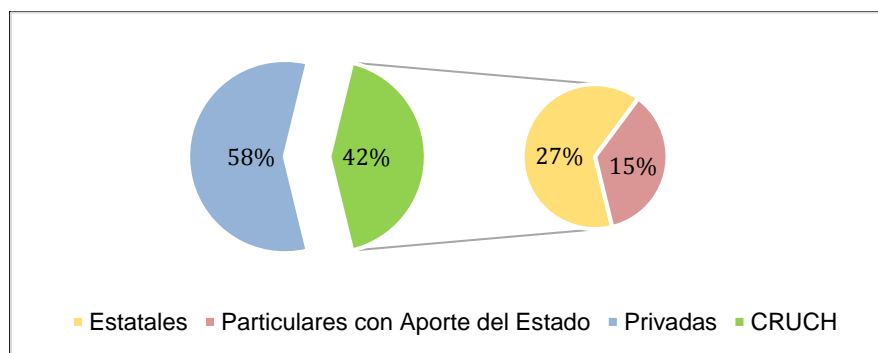


Ilustración 1: Clasificación Universidades Chilenas. Fuente: www.mifuturo.cl

La matrícula de primer año en el 2016 se distribuye en un 47,9% para las universidades del CRUCH y un 52,01% en universidades privadas. La matrícula total para las universidades es de aproximadamente unos 720 mil estudiantes.

Los grados otorgados por las universidades son bachiller, licenciatura, magíster y doctor y sus duraciones van desde los 2 años de bachiller, licenciatura 4 años, magíster 2 años y doctor 4 años.

1.2 Descripción General de la Universidad de Chile

La Universidad de Chile es la institución de educación superior más antigua del país, fue fundada en 1842, su primer Rector fue Andrés Bello, ha desarrollado políticas de acción destinadas a resolver los problemas nacionales y regionales que afectan nuestro país. Se define a sí misma como garante de la cultura clásica, humanista y

secular. Es así como su carácter nacional y pública, lo asume con compromiso y vocación de excelencia en la formación de personas, aportando además al desarrollo espiritual y material de la Nación, construyendo liderazgo en el desarrollo innovador de las ciencias y las tecnologías, las humanidades y las artes, a través de sus funciones de docencia, creación y extensión, con especial énfasis en la investigación y el postgrado.

En esta casa de estudios se han formado 20 de los 32 presidentes de la República, siendo también la universidad con la mayor cantidad de académicos y académicas con Premios Nacionales vigentes en su actividad universitaria, y nuestros 2 premios Nobel en Literatura Gabriela Mistral y Pablo Neruda, lo que convierte a la Universidad de Chile en la institución más importante a nivel superior en el país.

Su permanente compromiso con la calidad y excelencia se manifiesta en el nivel máximo de acreditación otorgado por la Comisión Nacional de Acreditación (CNA) en todas sus áreas y por un periodo de 7 años desde el 21 de diciembre del año 2011 hasta el 21 de diciembre del año 2018, de esta forma se confirma su indiscutido liderazgo como institución pública de excelencia en la educación superior en el país. La comunidad de la Universidad de Chile está compuesta por 3.637 académicos, 30.755 estudiantes en programas de Pregrado, 1.441 de Doctorado, 6.652 estudiantes de Magíster y 1.239 estudiantes de postítulo y profesionales especialistas, junto a 8.680 funcionarios de personal de colaboración en las unidades académicas, organismos centrales y Hospital Clínico.

Actualmente, la Universidad ofrece 69 programas de estudio, de los cuales 54 conducen a títulos profesionales y 15 a licenciaturas terminales, además de un programa de Bachillerato. Asimismo, cuenta con el sistema de postgrado más grande y complejo del país, compuesto por 38 programas de doctorado, 137 programas de magíster, 76 programas de postítulo y 14 cursos de especialización de postítulo. La Universidad se organiza en 14 Facultades, 4 Institutos, el Hospital Clínico de la Universidad de Chile José Joaquín Aguirre, y el Liceo Experimental Manuel de Salas. Las Facultades son: Artes; Arquitectura y Urbanismo; Ciencias; Ciencias Agronómicas; Economía y Negocios; Ciencias Físicas y Matemáticas; Ciencias Forestales y de la Conservación de la Naturaleza; Ciencias Químicas y Farmacéuticas; Ciencias Sociales; Ciencias Veterinarias y Pecuarias; Derecho; Filosofía y Humanidades; Medicina; Odontología. Los Institutos son: Instituto de Asuntos Públicos; Instituto de Estudios Internacionales; Instituto de la Comunicación e Imagen, e Instituto de Nutrición y Tecnología de los Alimentos.

Referente a su infraestructura, la Universidad cuenta con 167 hectáreas de terreno urbano, 683 mil m² de superficie construida en uso, 9.375 hectáreas de superficie agrícola, más de 3.800 computadores para uso de estudiantes, cerca de 3 millones de volúmenes en 48 bibliotecas y una biblioteca Digital, otorga acceso directo o referencial a más de 722.000 documentos: libros, tesis, revistas y artículos, entre otros. Incluye más de 42.000 objetos digitalizados de gran valor histórico como mapas, manuscritos, partituras, artesanías, fotografías, audios y películas. Además una gran institucionalidad cultural compuesta de Orquesta Sinfónica, Ballet Nacional, Coro de la Universidad de Chile, Teatro Nacional, Museo de Arte Contemporáneo, Museo de Arte Popular Americano, Archivo Andrés Bello, entre otros, además de las instalaciones culturales que se encuentran adscritas a las Facultades e Institutos. Asimismo, por su importancia como apoyo al proceso educativo, cuenta con modernas instalaciones para prácticas deportivas en los distintos campus que componen la institución.

1.2.1 Misión:

- La generación, desarrollo, integración y comunicación del saber en todas las áreas del conocimiento y dominios de la cultura, constituyen la misión y el fundamento de las actividades de la Universidad, conforman la complejidad de su quehacer y orientan la educación que ella imparte.
- La Universidad asume con vocación de excelencia la formación de personas y la contribución al desarrollo espiritual y material de la Nación. Cumple su misión a través de las funciones de docencia, investigación y creación en las ciencias y las tecnologías, las humanidades y las artes, y de extensión del conocimiento y la cultura en toda su amplitud. Procura ejercer estas funciones con el más alto nivel de exigencia.
- Es responsabilidad de la Universidad contribuir con el desarrollo del patrimonio cultural y la identidad nacionales y con el perfeccionamiento del sistema educacional del país.

1.2.2 Visión de futuro:

- Desarrollo consistente con la misión histórica, la naturaleza estatal y pública y el compromiso nacional de la universidad.
- Parámetros internacionales de excelencia académica.
- Respuesta creativa y eficaz a las condiciones y desafíos que plantea la globalización y la inserción del país en el orden mundial.

1.3 Problema u Oportunidad Identificada

La Universidad de Chile, ha identificado a la gestión de los procesos de la docencia como una de las prioridades de la Vicerrectoría de Asuntos Académicos (VAA). Siendo la docencia una de las funciones importantes de la misión institucional, priorizada por el Plan de Mejoramiento Institucional (PMI), la VAA se planteó el desafío de concretar etapas superiores de perfeccionamiento del modelo de gestión de la docencia, que incluyera el paradigma de gestión focalizada en el estudiante, en coherencia con el modelo educativo.

Se ha dado importancia a la revisión y seguimiento de los diferentes procesos involucrados, a la obtención automatizada de los resultados de los procesos, a su organización en plataformas de gestión, a la aplicación de indicadores y a la preparación de reportes de gestión que, al ser analizados periódicamente, puedan dar como resultado una mejor y fundamentada toma de decisiones para el aseguramiento de la calidad.

La implementación de estos procesos y el registro en los sistemas corporativos de la institución de cada una de las actividades y actores que en ellos participan, permite la configuración de bases de datos desde las cuales es factible extraer antecedentes concretos y validados para construir informes con estadísticas e indicadores de los resultados de esa gestión.

Como uno de los objetivos principales del PMI es descentralizar las acciones que ayuden a mejorar los indicadores solicitados por el Ministerio de Educación, centrándose principalmente a la deserción estudiantil, y con esto permita que la toma de decisiones se realice en las distintas Unidades Académicas con un lineamiento dirigido a nivel central. Para lograr lo anterior fue necesaria la creación de la Unidad Información para la Gestión Académica, la cual es la encargada de recopilar la información y construir reportes tanto para el Pregrado como para Postgrado y Postítulo. El análisis y reflexión sobre estos informes permitirá la elaboración de planes de mejoramiento tanto de los procesos como de los resultados de la gestión de la docencia, en el marco de políticas de desarrollo académico institucionales.

La deserción estudiantil conlleva costos altísimos a los estudiantes, sobre todo aquellos que sus familias no disponen de los recursos y debe solicitar créditos, quedando endeudados y debiendo pagar por no lograr obtener un grado o título. Por otro lado, también genera costos a la institución, ya que se determinan cupos de ingreso a cada Programa de Formación, los cuales son llenados y se espera que los estudiantes permanezcan hasta lograr su graduación y/o titulación.

Previo al proyecto, se han creado centros de aprendizaje en el Campus Sur y en la Facultad de Ciencias Químicas y Farmacéuticas, quienes aplican criterios distintos para seleccionar a los estudiantes que podrían desertar.

El proyecto ha generado grandes expectativas en las autoridades, debido a que los beneficios estudiantiles y económicos que permitirán la autosustentabilidad de los centros de aprendizajes.

1.4 Objetivos y Resultados Esperados del Proyecto

1.4.1 Objetivo General

Diseñar una estrategia para disminuir las deserciones tempranas de los estudiantes de pregrado de la Universidad de Chile, a través de la creación y aplicación de un modelo que permita detectar tempranamente a los posibles desertores de los estudiantes de pregrado, caracterizándolos por medio de una serie de atributos personales, socioeconómicos, grupo familiar y de rendimiento escolar.

1.4.2 Objetivos Específicos

- OE1: Recopilar y construir un repositorio de información que albergue toda la necesaria para la creación de informes de Gestión Académica.
- OE2: Encontrar el mejor modelo de predicción de deserción, utilizando distintos algoritmos y técnicas.
- OE3: Identificar los factores de riesgo que inciden en la deserción del estudiante, esto es encontrar las variables independientes que inciden en la predicción.
- OE4: Construir una interfaz de consulta de estudiantes que podrían desertar a través de una plataforma online.

1.4.3 Resultados Esperados

Tener un repositorio de información único que contenga la información necesaria para realizar la analítica para la clasificación de los estudiantes. Este repositorio se llenará inicialmente con información desde el 2004 en adelante y se actualizará a fines de enero de cada año.

Contrataciones de personal de acompañamiento, en primera instancia en la carrera de Arquitectura y a partir del segundo semestre comenzar con todo el campus Andrés Bello y Campus Norte.

Contar con un listado de estudiantes que podrían desertar, el cual debe estar publicado en línea, para que pueda ser consultado en todo momento por personal de las Secretarías de Estudios, el cual se actualizará todos los meses de enero.

1.5 Alcance

El presente proyecto aborda la deserción de todos los estudiantes de Pregrado de la Universidad de Chile de los años 2004 al año vigente, donde su vía de ingreso ha sido la Prueba de Selección Universitaria (PSU), por lo cual no considera las siguientes vías de ingreso: Beca Excelencia Académica (BEA), Ciclo Básico de Artes, Concurso Trabajadores, Convenio Cerro Navia, Convenio Étnico, Convenio Trabajadores FF.AA., Cupo Extranjeros, Deportista Destacado, Egresado o Titulado, Equidad de Género, Escuela de Desarrollo de Talentos, Estudiantes Ciegos, Etapa Básica, Exonerados, Ingreso Postgrado, Movilidad Estudiantil, Movilidad Estudiantil Internacional, Otras Instituciones Artísticas, Otros Ingresos Especiales, PAA, Prog. Complem. Lic. Area de la Salud, SIPEE, Transferencias Externas e Internas.

El presente estudio considera la deserción que se produce en el primer año, ya que es la más significativa, donde el impacto del primer año asciende entre un 10% y 20% de la matrícula de la cohorte y el piloto es realizado con la carrera de Arquitectura.

1.6 Riesgos Potenciales

A continuación, se exponen 15 riesgos potenciales del proyecto:

Tabla 1: Riesgos Potenciales

Tipo	Probabilidad	Impacto	Plan de Mitigación
1. Inconsistencias en la información	Alta	Aumento de tiempos de procesamiento, mal diagnóstico y resultados	Exigir a la parte técnica que realice correcciones de esto.
2. Reforma estudiantil	Alta	Educación gratuita podría aumentar la tasa de deserción.	Generar un programa de apoyo para los estudiantes con gratuidad.
3. Nuevas vías de ingreso	Alta	Baja certeza del comportamiento de permanencia	Programa de reforzamiento para nivelar conocimiento
4. Aumento de programas de formación técnica	Alta	Disminución del interés por carreras universitarias	Buscar instancias de acercar la universidad a los establecimientos de educación secundaria.

5. Desempleabilidad	Media	Aumento de postergaciones y/o renunciaciones	Buscar fuentes de financiamiento alternativo o facilidades de pago
6. Incorrecta selección de variables	Media	Un modelo que no responda a los objetivos	Contrastar resultados con los antiguos modelos.
7. Gestión del cambio	Media	Resistencia a las nuevas mejoras	Capacitación y acompañamiento del proceso
8. Movilización de estudiantes en unidades académicas	Media	Aumento en deserciones por cambio de institución	Reubicar clases en otras instalaciones.
9. Creación de nuevas universidades	Media	Menor número de postulaciones	Campaña comunicacional del prestigio de la universidad
10. Cambios de gobierno	Media	Cambios en legislaciones	Generar mesa de trabajo
11. Catástrofes Naturales	Media	Deterioro en infraestructura	Reubicación en otras instalaciones
12. Cambios en la PSU	Media	Baja certeza del comportamiento de permanencia	Generar mesa de trabajo
13. Acreditación de otras universidades	Media	Menor número de postulaciones	Reafirmación del prestigio de la universidad con estadística de empleabilidad de los egresados
14. Crisis Financiera	Baja	Aumento de postergaciones y/o renunciaciones	Buscar fuentes de financiamiento alternativo
15. Cambio de Autoridades de la Universidad	Baja	Puede cerrarse el proyecto	Presentar el proyecto a las nuevas autoridades.

CAPÍTULO 2: MARCO TEÓRICO

2.1 Metodología de Ingeniería de Negocios

Barros (2004) formuló esta metodología, la cual es la columna vertebral del MBE y tiene las siguientes etapas:

2.1.1 *Planteamiento estratégico*

En esta etapa se utilizan primeramente el Modelo Delta de Hax, el cual permite ver donde debiera estar posicionada la estrategia del proyecto, estos podrían ser Lock-in Sistémico, cuando el cliente no tiene ninguna otra opción; Mejor Producto, ya sea compitiendo por precio o calidad; y finalmente Solución Total al Cliente donde el centro es el servicio o producto que mantenga al cliente con altos niveles de satisfacción. Luego se utiliza el Balanced Score Card, lo cual permite realizar un mapa de la estrategia, donde su base e inicio es realizar actividades y/o mejoras en la perspectiva de aprendizaje y crecimiento, esto debiese influir en los procesos internos, donde se genera un impacto que afecta al cliente y finalmente y no menos importante, ya que está en la cúspide es la perspectiva financiera.

2.1.2 *Definición del Modelo de Negocios*

El modelo de negocios está compuesto por recursos claves como recursos humanos, tecnología y marcas; procesos claves como diseño de productos, marketing y capacitación, por la propuesta de valor al cliente, que contiene los clientes objetivos y el trabajo que se realizará; y finalmente la fórmula de utilidad la cual debe indicar cuánto dinero se puede ganar, cómo se asignan los costos y cuán rápido se usan los recursos para obtener el volumen deseado.

2.1.3 *Diseño de la Arquitectura de Procesos*

En esta etapa se busca determinar las relaciones que existen entre los macroprocesos de una empresa. Para ello se utilizan los patrones, esto se debe hacer con la situación actual y con la propuesta de rediseño.

2.1.4 *Diseño de los Procesos*

Es una etapa más detallada del punto anterior, acá se realizan los modelos en BPMN², describiendo muy bien los roles con las actividades que realizan.

² BPMN: Business Process Model and Notation

2.1.5 *Diseño de la Aplicación de Apoyo*

Se detalla cada componente de la Arquitectura TI, con sus respectivos Casos de Uso y pantallas de tal modo de que se puede apreciar cómo funcionará la solución propuesta.

2.1.6 *Construcción e Implementación*

Se desarrollan las aplicaciones diseñadas previamente y se implementan tanto los procesos como las aplicaciones TI de apoyo.

2.2 *Lógica de Negocios*

2.2.1 *Proceso de Extracción del Conocimiento*

Uno de los objetivos específicos del proyecto es obtener un modelo de predicción de deserción, el cual surge de la aplicación de un algoritmo, que se debe comprobar su consistencia y asegurar que es el mejor resultado, para esto es necesario aplicar la metodología Knowledge Discovery in Databases (KDD), la que permite descubrir patrones de comportamiento de los datos y extraer conocimiento de ellos.

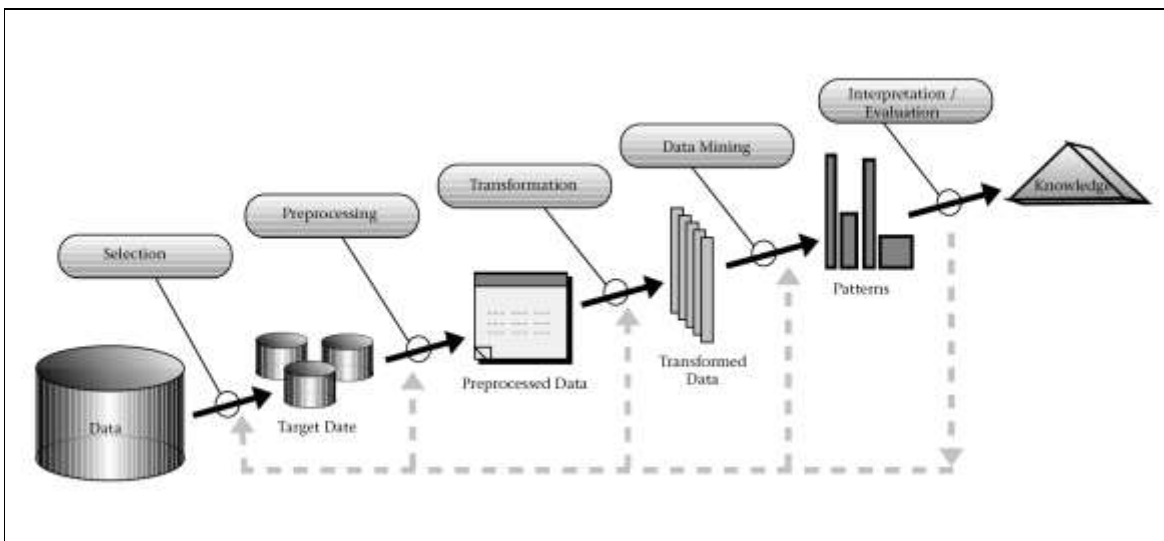


Ilustración 2: Proceso de Extracción del Conocimiento (KDD).

A continuación, se explican los pasos del KDD:

- **Selección:** establecer los objetivos del negocio, los objetivos de aplicar el proceso y establecer los criterios de éxito, para lo cual se obtienen los datos desde distintas fuentes para luego seleccionar aquellos que permitan responder dichos objetivos. Se exploran los datos, verificando la calidad y estableciendo rangos.

- **Pre-Procesamiento:** se preparan y limpian los datos seleccionados anteriormente, utilizando variadas estrategias para los datos inconsistentes, faltantes o fuera de rango, con el fin de dejarlos de una forma manejable para fases posteriores.
- **Transformación:** reducción y proyección de datos, encontrando características útiles, con estos métodos se puede reducir el número de variables o se pueden encontrar representaciones invariables.
- **Minería de datos:** buscar patrones de interés en una representación particular, incluyendo reglas como por ejemplo de clasificación, árboles y segmentación. Para esto se debe alinear el objetivo del proceso con un método particular de minería de datos, tal como, clasificación, regresión, segmentación o predicción. Además, se debe hacer un análisis exploratorio, selección del modelo e hipótesis, depende de lo que se haya elegido en el punto anterior, por ejemplo, para el caso de predicción, se puede comprobar los Árboles de Decisión, Support Vector Machines, Regresión Logística y Redes Neuronales.
- **Interpretación y Evaluación:** se interpretan los patrones obtenidos, y en caso de quedar alguna duda se puede volver a cualquier punto anterior. Además, actuar sobre el conocimiento descubierto, ya sea publicando los resultados o incorporando el conocimiento en otro sistema para acciones futuras.

2.2.2 Modelos de Minería de Datos

Existen variadas formas de aplicar la minería de datos, esto dependerá del proyecto, puede ser para explicar el pasado o predecir el futuro.

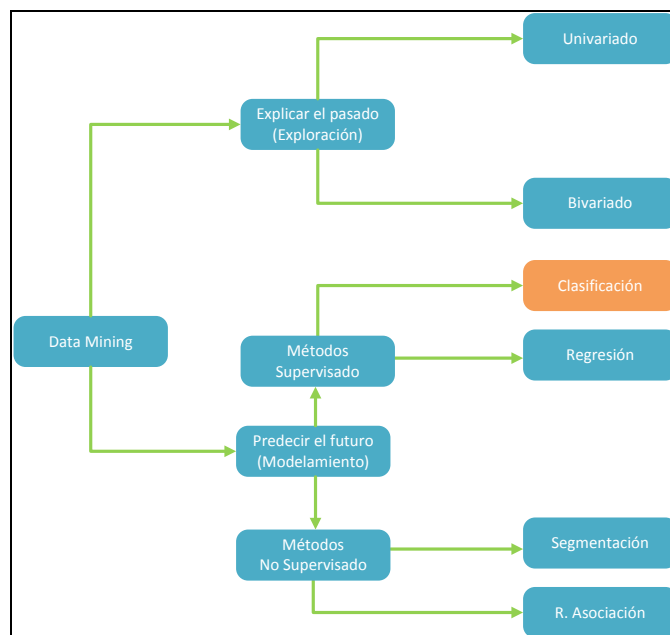


Ilustración 3: Modelos de Data Mining.

El presente proyecto utiliza modelos de clasificación, los cuales se explicarán a continuación.

2.2.3 Métodos No Supervisados

Los modelos no supervisados son aquellos que no tienen una variable objetivo, se utiliza principalmente para segmentación y reglas de asociación, un ejemplo es buscar patrones de datos que permitan caracterizar el comportamiento de clientes.

2.2.4 Métodos Supervisados

Los modelos supervisados son aquellos que tienen una variable objetivo (dependiente) que permite etiquetar, donde todas las otras variables (independientes) explican el comportamiento en función del objetivo formulado. Estos modelos se pueden separar en dos grupos, de clasificación y de regresión, el proyecto se centrará en la primera de estas.

2.2.4.1 Redes Neuronales

Las Redes Neuronales son un modelo supervisado que se genera de la base del funcionamiento de neuronas en el cerebro humano, en la cual se replica su proceso biológico de las conexiones neuronales, su capacidad de autoaprendizaje del sistema nervioso y su buen funcionamiento de la conectividad e interdependencia de los procesos, con el fin de construir un modelo basado en esta analogía para obtener sus mismos buenos resultados.

McCulloch y Pitts (McCulloch & Pitts, 1943) diseñaron un modelo abstracto y simple de una neurona artificial, este es el elemento básico de procesamiento en una red neuronal. En donde el algoritmo del esquema siguiente, muestra la capacidad de la neurona de calcular de manera limitada una suma ponderada de sus entradas y posteriormente le aplica la función de activación con lo cual le transmitirá una señal a la siguiente neurona.

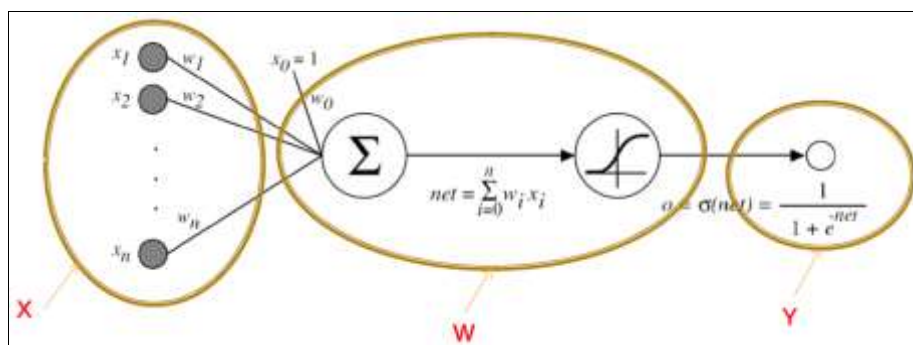


Ilustración 4: Detalles de una neurona.

La fórmula se puede simplificar a la siguiente expresión $Y = F(X, W)$,

Donde:

Y = vector formado por las salidas de la red

F = función continua no lineal

X = vector de entrada a la red

W = conjunto de todos los parámetros de la red (pesos y umbrales)

2.2.4.2 Support Vector Machines

Las máquinas de vectores de soporte (Support Vector Machines o SVM) son un conjunto de algoritmos desarrollados por Vladimir Vapnik (Vapnik, 1995).

Al igual que las redes neuronales, SVM aprende de un conjunto de entrenamiento, luego se aplica a un grupo de datos nuevos y busca clasificar de forma óptima a través de la generación de hiperplanos separadores, resolviendo un problema de optimización, el cual tiene dos objetivos, minimizar el error en la separación del conjunto de entrenamiento y maximizar el margen de separación. Esta técnica permite cumplir los dos objetivos anteriores minimizando el riesgo estructural.

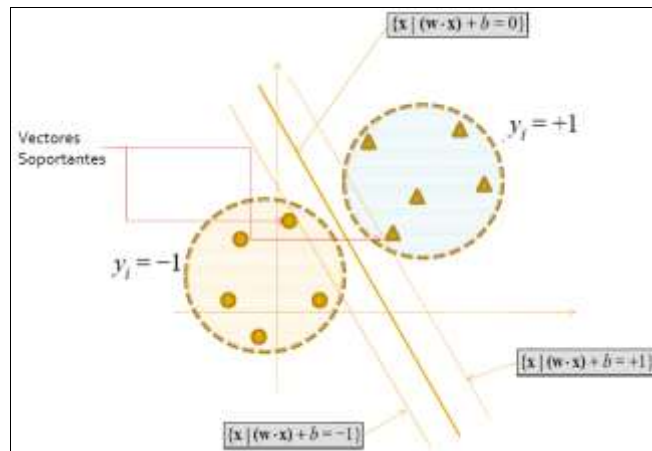


Ilustración 5: Hiperplano SVM

2.2.5 Árbol de decisiones

Son modelos supervisados que son de fácil interpretación, debido a su extenso y común uso. Ya se hizo la analogía de las redes neuronales con las neuronas del cerebro, en esta oportunidad la analogía es con un árbol común, el árbol más ilustrativo sería un pino. Este modelo considera el total de los datos y los comienza a clasificar desde la copa hasta la base, siempre en base a preguntas las cuales son binarias, 1 o 0, sí o no, dado lo anterior se utiliza mucho también para la programación por su fácil programación.

Un árbol de decisión se compone por nodos, que son las variables, los vectores que son las decisiones que se deben tomar y las hojas representan los resultados.

A continuación, se muestra un ejemplo:

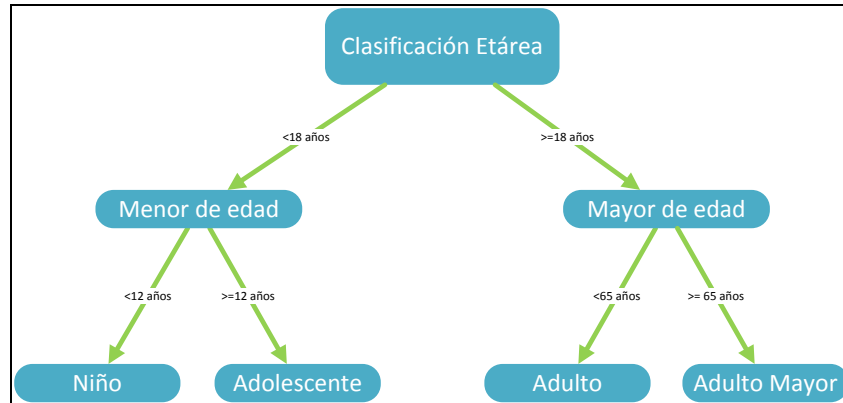


Ilustración 6: Ejemplo Árbol de Decisión. Fuente: Elaboración propia.

2.2.6 Regresión Logística

Tal como los modelos anteriormente mencionados, la regresión logística se utiliza para predecir una variable dependiente en función un grupo de variables independientes.

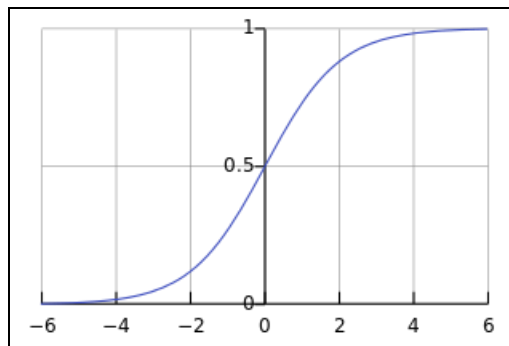


Ilustración 7: Función Logística.

La regresión logística tiene cuatro fases:

- Selección de variables de análisis
- Estimación de pesos y coeficientes de las variables
- Clasificación de los casos
- Análisis de resultados

$$Y = \frac{1}{1 + e^{-(B_0 + B_1 X_1 + B_2 X_2 + \dots + B_k X_k)}} \quad 0 < Y < 1$$

Ilustración 8: Función Regresión Logística.

Donde:

B_0 = posición de la curva sobre el eje horizontal

B_1 = pendiente de la curva medida en la zona de inflexión

X_k = ponderación proporcional para cada variable

2.3 Deserción Universitaria

A lo largo de la historia se han realizado diversos estudios sobre deserción universitaria y existe una investigación que recopila los modelos de los autores más connotados, Modelo Conceptual para la Deserción Estudiantil Universitaria Chilena (Díaz, 2008).

La matriz topológica que se muestra a continuación, permite hacer un análisis multidimensional y donde se relaciona las variables con sus autores y la relación entre autores.

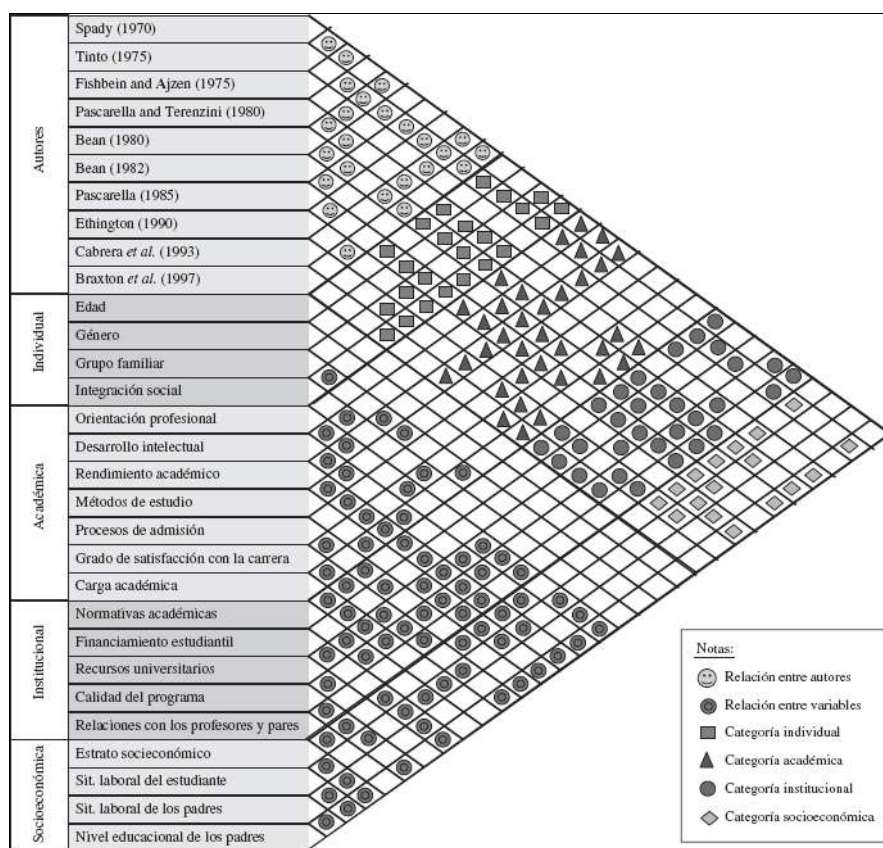


Ilustración 9: Matriz topológica de la sistematización de la literatura relacionada con la deserción estudiantil. Fuente: Christian Díaz Peralta

CAPÍTULO 3: PLANTEAMIENTO ESTRATÉGICO Y MODELO DE NEGOCIOS

3.1 Posicionamiento Estratégico

El posicionamiento estratégico del proyecto se basa en el modelo Delta, (Hax 2010), está situado entre redefinir la integración con el cliente y entregar una solución total al cliente, es decir, apuntando directamente a la integración al cliente, esto explica por qué la propuesta de valor es brindarle un muy buen servicio.

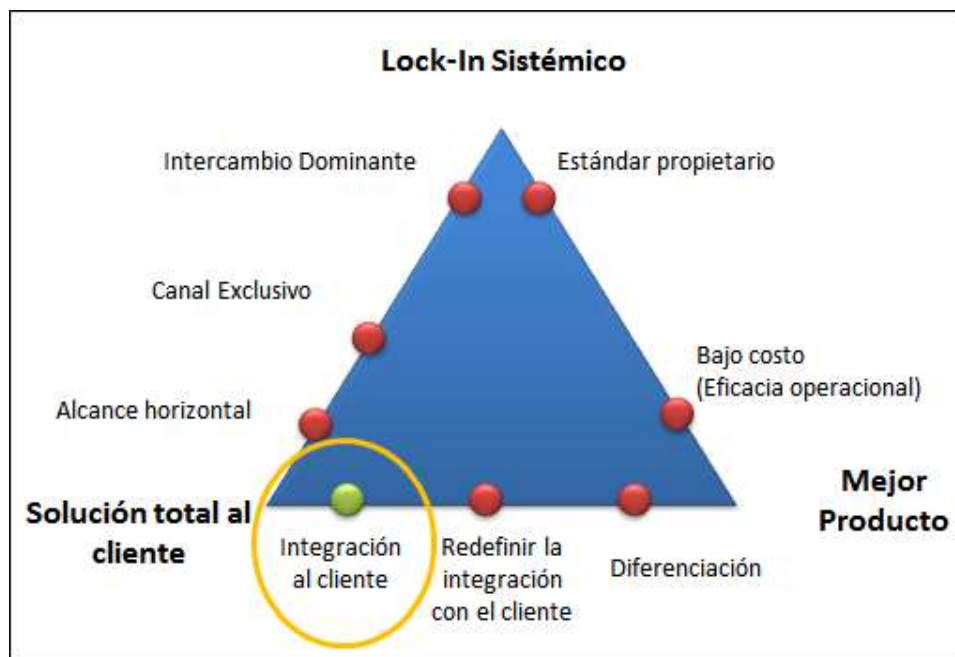


Ilustración 10: Posicionamiento Estratégico. Fuente: Elaboración Propia.

El modelo que resulta del proyecto, permite clasificar a los estudiantes recién ingresados de acuerdo a si continuarán o desertarán por algún motivo, basándose en la data histórica y así definir una estrategia de captación y retención hacia las necesidades especiales de cada estudiante.

3.2 Mapa Estratégico

A continuación, se presenta el mapa estratégico del proyecto, el cual fue construido a partir de reuniones establecidas con los stakeholders.

Para que el proyecto tenga una exitosa implementación se debe comenzar desde la base, para eso se debe integrar y adaptar el proceso de capacitación, donde el personal adquirirá nuevas competencias, las cuales van de la línea de los objetivos

estratégicos. Una vez que se haya realizado con éxito lo anterior se debe realizar la integración de los procesos y tecnología para optimizar el proceso de gestión académica con lo cual se optimizan los recursos y se mantiene una estrecha comunicación con el estudiante para obtener un aumento en la retención con el consiguiente aumento del ingreso por concepto de arancel.

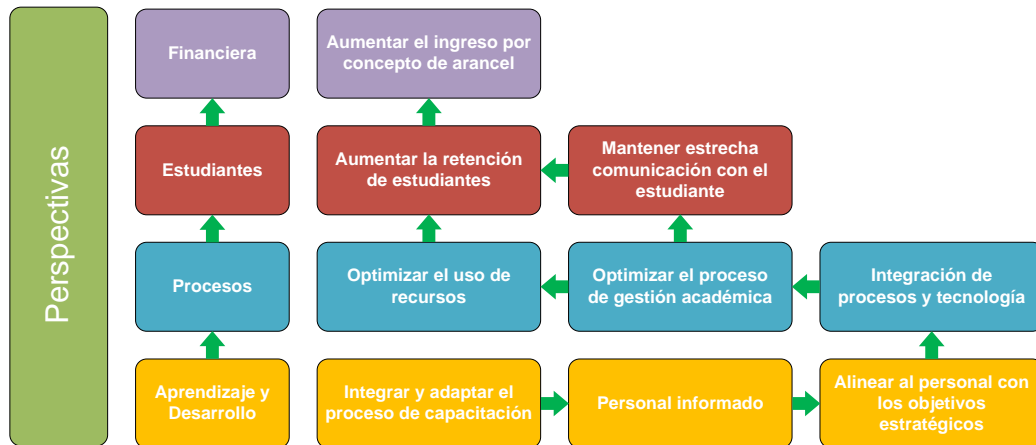


Ilustración 11: Mapa Estratégico. Fuente: Elaboración Propia.

3.3 Modelo de Negocios

El modelo de negocios, de acuerdo a la estructura propuesta por Johnson (2008), debe contener la propuesta de valor para el cliente, fórmula de utilidades, recursos claves y procesos claves.

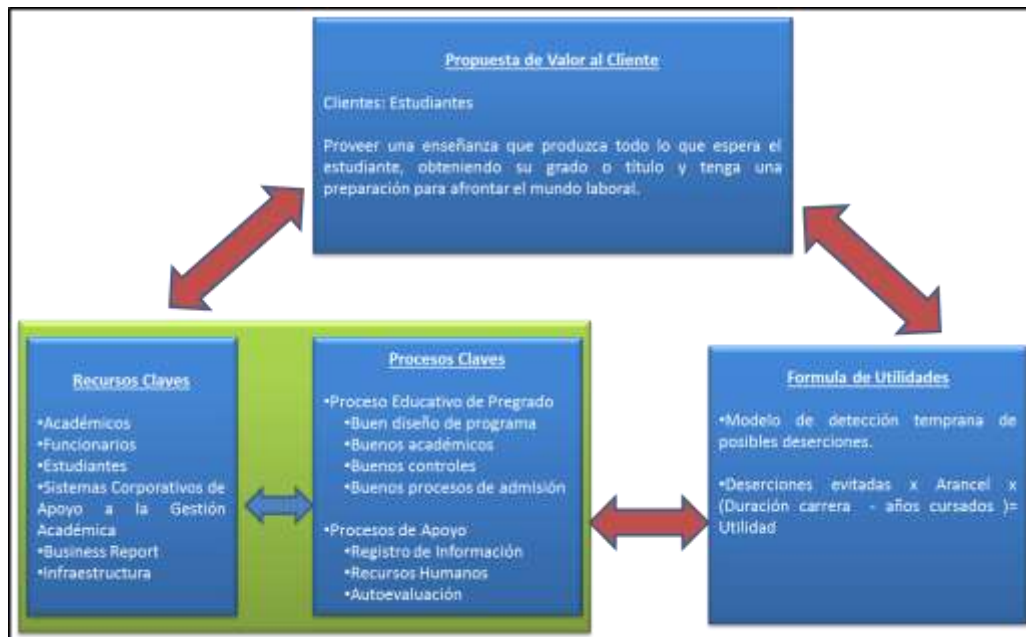


Ilustración 12: Modelo de Negocios. Fuente: Elaboración Propia.

3.3.1 Propuesta de Valor al Cliente

La propuesta de valor del proyecto es que la Universidad de Chile se preocupe de proveer una enseñanza que le permita entregar lo que es ofrecido tanto en los reglamentos como en los discursos y que es lo que el estudiante realmente espera obtener al ingresar a una de las universidades más prestigiosas, además se debe considerar que el estudiante ingresa para obtener conocimiento y por el producto final que es el grado o título, pero todo esto no serviría si después enfrentado en el mundo laboral no tiene las herramientas para desenvolverse adecuadamente.

3.3.2 Recursos claves

Los recursos claves son los académicos que son quienes imparten la docencia y son el vínculo de la entrega del conocimiento, los funcionarios quienes con su trabajo hacen que la Universidad marche adecuadamente, los estudiantes quienes son los clientes, pagan por un producto y servicio y financian la universidad, los sistemas corporativos de apoyo a la gestión académica junto a Business Report, el primero, que permite registrar la información y el segundo obtener conocimiento de los datos y poder hacer gestión con ellos para mejorar los servicios ofrecidos por la institución y por último y no por eso menos importante la infraestructura que alberga a los estudiantes y les brinda comodidad y el ambiente adecuado para recibir el conocimiento.

3.3.3 Procesos claves

Los procesos claves son el proceso educativo de pregrado el cual debe tener principalmente un buen diseño del programa, las horas cronológicas adecuadas en los mejores horarios y una correcta distribución de malla para balancear la carga académica, además de buenos académicos, los cuales cada vez son requeridos con grados más altos, si antes era requisito mínimo poseer un magíster, hoy se exige tener un grado de doctor, además existe la instancia de la evaluación docente, donde se plasma como los estudiantes perciben el trabajo realizado por los académicos.

Utilizando los recursos claves de sistemas corporativos y Business Report, se debe realizar el seguimiento de los procesos, al ser una tarea permanente, éste se transforma en un proceso en sí, que cobra gran importancia para lograr la propuesta de valor controlando que se vaya ejecutando de acuerdo a lo planificado.

Existen procesos de apoyo como el registro de la información lo cual se debe asegurar que se encuentre disponible 24x7, los recursos humanos que permiten ejecutar el proyecto y un proceso de autoevaluación que mida el avance y logro de objetivos del proyecto educativo.

3.3.4 Fórmula de utilidades

$$Utilidad = \delta * \alpha * (\gamma - \varphi)$$

Donde

δ = *Deserciones evitadas*

α = *Arancel del Programa de Estudios*

γ = *Duración del programa de estudios*

φ = *Años cursados en el programa de estudios*

La fórmula de utilidades se puede expresar de la siguiente manera, un estudiante que según el modelo indique que va a desertar, se le aplica reforzamiento y se evita que el estudiante deserte, esto haría sumar 1 el valor de δ , ya que sería una deserción evitada, luego ésta se multiplica por el arancel del programa de estudios que está matriculado el estudiante (α), todo esto se debe multiplicar por la diferencia entre la duración del programa de estudios (γ) y los años cursados que lleva el estudiante (φ), esto último se explicaría como la permanencia futura del estudiante.

En el capítulo de financiamiento se explica con mayor detalle la fórmula de utilidades.

3.3.5 Amenazas de nuevos competidores

En el ámbito de la educación superior, si bien existe una amplia oferta educacional que no logra satisfacer la demanda, las barreras de entradas son muy altas, lo que implica que la amenaza de que surjan nuevos competidores sea muy baja.

3.3.6 Amenazas de productos y servicios sustitutos

La amenaza de productos y servicios sustitutos es baja, podrían considerarse los programas online y algunos Institutos, sin embargo no impacta al mercado de la Universidad de Chile.

3.3.7 Poder de Negociación de clientes

El poder de negociación de los clientes (estudiantes) es media, ya que estos han tomado mucha fuerza en los últimos años y al organizarse son capaces de producir cambios, sin embargo se deben atener a los estatutos universitarios.

3.3.8 Poder de Negociación de Proveedores

Cada vez entran más académicos al sistema, obteniendo grados a muy baja edad, subiendo los estándares de la universidad, esto genera una competitividad cada vez más grande lo que permite aún que la negociación de los académicos sea baja.

3.3.9 Rivalidad entre Competidores

La rivalidad entre competidores es alta, ya que existen Universidades del CRUCH como la Pontificia Universidad Católica y algunas privadas como la Universidad Adolfo Ibáñez, que han hecho que la Universidad de Chile deba invertir en difusión y competir tramo a tramo con ellas.

3.4 Análisis FODA

3.4.1 Fortalezas

- Alta competencia del cuerpo académico.
- Recoge los mejores estudiantes del país.
- Alta calidad de oferta académica de Pregrado y Postgrado.
- Capacidad de internacionalización.

3.4.2 Debilidades

- Falta de comunicación entre las diferentes Unidades Académicas.
- Alta dependencia financiera de la obtención de recursos propios a través de proyectos.
- Tendencia del cuerpo académico a transformarse en gestores de financiamiento institucional.

3.4.3 Oportunidades

- Concursos de financiamiento.
- Existencia de redes internacionales vinculadas a los procesos de formación de estudiantes e investigación.

3.4.4 Amenazas

- Gran crecimiento del sistema universitario privado.
- Burocratización creciente del sistema de financiamiento y compras, comparado con las universidades privadas.

CAPÍTULO 4: ANÁLISIS SITUACIÓN ACTUAL

La arquitectura de procesos actual de la Universidad de Chile, utilizando la metodología de ingeniería de negocios propuesta por Barros (2004), permite adecuar los patrones de negocios desde los 4 macroprocesos hasta el proceso específico de interés del proyecto.

4.1 Macroproceso 1: Cadena de Valor.

Procesos que entregan el servicio de educación al estudiante, desde la postulación y admisión, pasando por matrícula, inscripción de asignaturas y terminando con la graduación.

4.2 Macroproceso 2: Desarrollo de nuevas capacidades.

Desarrollo de nuevos programas, los nuevos programas de pregrado responden a necesidades de la institución y es una decisión de alto nivel donde el rector está involucrado, por otro lado los programas de postgrado se centran en una necesidad de las Unidades Académicas las cuales formulan el nuevo programa y deben contar con la supervisión y aprobación del Departamento de Postgrado y Postítulo.

4.3 Macroproceso 3: Planificación del Negocio.

Planificación de la Universidad, contempla todos los procesos y actividades de planificación que realizan los altos directivos, entiéndase Rector, Prorector, Vicerrectores, Consejo Universitario y Decanos, quienes en conjunto con sus respectivos equipos bajan líneas de planificación hasta el más bajo nivel directivo.

4.4 Macroproceso 4: Procesos de Apoyo.

Procesos de Apoyo a los otros macroprocesos, contempla Aranceles, Recursos Humanos, Matrícula, donde intervienen funcionarios del área TI, de las unidades académicas y de los Departamentos de Pregrado y Postgrado y Postítulo.

A continuación se muestran los macroprocesos de la Universidad de Chile.

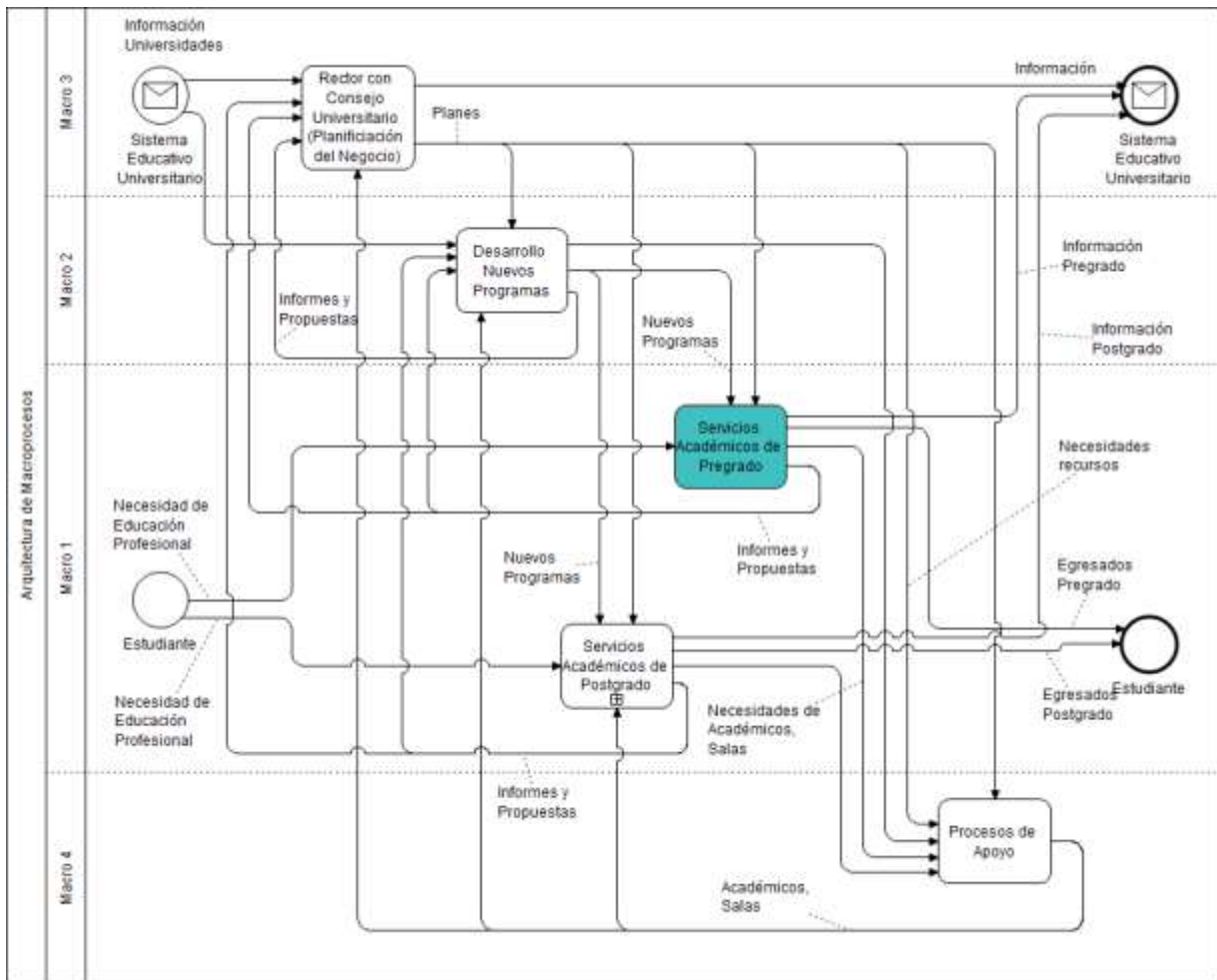


Ilustración 13: Macroprocesos Universidad de Chile. Fuente: Elaboración Propia.

4.5 Servicios Académicos de Pregrado

Los servicios académicos de pregrado, se preocupan tanto de la administración de la relación con el estudiante, la administración de la relación con el académico, la gestión de los procesos académicos y los procesos propiamente tales.

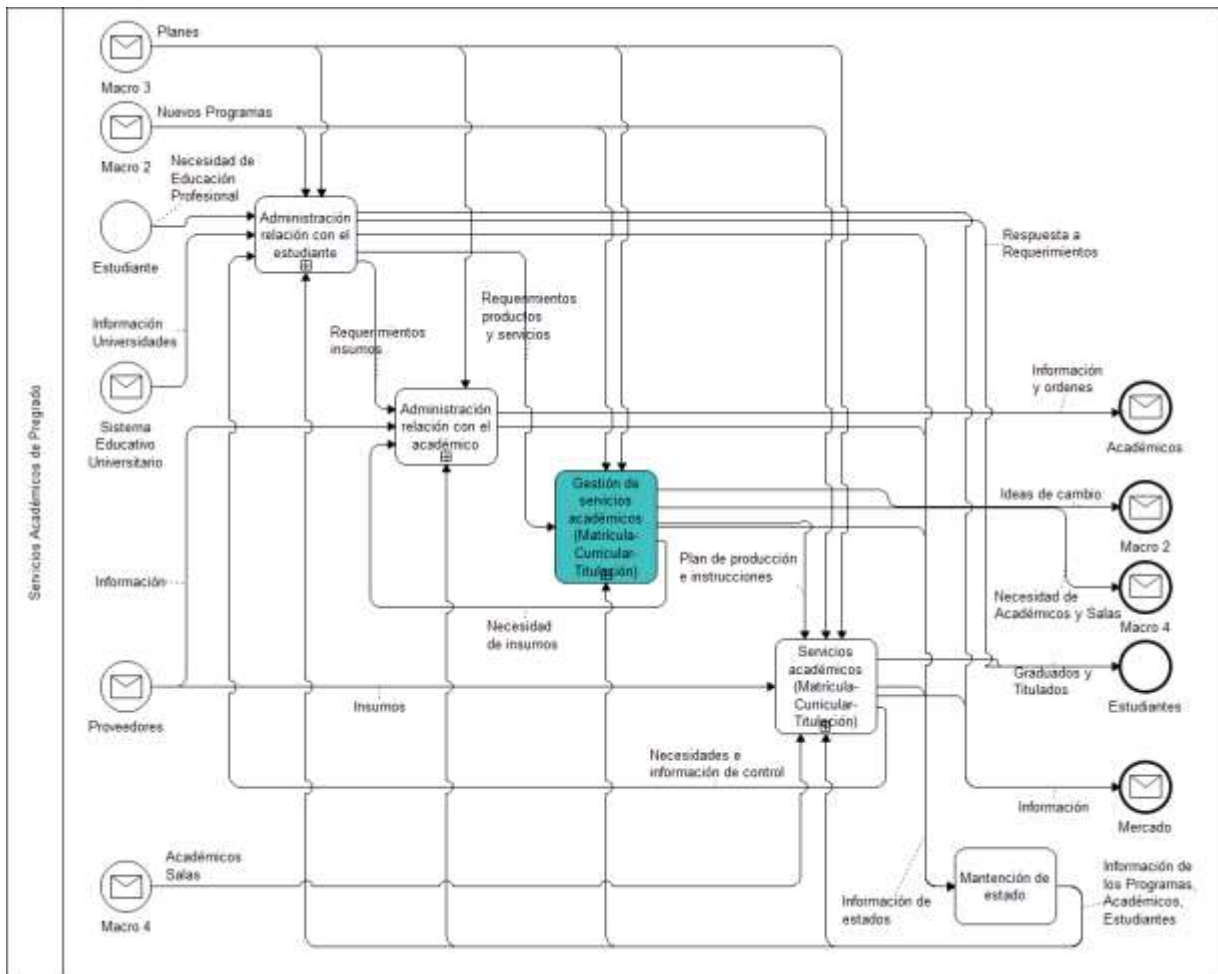


Ilustración 14: Servicio académicos de Pregrado. Fuente: Elaboración Propia.

4.5.1 Administración de relación con el estudiante

La administración de la relación con el estudiante se realiza principalmente en las unidades académicas, pero en caso de que éstas no puedan resolver alguna situación es escalada a nivel del Departamento de Pregrado, por ejemplo, solicitudes de reincorporación, problemas atendibles, entre otros.

4.5.2 Administración de relación con el académico

La administración de la relación con los académicos se realiza principalmente en los departamentos de las unidades académicas, contrataciones, asignación de cursos, evaluación académica, entre otros.

4.5.3 Gestión de Servicios Académicos

A nivel central existe en el Departamento de Pregrado una Unidad de Gestión Académica la cual gestiona los procesos de los servicios académicos de pregrado tales como postulación, matrícula, inscripción de asignaturas y registro de calificaciones, quienes tienen comunicación directa con las secretarías de estudios.

4.5.4 Servicios Académicos

Son todos los servicios que la Universidad de Chile provee al estudiante para obtener conocimiento, desde su postulación hasta recibir su grado o título.

4.6 Gestión de Servicios Académicos

La gestión de servicios académicos de pregrado, se preocupan tanto de la implementación de nuevos programas, de la planificación y control de procesos académicos y debe decidir la entrega de los procesos académicos.

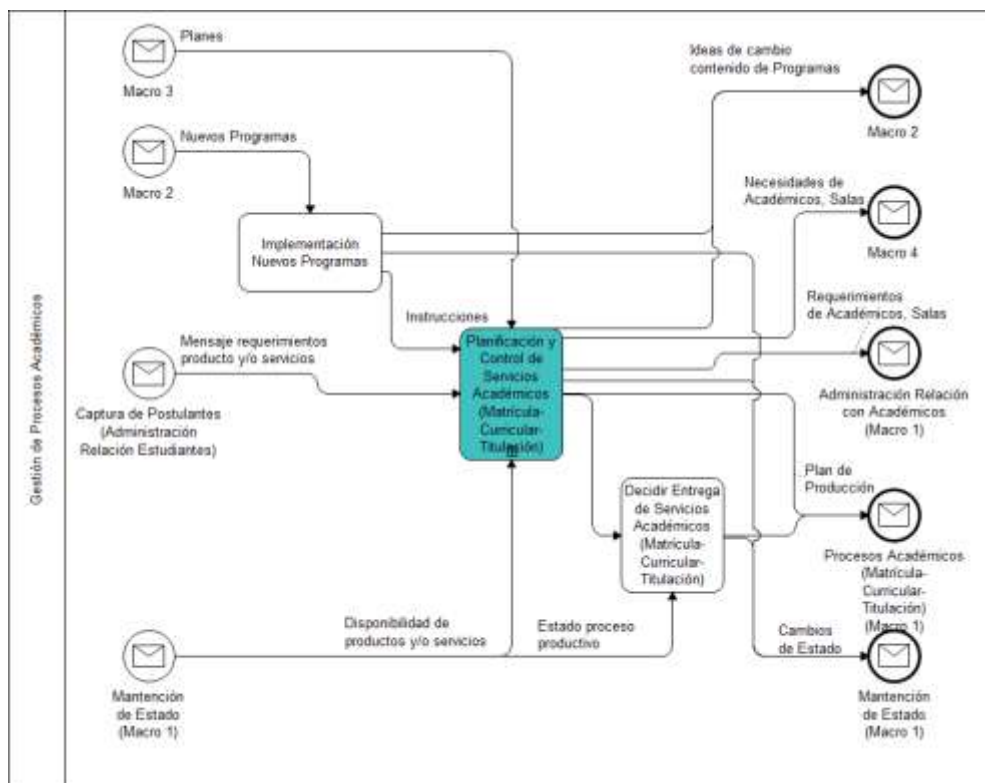


Ilustración 15: Gestión de servicios académicos. Fuente: Elaboración Propia.

4.6.1 Implementación de nuevos programas

El Departamento de Pregrado solicita la creación del programa en el sistema GUIA Matrícula y en U-Campus, define las vacantes y prepara la admisión para el próximo período académico.

4.6.2 Planificación y control de servicios académicos

Aquí es donde el proyecto deberá ejecutarse, ya que no se encuentra desarrollado el seguimiento del estudiante.

4.6.3 Decidir entrega de procesos académicos

Dada la planificación del punto anterior, en esta actividad se debe decidir cuándo se hará la entrega de los procesos de matrícula, curricular y titulación.

4.7 Diagnóstico de la Situación Actual

Este año se ha puesto sobre el tapete el tema de la deserción y se ha realizado el primer estudio el cual caracteriza al estudiante una vez que ya ha desertado, lo que permite de alguna forma entender el motivo por el cual se están yendo.

Se debe crear un proceso de seguimiento, el cual va acompañando al estudiante durante su primer año y de esta manera evitar su deserción, ya que en la actualidad no existe un proceso a nivel central que permita realizar un correcto seguimiento del estudiante, lo cual se refleja en los niveles de deserción y extensos tiempos de permanencia en los programas de formación, lo anterior es una oportunidad para realizar un modelo de predicción basado en la información que el estudiante trae desde el proceso de postulación y así poder evitar que renuncie o sea eliminado del programa.

Se ha podido comprobar en otras instancias que los portales de información con ingreso diferenciado y administrado centralmente como el SIGA, han tenido una buena recepción, de acuerdo a sus experiencias y resultados, dado lo anterior, implementar el seguimiento del estudiante con esta plataforma se traduce en una gran oportunidad.

4.8 Cuantificación del Problema u Oportunidad

4.8.1 Deserción en la Educación Superior³

Según la Serie Evidencias del Ministerio de Educación (2012), se indica que en los últimos 30 años, el sistema de educación superior experimentó un crecimiento que significó pasar de una matrícula de pregrado de 165 mil estudiantes a principios de los años ochenta, a una matrícula de más de 1 millón de estudiantes el año 2012.

Este desarrollo es, sin lugar a dudas, un logro importante. De acuerdo a la OECD (2009), gracias a este fenómeno el sistema de educación superior pasó de ser un sistema de elite a uno masivo. Sin embargo, este aumento en la matrícula trajo consigo un cambio en el perfil y en las necesidades de los estudiantes. Frente a ello, se hicieron más patentes otros problemas como es el caso de la deserción.

En Chile, más del 50% de quienes se matriculan en la educación superior no concluyen el programa en el que se matricularon inicialmente. Esta situación genera importantes pérdidas de eficiencia para el Estado y las instituciones, así como disminución de oportunidades para los estudiantes y sus familias.

4.8.2 Deserción en la Universidad de Chile

El concepto de deserción no es nuevo en la Universidad de Chile, desde más de una década se realizaban informes que se disponibilizaban a las grandes autoridades, sin embargo se realizaron cambios tanto en altas direcciones del soporte tecnológico, como por su parte la VAA comienza a incorporar nuevos criterios y definiciones que permitan la creación de reportes periódicos y que sean distribuidos a quienes participan en los procesos que inciden en la mejora de la retención.

Se realizan dos mediciones de tasa de deserción a nivel Institucional, una que involucra el Programa Académico de Bachillerato y otra que no, esto debido a que este programa es utilizado por estudiantes que se transfieren a otros programas y el 50,8% no se matricula en segundo año para terminar con un grado de Bachiller, muchos de ellos consideran este programa como momentáneo, donde inscriben asignaturas que le pueden servir para su próximo plan y así ingresar el próximo año, ya sea por medio de transferencia interna o rindiendo nuevamente la PSU al Programa de Formación deseado.

³ Serie Evidencias Ministerio de Educación (2012)

A continuación se puede apreciar la tasa de deserción de los programas de formación de pregrado excluyendo el Programa Académico de Bachillerato.

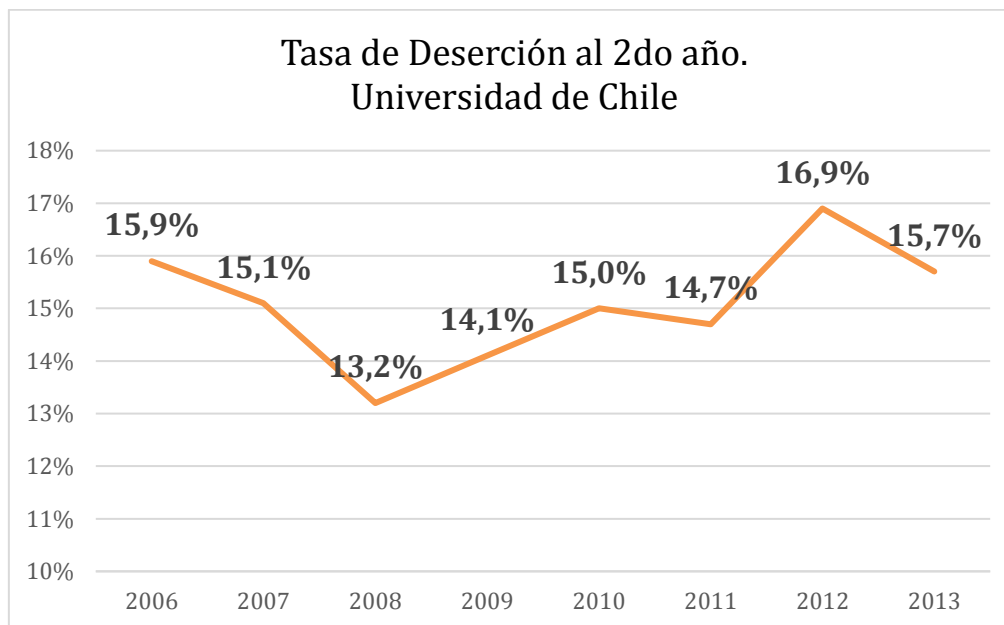


Ilustración 16: Tasa de Deserción. Fuente: Anuario U. de Chile. Síntesis 2006-2013

CAPÍTULO 5: PROPUESTA DE DISEÑO DE PROCESOS

Continuando con la metodología de Ingeniería de Negocios, para lograr los objetivos del proyecto es necesario realizar un rediseño en la ubicación específica de Planificación y Control de Servicios Académicos e implementar un seguimiento del estudiante que permita que éste no se retire tanto por motivos personales como académicos.

5.1 Direcciones de Cambio y Alcance

A continuación se detallan las variables de diseño, las cuales permiten contrastar la situación actual con la propuesta y que los esfuerzos de rediseño concuerden con los objetivos del proyecto.

Tabla 2: Variables de Estructura de Empresa y Mercado

Variable	Situación Actual	Situación Propuesta
a.1. Servicio integral al cliente	NO	SI
a.2. Lock-in sistémico	NO	NO
a.3. Integración con proveedores	NO	NO
a.4. Estructura interna: centralizada o descentralizada	Descentralizada	Centralizada
a.5. Toma de decisiones: centralizada o descentralizada	Descentralizada	Centralizada

La tabla anterior muestra la concordancia de los tres primeros puntos con el posicionamiento estratégico ya expuesto en el capítulo 3, donde la situación propuesta va dirigida hacia el servicio integral del estudiante. Si bien, las unidades académicas con sus respectivos programas de formación trabajan muy coordinados con el Departamento de Pregrado en cuanto a los procesos académicos, tales como admisión, matrícula y registro de calificaciones, lo que se podría entender como una estructura interna centralizada, finalmente la toma de decisiones y la estructura para el área específica del proyecto se realiza de forma descentralizada, dado lo anterior el proyecto busca generar las instancias de coordinación para una toma de decisiones en conjunto con el Departamento de Pregrado en el ámbito de gestión y mejora de los indicadores propuestos.

Tabla 3: Variables de Anticipación

Variable	Situación Actual	Situación Propuesta
b.1. Planificación con Centros de Aprendizaje	NO	Planificar el año de acuerdo al listado de estudiantes que se deben nivelar para retener.
b.2. Modelo Predictivo riesgo de deserción	NO	Modelo basado en BI y que determina automáticamente el riesgo de deserción
b.3. Monitoreo de centros de aprendizajes	NO	Permite ver si se está ejecutando bien las acciones de mejoramiento

Tabla 4: Variables de Coordinación

Variable	Situación Actual	Situación Propuesta
c.1. Reglas	Reglas informales	Reglas formales con apoyo Computacional para la prevención de deserción
c.2. Jerarquía	Se usa cuando los Secretarios de Estudios deben decidir por las solicitudes de los estudiantes	Se usa cuando los Secretarios de Estudios deben decidir por las solicitudes de los estudiantes
c.3. Colaboración	Existe colaboración, pero no queda registrado.	Colaboración a través de seguimiento de Centros de Aprendizaje
c.4. Partición	NO	NO

Tabla 5: Variables de Prácticas de Trabajo

Variable	Situación Actual	Situación Propuesta
d.1. Lógica de negocio automatizada o semiautomatizada.		
• Elaboración de Informes de Deserción	NO	Lógica 100% automatizada, para la confección de documentos online
• Revisión de Documentos transversales en línea	NO	Lógica 100% automatizada, consulta web sobre información de matrícula, entre otros

d.2. Lógica de apoyo a actividades tácitas	NO	Lógica de medición de KPI's para apoyar planificación Estratégica
d.3. Procedimientos de comunicación e integración	NO	Utilización de Google Drive y Trello para la comunicación en línea, lo cual incentiva la colaboración
d.4. Lógica y procedimientos de medición de desempeño y control	NO	Se mide desempeño a través de KPI's

Tabla 6: Variables de Integración de Procesos Conexos

Variable	Situación Actual	Situación Propuesta
e.1. Proceso aislado	SI	NO
e.2. Todos o la mayor parte de los procesos de un macroproceso	NO	NO
e.3. Dos o más macros que interactúan	NO	NO

Tabla 7: Variables de Mantenimiento Consolidada de Estado

Variable	Situación Actual	Situación Propuesta
f.1. Datos propios	SI	SI
f.2. Integración con datos otros sistemas de la empresa	SI	SI, GUIA Matrícula, Becas y Beneficios entre otros.
f.3. Integración con datos de sistemas otras empresas	SI	Información del DEMRE ⁴

5.2 Diseño Detallado de Procesos TO BE

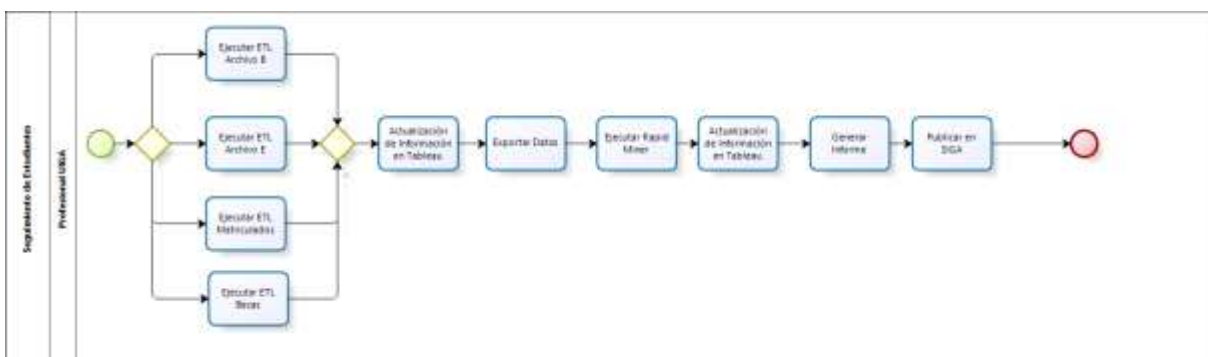


Ilustración 17: Generación de listas de estudiantes.

⁴ Departamento de Evaluación, Medición y Registro Educativo

El proceso se inicia cuando la información del DEMRE es entregada, aproximadamente el 12 de enero, sin embargo los estudiantes nuevos finalizan su proceso de matrícula a fines de enero. El profesional de la Unidad de Información para la Gestión Académica (UIGA) realiza la carga de los datos, Archivo B, Archivo E, Matriculados y Becas a archivo base, el cual para que sea considerado en los reportes se debe actualizar en Tableau.

Luego se exporta un archivo Excel, el cual es leído por RapidMiner para usarse como archivo de predicción, es aquí donde ocurre la lógica de la aplicación, ya que una vez ejecutado, se genera un nuevo archivo Excel, el cual es leído por Tableau para entregar un reporte con el listado de los estudiantes que podrían desertar.

El reporte queda disponible en el portal SIGA, donde el secretario de estudios podrá ver el listado de estudiantes.

5.3 Diseño de Lógica de Negocios

Se realizó una prueba de concepto, utilizando el Proceso de Extracción de Conocimiento (KDD), donde se probaron 4 algoritmos de data mining distintos, Support Vector Machines (SVM), Árboles de Decisión, Regresión Logística y Redes Neuronales en el software de minería de datos RapidMiner, estos fueron seleccionados de acuerdo a los buenos resultados que han tenido en otros estudios.

El proceso de RapidMiner es el siguiente:

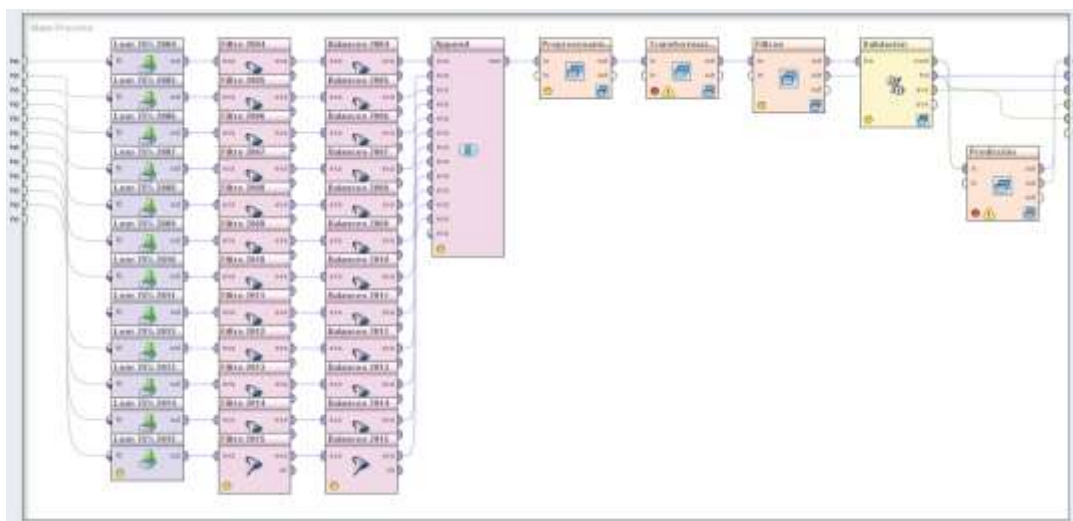


Ilustración 18: Proceso en RapidMiner. Fuente: Elaboración Propia

5.3.1 Selección de datos:

En esta primera etapa se deben seleccionar los datos que contendrá el modelo, para ello se utiliza la información recopilada de: DEMRE, proceso de matrícula, becas y beneficios, situación académica y finalmente el cálculo de algunas variables cruzadas.

Se construye un datamart⁵, este punto ha sido de un gran esfuerzo, lo cual ha demandado una gran cantidad de tiempo, ya que no se contaba con uno, el cual permita hacer los cruces para por ejemplo unir la información de postulación con la de matrícula, permitiendo caracterizar al estudiante con sus datos obtenidos del proceso de postulación. (Anexo 2)

Las fuentes de datos con sus respectivos campos son:

Tabla 8: Fuentes de datos con sus respectivos campos

Postulación DEMRE Archivo B: Información Socioeconómica		
Cuantos Trabajan	De Proseguir Estudios	Tipo Organismo donde Trabajan
Viven Sus Padres	Situación Ocupacional	Horas que dedica al Trabajo
Horario Trabajo	Ocupación Principal	Quien es el Jefe de Familia
Grupo Familiar	Cobertura de Salud	Tiene Trabajo Remunerado
Quien Financia	Cuantos Estudian	Educación de los Padres
Rama Actividad	De Proseguir Estudios	Tipo Organismo donde Trabajan
Postulación DEMRE Archivo E: Postulaciones Efectivas		
Estado Postulación	Promedio LM	Nacionalidad
Situación Egreso	Lenguaje y Comunicación	Grupo Dependencia
Año Egreso	Matemáticas	Rama Educacional
NEM ⁶	Historia y Ciencias Sociales	Ingreso Bruto Familiar
Puntaje Ponderado	Ciencias	Preferencia
Matrícula		
Cohorte	Fecha de Nacimiento	Sexo
Región de Origen	Origen Étnico	
Beca		
Tiene Beca		
Situación Académica		
Situación Académica	Tipo Causa	Motivo
Campos calculados		
Identificador Nombre	Deserción	Educ Padre
Educ Madre	Educ Máxima	Edad

⁵ Datamart: subconjunto de datos para la toma de decisiones en un área específica.

⁶ NEM: Nota Enseñanza Media

Estos datos son cargados a Tableau para hacer el cruce entre fuentes de datos, con eso se obtiene el archivo que es cargado en RapidMiner

RapidMiner lee el archivo Excel que contiene el 75% de la información de los estudiantes matriculados nuevos desde el año 2004 al 2015 de la carrera de Arquitectura. Existe una caja para cada año la cual se diferencia solamente por su nombre, ya que el filtro se realiza en la caja de la actividad siguiente y existe una caja para cada año desde 2004 a 2015.

El 25% restante es apartado para que luego sea utilizado como entrada para la predicción, lo anterior permite probar cómo se comporta el modelo a lo largo de los años, esto no ocurriría si los datos son utilizados para el entrenamiento del modelo.

Posteriormente se balancea por año, esta actividad es muy importante, ya que previamente se tiene un conjunto de datos donde la mayoría no deserta, es una proporción de un 80% aproximadamente, lo que desbalancea la muestra de acuerdo a la variable objetivo y no favorece a la detección de estudiantes que desertan. Para resolver esto se balancea extrayendo aleatoriamente tuplas del universo que no deserta para dejar en la misma cantidad ambos valores de la variable.

Tabla 9: Balanceo por año.

Año	Desertan	Año	Desertan
2004	28	2010	54
2005	40	2011	50
2006	36	2012	59
2007	34	2013	60
2008	37	2014	53
2009	43	2015	57

Para finalizar esta etapa se unen todos los datos que han ingresado al modelo, si recordamos es el 75% de cada uno de los años. Esto se realiza porque debe ser un conjunto de datos el que debe ir hacia adelante del proceso.

5.3.2 Preprocesamiento de Datos:

En esta actividad se realiza la limpieza de datos, donde se ha utilizado reemplazar los valores perdidos por el promedio, como se muestra a continuación.



Ilustración 19: Preprocesamiento de datos. Fuente: Elaboración Propia

5.3.3 Transformación:

En esta actividad, como bien dice su nombre se transforman para adaptarse a los distintos algoritmos, de modo que las variables sean más fáciles de entender.

Se utilizó Nominal a Binominal la cual transforma las variables que tienen categorías, en variables con 2 posibles valores 1 o 0. Luego se transforman las variables en un rango numérico y finalmente se normalizan, es decir se dejan dentro de un rango 0 y 1.



Ilustración 20: Transformación. Fuente: Elaboración Propia

Además se utilizan filtros para seleccionar y remover variables que no aporten a la variable objetivo de deserción, por ejemplo aquellas que no cumplen con una desviación menor a 0,01.

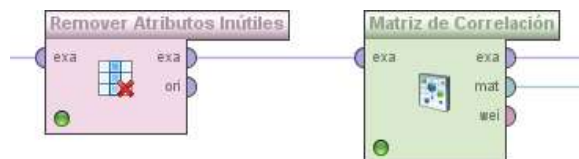


Ilustración 21: Filtros. Fuente: Elaboración Propia

Se ejecuta la matriz de correlación la cual normaliza el peso de las variables con respecto entre cada una de ellas y la variable objetivo, lo que permite saber si una variable indica lo mismo que otra y así eliminar una de las dos, mientras menos

variables tenga el modelo, más fácil de explicar es y si existe una variable que no aporte a la toma de decisiones no es necesario que ésta sea considerada.

5.3.4 Minería de Datos:

Es el subproceso donde se aplica el modelo, es el corazón del data mining, busca patrones de comportamiento para poder explicar el modelo, es aquí donde se probaron los 4 algoritmos expuestos anteriormente y fueron las Redes Neuronales las que mejor comportamiento tuvieron. Una vez que se determinó el algoritmo, se ajustaron los pesos de las redes, número de ciclos de aprendizaje y otras variables sensibles para ir obteniendo un mejor resultado.

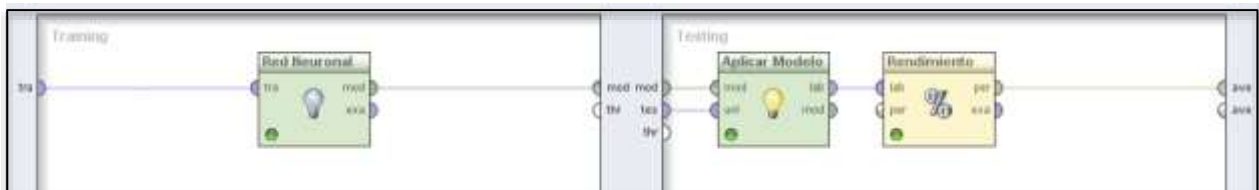


Ilustración 22: Validación. Fuente: Elaboración Propia

Se debe recordar que se está trabajando con el 75% de los datos y con balanceo, en esta parte del proceso se trabaja con un 70% para entrenar los datos y un 30% para testear. Se ejecuta una red neuronal con 1500 ciclos de entrenamiento, con una tasa de aprendizaje de 0,3 y un momentum de 0,2. Luego se aplica el modelo con el grupo de testeo y finalmente se mide el rendimiento, seleccionando la precisión del modelo y el área bajo la curva (AUC por sus siglas en inglés)

La predicción contiene varias actividades, primero lee la hoja del archivo que contiene el 25% restante de los años 2004 a 2015, pasa por el mismo proceso de transformación anterior, se le aplica el modelo que ya ha sido entrenado y testeado, luego se le mide el rendimiento de esta predicción y finalmente se escribe un archivo Excel, en el cual a la variable objetivo se le ha asignado un valor entre 0 y 1 el cual si está más cerca del 1 desertará el próximo año y más cerca del 0 se mantendrá matriculado en el mismo programa.

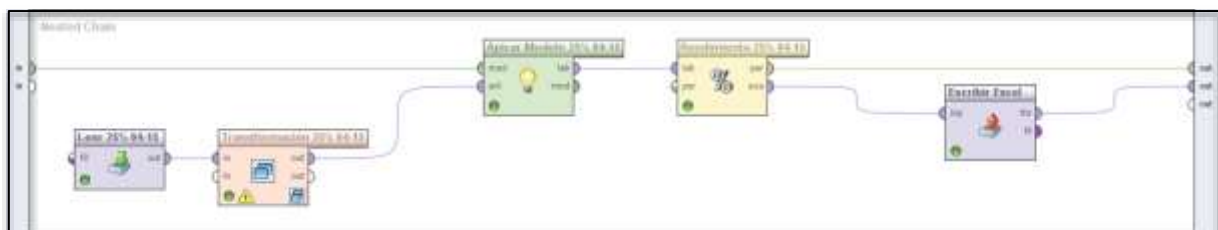


Ilustración 23: Predicción. Fuente: Elaboración Propia

5.3.5 Interpretación y Evaluación:

Las variables que se utilizan en el modelo son las siguientes:

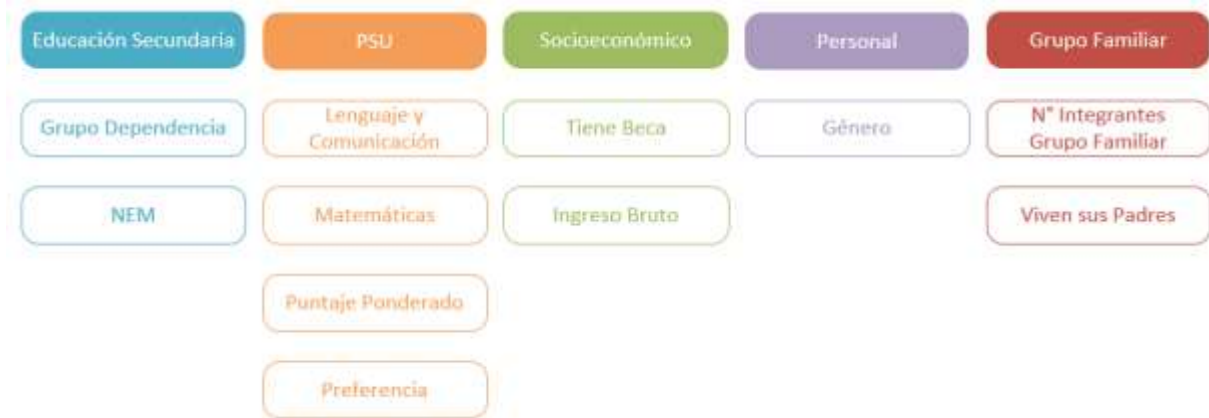


Ilustración 24: Variables utilizadas en el modelo. Fuente: Elaboración Propia

A continuación se muestra el resultado del modelo obtenido

Tabla de rendimiento de clasificación multiclase:

Multiclass Classification Performance			
accuracy: 51.68%			
	true 0.0	true 1.0	class precision
pred. 0.0	172	47	78.54%
pred. 1.0	241	136	36.07%
class recall	41.65%	74.32%	

Ilustración 25: Matriz de Confusión. Fuente: Elaboración Propia

El resultado es sobresaliente, ya que el modelo detecta al 74,32% de los estudiantes que desertaron, lo cual es un muy buen modelo.

5.3.5.1 Propuestas de Modelos.

Se realizaron más de 100 combinaciones distintas, luego para comparar cada uno de los resultados se confeccionó una planilla de cálculo, la cual genera la matriz de confusión para cada uno de los modelos a comparar. (Anexo 3)

A continuación se detallarán dos propuestas realizadas por académicos del magíster, para encontrar el mejor modelo:

5.3.5.1.1 Propuesta 1: Prof. Ezequiel Muñoz.

Se solicitó realizar los siguientes 24 casos, todos ellos considerando un 75% de las tuplas escogidas aleatoriamente y dejando un 25% para la predicción, cada uno de estos casos se realiza con y sin balanceo.

		Entrenamiento y testeo															Predicción									
		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	
Caso 1	Con balanceo	75%															25%									
Caso 2	Sin balanceo	75%															25%									
Caso 3	Con balanceo	75%															25%									
Caso 4	Sin balanceo	75%															25%									
Caso 5	Con balanceo	75%															25%									
Caso 6	Sin balanceo	75%															25%									
Caso 7	Con balanceo	75%															25%									
Caso 8	Sin balanceo	75%															25%									
Caso 9	Con balanceo	75%															25%									
Caso 10	Sin balanceo	75%															25%									
Caso 11	Con balanceo	75%															25%									
Caso 12	Sin balanceo	75%															25%									
Caso 13	Con balanceo	75%															25%									
Caso 14	Sin balanceo	75%															25%									
Caso 15	Con balanceo	75%															25%									
Caso 16	Sin balanceo	75%															25%									
Caso 17	Con balanceo	75%															25%									
Caso 18	Sin balanceo	75%															25%									
Caso 19	Con balanceo	75%															25%									
Caso 20	Sin balanceo	75%															25%									
Caso 21	Con balanceo	75%															25%									
Caso 22	Sin balanceo	75%															25%									
Caso 23	Con balanceo	75%															25%									
Caso 24	Sin balanceo	75%															25%									

Ilustración 26: Casos de prueba, Propuesta 1. Fuente: Elaboración Propia

Se confeccionó en RapidMiner un gran proceso, el cual sólo se van desactivando cajas, las cuales permiten realizar las 24 combinaciones mencionadas. En caso de que sea sin balanceo, toda la tercera columna está apagada y en caso de que se quiera considerar menos años se van apagando las 3 cajas del año respectivo, es decir, Leer Excel del Año, Filtro Año y Balanceo Año. Como se muestra en la siguiente ilustración.

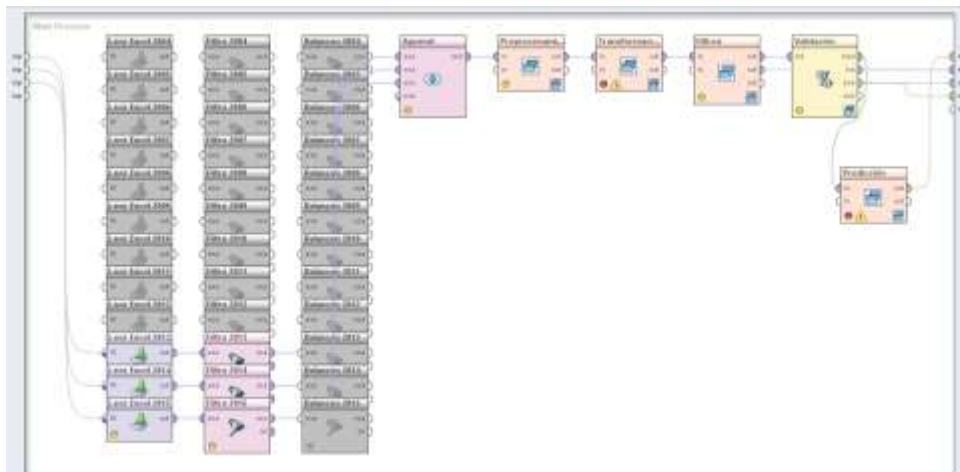


Ilustración 27: Propuesta 1. Fuente: Elaboración Propia

Los resultados fueron más imprecisos que el seleccionado en el proyecto.

5.3.5.1.2 Propuesta 2: Prof. Luciano Villarroel.

Se solicitó realizar los siguientes 18 casos, todos ellos considerando 3 años seguidos los cuales componen un dataset, éstos son utilizados para predecir el cuarto año, repitiendo esta acción 9 veces con la información balanceada por año y el mismo número de casos sin balancear, término descrito anteriormente. Luego cada caso genera una matriz de confusión, la cual muestra los valores acertados tanto de los estudiantes que permanecerán y los que desertarán. Para finalmente promediar estas 9 matrices y obtener una matriz de confusión total.

		2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Caso 1	Con balanceo		Dataset 1										
Caso 2	Con balanceo			Dataset 2									
Caso 3	Con balanceo				Dataset 3								
Caso 4	Con balanceo					Dataset 4							
Caso 5	Con balanceo						Dataset 5						
Caso 6	Con balanceo							Dataset 6					
Caso 7	Con balanceo								Dataset 7				
Caso 8	Con balanceo									Dataset 8			
Caso 9	Con balanceo										Dataset 9		
Caso 10	Sin balanceo		Dataset 1										
Caso 11	Sin balanceo			Dataset 2									
Caso 12	Sin balanceo				Dataset 3								
Caso 13	Sin balanceo					Dataset 4							
Caso 14	Sin balanceo						Dataset 5						
Caso 15	Sin balanceo							Dataset 6					
Caso 16	Sin balanceo								Dataset 7				
Caso 17	Sin balanceo									Dataset 8			
Caso 18	Sin balanceo										Dataset 9		

Ilustración 28: Casos de Prueba, Propuesta 2. Fuente: Elaboración Propia

Se confeccionó en RapidMiner un gran proceso, el cual genera 9 matrices de confusión las que se encuentran balanceadas y para obtener las no balanceadas se deben apagar las cajas de la cuarta columna de la siguiente ilustración.



Ilustración 29: Propuesta 2. Fuente: Elaboración Propia

Los resultados fueron más imprecisos que el seleccionado en el proyecto.

CAPÍTULO 6: PROPUESTA DE APOYO TECNOLÓGICO

6.1 Especificación de Requerimientos

Esta sección incluye los requerimientos que inciden directamente al uso del componente de apoyo tecnológico. En una primera instancia se indicarán los requerimientos funcionales y posteriormente los no funcionales.

6.1.1 *Requerimientos Funcionales*

- Los archivos B y E de postulación DEMRE, deben cargarse a Tableau, en formato de hojas Excel.
- La lectura de la base de datos de matrícula y becas debe realizarse a través de consultas en SQL Navigator.
- Los cruces de información se realizarán mediante Tableau.
- Se cargará un solo archivo de datos en formato Excel a RapidMiner, donde se ejecutará el modelo construido previamente.
- La actualización del reporte debe hacerse a través del complemento “Import Sheet” y utilizar la opción de Update All Conexions en la hoja de cálculo de Google.
- El listado de estudiantes que podrían desertar se publicará en el portal SIGA.
- El reporte debe tener 4 columnas RUT, Apellido Paterno, Apellido Materno y Nombres del estudiante.

6.1.2 *Requerimientos No Funcionales*

- La solución debe adaptarse al formato y color de los CSS de las páginas web institucionales.
- La carga de datos se realiza los últimos días de enero, cuando se disponga de toda la matrícula de estudiantes nuevos.
- Se dará acceso a funcionarios pertenecientes al Departamento de Pregrado, Decanatos y Secretarías de Estudios.

6.2 Arquitectura Tecnológica

El presente proyecto busca entregar una solución económica, amigable y que pueda ser implementado por agentes del negocio y no necesariamente técnicos, para posteriormente replicarla en diversas áreas que necesiten desplegar información para la toma de decisiones y que el área de negocios pueda ser lo más independiente del área de TI en la creación de sus propios indicadores. Se utilizaron principalmente las siguientes 6 herramientas tecnológicas: SQL Navigator, Tableau Desktop, Trello, Rapid Miner, Excel y el Sistema de Información para la Gestión Académica (SIGA) que

está compuesto por Google Sites y Awesome Table y despliega información en PDF o formato Tableau Reader.

Cabe destacar que esta arquitectura es una solución para poder ser implementada en el corto plazo, la cual se montará posteriormente en el Sistema de Indicadores de Gestión Integrados, proyecto que está siendo desarrollado en conjunto por la Vicerrectoría Económica y de Gestión Institucional, Vicerrectoría de Asuntos Académicos y el Consejo de Evaluación.

Se han tenido reuniones donde se ha valorado el trabajo realizado y en ningún momento se considera como un doble trabajo, ya que las definiciones y pruebas se pueden validar con el proceso 2017.

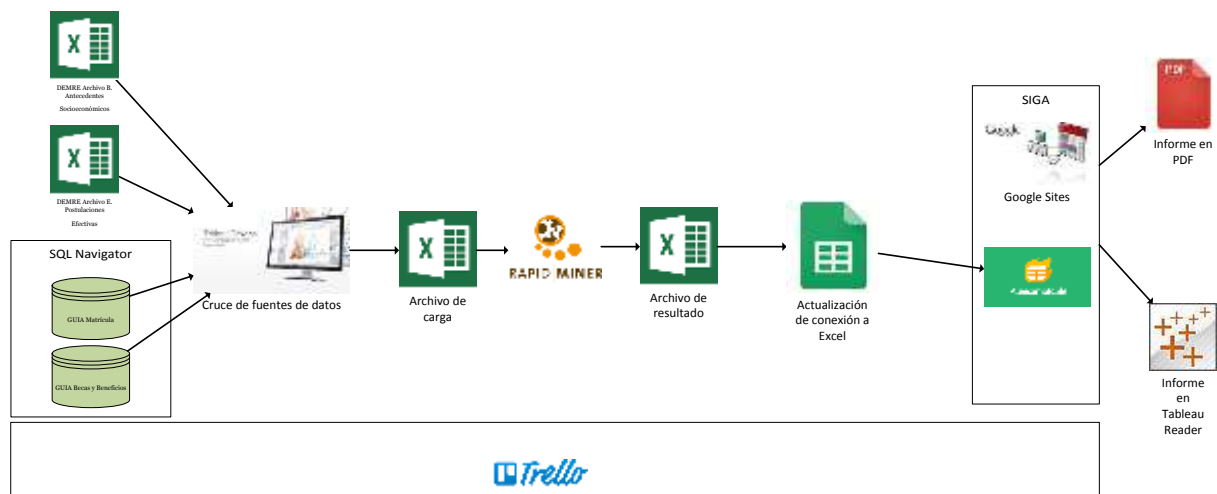


Ilustración 30: Arquitectura Tecnológica. Fuente: Elaboración propia.

6.2.1 SQL Navigator

La Universidad de Chile posee una base de datos central y un par de “vistas en Excel” que le permiten al usuario de nivel de negocio acceder a la información. A lo largo del proyecto se ha podido verificar que dichas vistas tienen algunos “detalles” lo cuales deben ser corregidos, por lo cual se optó por obtener la información desde el origen, para ello se utilizó SQL Navigator, donde se modificaron algunas consultas a la base de datos que son obtenidas los días 5 del mes siguiente.

Se extrae información de matrícula y de becas.

6.2.2 Trello

Para tener un mejor ordenamiento de las tareas a realizar, se ha utilizado Trello, la cual es una herramienta en línea, que permite la creación de tareas y Checklist, los cuales pueden ser revisados y modificados simultáneamente en el computador,

smartphone o tablet. Se han documentado todos los pasos para la actualización de los archivos bases que son cargados posteriormente en Tableau. Se utiliza transversalmente durante todo el proceso.

- Actualizar Base GUIA Matrícula
- Actualizar Postulación Archivo B (DEMRE)
- Actualizar Postulación Archivo E (DEMRE)
- Actualizar Becas a través de SQL Navigator
- Actualizar “Awesome Table” en Google Sites
- Otorgar privilegios a usuarios en SIGA

6.2.3 *Tableau Desktop*

Esta herramienta es utilizada dos veces, la primera para realizar cruces de información y así obtener el archivo que necesita RapidMiner y la segunda para elaborar los informes de seguimiento.

6.2.4 *Google Sites*

La Universidad de Chile tiene un contrato con Google que implica la disponibilidad de todas las herramientas del paquete de “Google Apps for Work”, dentro de las cuales se encuentra Google Sites, con el cual se ha construido el Sistema de Información para la Gestión Académica (SIGA), donde se otorgan los permisos de acceso específicos según nivel de detalle de la información.

Se crean carpetas en Google Drive, la cual permite sincronizar los permisos con el portal creado en Google Sites.

Para el acceso de usuarios se debe seguir cada paso del Checklist disponible en Trello.

6.2.5 *Integraciones con otros sistemas*

A modo de complemento, una vez obtenido el listado de estudiantes que podrían desertar, se puede buscar la información adicional en el Portafolio del Estudiante, el cual el autor ha sido parte de las definiciones y concepción.

6.3 Diseño de la Aplicación

A continuación se detallará el diseño de la aplicación a través de la ilustración anterior, donde se agregan los casos de uso para ver su radio de acción.

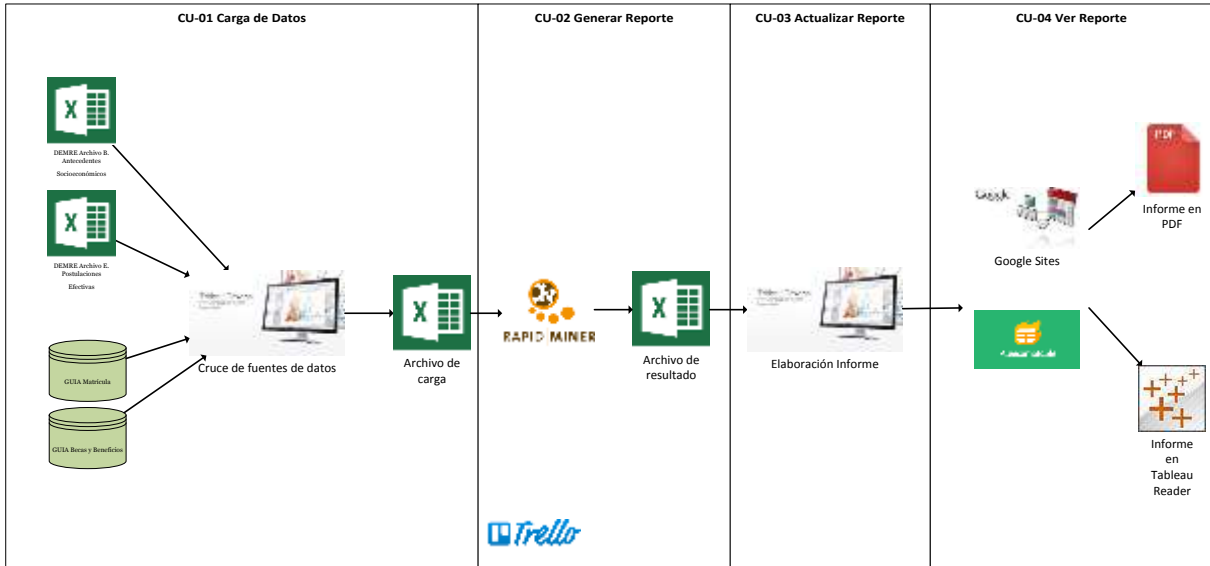


Ilustración 31: Arquitectura Tecnológica con Casos de Usos. Fuente: Elaboración propia.

6.3.1 Casos de Uso

Los casos de uso para el proyecto son 4. A continuación se dará una breve descripción y el detalle se puede ver en el anexo 1.

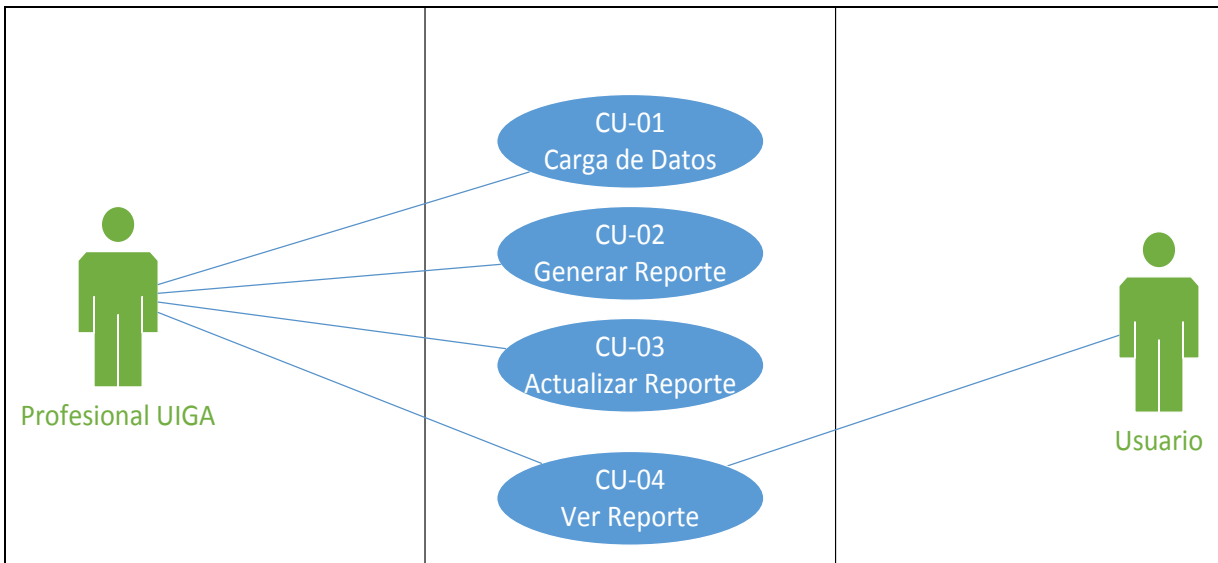


Ilustración 32: Casos de Usos. Fuente: Elaboración propia.

6.3.1.1 CU-01 Carga de Datos

Profesional UIGA ejecuta los ETL⁷ y carga los datos a Tableau Desktop.

6.3.1.2 CU-02 Generar Reporte

Profesional UIGA genera el reporte con los datos cargados en el punto anterior, para esto ejecuta RapidMiner el cual genera un archivo Excel.

6.3.1.3 CU-03 Actualizar Reporte

Profesional UIGA actualiza el reporte desde hoja de cálculo de Google, la cual lee el archivo Excel generado en el punto anterior.

6.3.1.4 CU-04 Ver Reporte

Usuario ingresa al portal SIGA y busca el listado de estudiantes que podrían desertar.

⁷ ETL: Proceso de extracción, transformación y carga de datos.

CAPÍTULO 7: EVALUACIÓN DEL PROYECTO

El proyecto ha traído diversos beneficios, entre ellos existen algunos que no estaban presupuestados como los que se muestran a continuación:

7.1 Beneficios colaterales del proyecto

- Creación del Sistema de Información para la Gestión Académica (SIGA) el cual permite realizar un seguimiento académico a los programas de estudios llegando al detalle de los estudiantes, implementado para todos programas de Doctorado, Magíster, Título de Profesional Especialista y Curso de Especialización de Postítulo, en las áreas de Postulación, Matrícula y Graduación, obteniendo indicadores como tasas de retención, tasas de graduación, uso de sistema de postulación en línea, tiempos de permanencia y reportes de corrección de información.
- Creación de formularios de registro y reportes estandarizados para áreas que no existe apoyo tecnológico. Caso de éxito: Encuesta de Actividades a Estudiantes PACE.
- Creación de comparador de modelos en hojas de Cálculo de Google.
- Creación de plataforma de seguimiento financiero y de avance de proyectos.

7.2 Definición de Beneficios Y Costos

El horizonte del proyecto es de 3 años, principalmente por la Reforma estudiantil.

El arancel promedio de los estudiantes del año 2017 es de \$3.524.079

	2006	2007	2008	2009	2010	2011	2012
1%	\$ 21.996.000	\$ 43.617.600	\$ 66.580.800	\$ 90.199.200	\$ 113.880.000	\$ 121.336.800	\$ 130.509.600
5%	\$ 109.980.000	\$ 218.088.000	\$ 332.904.000	\$ 450.996.000	\$ 569.400.000	\$ 606.684.000	\$ 652.548.000
10%	\$ 219.960.000	\$ 436.176.000	\$ 665.808.000	\$ 901.992.000	\$ 1.138.800.000	\$ 1.213.368.000	\$ 1.305.096.000
100%	\$ 2.199.600.000	\$ 4.361.760.000	\$ 6.658.080.000	\$ 9.019.920.000	\$ 11.388.000.000	\$ 12.133.680.000	\$ 13.050.960.000
	705	1398	2134	2891	3650	3889	4183
Estudiantes que desertan	705	693	736	757	759	944	987
		705	693	736	757	759	944
			705	693	736	757	759
				705	693	736	757
					705	693	736

Ilustración 33: Estudiantes que desertan en Pregrado por año. Fuente: Elaboración propia.

El cálculo que se ha realizado es que cada año existen estudiantes que desertan y esos se encuentran en la 6ta línea (2006:705, 2007:693, 2008:736, etc), por este concepto se deja de percibir los ingresos arancelarios de estos estudiantes desertados por los años que quedan según la duración del programa, es por esto que se ha

sumado los estudiantes desertados por los siguientes años. Cabe destacar que el año 2009 es un buen año para considerar ya que tiene 4 años de acumulación de estudiantes desertados, lo cual indica que se dejan de percibir 2891 ingresos que equivalen a \$9.019.920.000, si a partir de esto se logra retener el 1% de estos estudiantes (no es lo mismo que 1% de retención), se percibe \$90.199.200 sólo en un año.

7.2.1 *Inversión Inicial del Proyecto*

El proyecto ha tenido cambios desde que fue formulado inicialmente, en su oportunidad requería una mayor inversión inicial, sin embargo hoy se encuentran disponibles nuevas tecnologías que permiten una independencia con respecto al área de tecnologías de información, lo que a su vez hace que el proyecto sea elaborado en una unidad sin mayores inversiones. A continuación se detalla la inversión inicial del proyecto que son las reuniones de coordinación y puesta en marcha.

Tabla 10: Inversión Inicial del Proyecto

Stakeholder	Remuneración	% de dedicación	Remuneración en Proyecto
Vicerrectora de Asuntos Académicos	\$6.800.000	5%	\$340.000
Directora Departamento de Pregrado	\$4.200.000	5%	\$210.000
Jefe Área Gestión Académica Pregrado	\$2.400.000	10%	\$240.000
Coordinador Unidad Gestión de Proyectos Académicos	\$2.800.000	20%	\$560.000
Coordinador Académico Pregrado	\$1.100.000	25%	\$275.000
Jefe Unidad de Información para Gestión Académica	\$1.500.000	35%	\$525.000
Mensual			\$2.150.000
Total 4 meses			\$8.600.000

7.2.2 *Costos de mantención*

La aplicación necesita actualizaciones mensuales que requieren muy poco tiempo.

Tabla 11: Costos de mantención

Stakeholder	Remuneración	% de dedicación	Remuneración en Proyecto
Jefe Unidad de Información para Gestión Académica	\$1.500.000	5%	\$75.000
Mensual			\$75.000
Total Año			\$900.000

7.2.3 Costos Directos por Profesionales

Las acciones que se realizarán a cada posible desertor, aún no han sido elaboradas por los expertos, sin embargo se consideran horas de profesionales para trabajar en ello de acuerdo a la magnitud de estudiantes que desertan por Unidades Académicas, es decir, aquella Unidad Académica que tiene más estudiantes desertores, tendrá más horas de profesionales para lograr su retención.

Tabla 12: Costos Directos por Profesionales

Profesional	Estudiantes Promedio Deserción	HH	Remuneración en Proyecto Año 1	HH	Remuneración en Proyecto Año 2	HH	Remuneración en Proyecto Año 3
AU ⁸	123	44	\$1.000.000	88	\$2.000.000	132	\$3.000.000
AR ⁹	54	32	\$727.273	64	\$1.454.545	96	\$2.181.818
CS ¹⁰	84	44	\$1.000.000	88	\$2.000.000	132	\$3.000.000
AG ¹¹	35	32	\$727.273	64	\$1.454.545	96	\$2.181.818
FM ¹²	63	44	\$1.000.000	88	\$2.000.000	132	\$3.000.000
FR ¹³	7	12	\$272.727	24	\$545.455	36	\$818.182
QF ¹⁴	53	32	\$727.273	64	\$1.454.545	96	\$2.181.818
SO ¹⁵	38	32	\$727.273	64	\$1.454.545	96	\$2.181.818
VT ¹⁶	18	12	\$272.727	24	\$545.455	36	\$818.182
DE ¹⁷	29	22	\$500.000	44	\$1.000.000	66	\$1.500.000
EN ¹⁸	69	44	\$1.000.000	88	\$2.000.000	132	\$3.000.000
FH ¹⁹	54	32	\$727.273	64	\$1.454.545	96	\$2.181.818
MD ²⁰	80	44	\$1.000.000	88	\$2.000.000	132	\$3.000.000
OD ²¹	16	12	\$272.727	24	\$545.455	36	\$818.182
AP ²²	13	12	\$272.727	24	\$545.455	36	\$818.182
CI ²³	16	12	\$272.727	24	\$545.455	36	\$818.182
	Mensual		\$10.500.000		\$21.000.000		\$31.500.000
	Total Año		\$126.000.000		\$252.000.000		\$378.000.000

⁸ AU: Facultad de Arquitectura y Urbanismo

⁹ AR: Facultad de Artes

¹⁰ CS: Facultad de Ciencias

¹¹ AG: Facultad de Ciencias Agronómicas

¹² FM: Facultad de Ciencias Físicas y Matemáticas

¹³ FR: Facultad de Ciencias Forestales y Conservación de la Naturaleza

¹⁴ QF: Facultad de Ciencias Químicas y Farmacéuticas

¹⁵ SO: Facultad de Ciencias Sociales

¹⁶ VT: Facultad de Ciencias Veterinarias y Pecuarias

¹⁷ DE: Facultad de Derecho

¹⁸ EN: Facultad de Economía y Negocios

¹⁹ FH: Facultad de Filosofía y Humanidades

²⁰ MD: Facultad de Medicina

²¹ OD: Facultad de Odontología

²² AP: Instituto de Asuntos Públicos

²³ CI: Instituto de Comunicación e Imagen

El proyecto no considera comprar nuevos servidores, ya que esta información ya se encuentra utilizándolos, lo cual existirá con o sin proyecto.

7.3 Flujo de Caja

Se consideró la tasa social de descuento del Ministerio de Desarrollo Social

$$TSD = \alpha * tp + \beta * q + \gamma * CMgx$$

Donde

tp: Tasa de Ahorro Interno: Se utiliza tasa interés de captación (indicada por Banco Central). De acuerdo a la opinión del panel de expertos, la mejor proyección disponible para esta tasa en el mediano-largo plazo corresponde al promedio de los últimos 13 años (2000-2012); es decir, 4,07%.

q: Rentabilidad de la inversión privada: a partir de una muestra de empresas que presentan sus Estados de Resultados a la SVS, se estimó la rentabilidad como la razón entre Resultado Operacional/Activo Operacional para el periodo 1990 – 2009, resultando una rentabilidad promedio de 7,5% anual.

CMgx: Costo marginal del endeudamiento externo: para estimar el costo marginal del endeudamiento externo se requiere conocer el costo medio del endeudamiento externo (*CMex*) y la elasticidad del ahorro externo con respecto a éste último. A su vez, para estimar el *CMex* se requiere contar con datos sobre: tasa interés internacional, riesgo país, variación porcentual del tipo de cambio nominal e inflación esperada.

Tabla 13: Tasa Social de Descuento. Fuente: Ministerio de Desarrollo Social

Escenario	α	β	γ	Tasa ahorro interno	Rendimiento inversión privada	Costo Marginal Endeudamiento Externo	TSD
1	0.45%	87.41%	12.14%	4.07	7.48	3.29	6.96
2	0.45%	87.41%	12.14%	4.07	7.48	5.49	7.22
3	21.08%	69.29%	9.63%	4.07	7.48	3.29	6.36
4	21.08%	69.29%	9.63%	4.07	7.48	5.49	6.57
5	0.20%	94.44%	5.37%	4.07	7.48	3.29	7.25
6	0.20%	94.44%	5.37%	4.07	7.48	5.49	7.37
7	10.53%	84.65%	4.81%	4.07	7.48	3.29	6.92
8	10.53%	84.65%	4.81%	4.07	7.48	5.49	7.02
9	0.45%	87.98%	11.58%	4.07	7.48	3.29	6.98
10	0.45%	87.98%	11.58%	4.07	7.48	5.49	7.23
11	21.10%	69.73%	9.18%	4.07	7.48	3.29	6.38
12	21.10%	69.73%	9.18%	4.07	7.48	5.49	6.58
13	0.20%	94.70%	5.10%	4.07	7.48	3.29	7.26
14	0.20%	94.70%	5.10%	4.07	7.48	5.49	7.37
15	10.51%	84.92%	4.57%	4.07	7.48	3.29	6.93
16	10.51%	84.92%	4.57%	4.07	7.48	5.49	7.03

Se utilizará el escenario 8 con una tasa social de descuento del 7,02%.

Se consideró como supuesto que el primer año se puede retener el 2% de los estudiantes que podrían desertar, los cuales son aproximadamente 58, para ello se consideró ciertas horas de trabajo de profesionales. Para el segundo año se espera duplicar esa cantidad, poder retener el 4%, algo así como 116 estudiantes utilizando el doble de horas de profesionales. Para el tercer año se espera un 6%, el cual corresponde aproximadamente 174 estudiantes con el triple de horas de profesionales contratados en el primer año.

Tabla 14: Flujo de Caja del Proyecto

	Año 0	Año 1	Año 2	Año 3
Ingresos		2%	4%	6%
Ingresos por estudiantes recuperados		\$180.960.000	\$361.920.000	\$542.880.000
Costos Directos				
Profesionales		-\$126.000.000	-\$252.000.000	-\$192.000.000
Costos Mantención		-\$900.000	-\$900.000	-\$900.000
Resultado Operacional		\$54.060.000	\$109.020.000	\$349.980.000
Gastos en Administración				
Capacitación		-\$20.000.000	-\$20.000.000	-\$20.000.000
Resultado Operacional No		\$ 34.060.000	\$89.020.000	\$329.980.000
Inversión Inicial	-\$8.600.000			
Flujo de caja	-\$8.600.000	\$34.060.000	\$89.020.000	\$329.980.000

Tasa Social de Descuento 7,02%

VAN \$370.161.270

TIR 547%

El proyecto tiene una Valor Actual Neto de \$370.161.270, considerando una inversión de \$8.600.000 y una Tasa Interna de Retorno de 547% es un proyecto muy rentable.

7.4 Análisis de Sensibilidad

7.4.1 Análisis de Sensibilidad 1

Se consideró como supuesto que el primer año se puede retener el 1,04% de los estudiantes que podrían desertar, los cuales son aproximadamente 30, para ello se consideró ciertas horas de trabajo de profesionales. Para el segundo año se espera duplicar esa cantidad, poder retener el 2,08%, algo así como 60 estudiantes utilizando el doble de horas de profesionales. Para el tercer año se espera un 3,11%, el cual corresponde aproximadamente 90 estudiantes con el triple de horas de profesionales contratados en el primer año.

Tabla 15: Flujo de Caja del Análisis de Sensibilidad 1

	Año 0	Año 1	Año 2	Año 3
Ingresos		1,04%	2,08%	3,11%
Ingresos por estudiantes recuperados		\$93.600.000	\$187.200.000	\$280.800.000
Costos Directos				
Profesionales		-\$126.000.000	-\$252.000.000	\$192.000.000
Costos Mantención		-\$900.000	-\$900.000	-\$900.000
Resultado Operacional		.\$33.300.000	-\$65.700.000	\$87.900.000
Gastos en Administración				
Capacitación		-\$20.000.000	-\$20.000.000	-\$20.000.000
Resultado Operacional No		-\$53.300.000	-\$85.700.000	\$67.900.000
Inversión Inicial	-\$8.600.000			
Flujo de caja	-\$8.600.000	-\$53.300.000	-\$85.700.000	\$67.900.000

Tasa Social de Descuento 7,02%

VAN -\$77.833.942

TIR -43%

El proyecto tiene un Valor Actual Neto de -\$77.833.942, considerando una inversión de \$8.600.000 y una Tasa Interna de Retorno de -43%.

7.4.2 Análisis de Sensibilidad 2

Se consideró como supuesto que el primer año se puede retener el 3% de los estudiantes que podrían desertar, los cuales son aproximadamente 87, para ello se consideró ciertas horas de trabajo de profesionales. Para el segundo año se espera duplicar esa cantidad, poder retener el 6%, algo así como 174 estudiantes utilizando el doble de horas de profesionales. Para el tercer año se espera un 9%, el cual corresponde aproximadamente 261 estudiantes con el triple de horas de profesionales contratados en el primer año.

Tabla 16: Flujo de Caja del Análisis de Sensibilidad 2

	Año 0	Año 1	Año 2	Año 3
Ingresos		3%	6%	9%
Ingresos por estudiantes recuperados		\$271.440.000	\$542.880.000	\$814.320.000
Costos Directos				
Profesionales		-\$126.000.000	-\$252.000.000	\$192.000.000
Costos Mantención		-\$900.000	-\$900.000	-\$900.000
Resultado Operacional		\$144.540.000	\$289.980.000	\$621.420.000
Gastos en Administración				
Capacitación		-\$20.000.000	-\$20.000.000	-\$20.000.000
Resultado Operacional No		\$124.540.000	\$269.980.000	\$601.420.000
Inversión Inicial	-\$8.600.000			
Flujo de caja	-\$8.600.000	\$124.540.000	\$269.980.000	\$601.420.000

Tasa Social de Descuento 7,02%

VAN \$834.156.310

TIR 1562%

El proyecto tiene un Valor Actual Neto de \$834.156.310, considerando una inversión de \$8.600.000 y una Tasa Interna de Retorno de 1562%.

CAPÍTULO 8: CONCLUSIONES

En el presente capítulo se presentan las conclusiones del proyecto, principalmente en 4 líneas, en cuanto a la metodología, ejecución del proyecto, resultados del proyecto y trabajos futuros.

8.1 Conclusiones sobre metodología

La arquitectura de procesos de la metodología de Ingeniería de Negocios, permitió entender de mejor forma el problema a resolver y lograr ubicarlo dentro de la institución, observando las interacciones entre las distintas áreas sin dejar de lado los procesos de planificación, creación y mantención de estado que son los que acompañan a la cadena de valor.

Todas las actividades de la metodología son un aporte para la elaboración del proyecto y permiten llevar desde la estrategia hasta los procesos detallados, pasando por el posicionamiento estratégico, mapa estratégico y modelo de negocios.

El proceso de extracción del conocimiento, es bastante detallado y fue una guía ilustrativa de cómo obtener un modelo predictivo. Desde la selección correcta de las variables, haciendo transformaciones y filtros, hasta aplicar los distintos algoritmos y cómo deben ser evaluados para elegir el mejor modelo.

8.2 Conclusiones sobre la ejecución del proyecto

A lo largo de la ejecución de este proyecto se han obtenido beneficios que no estaban estimados, como la creación del portal SIGA, los formularios de registro con posteriores reportes de análisis y finalmente el comparador de modelos.

La construcción del datamart, recopiló información de distintas fuentes de datos, por ejemplo en matrícula existen atributos interesantes de la vida del estudiante como las situaciones académicas, que son estados del estudiantes en el transcurso del año, tales como postergación, renuncia o eliminación académica, entre otras, las cuales podrían haber permitido caracterizar mejor al estudiante, sin embargo, el registro de éstas generalmente se hace a fin de año, por lo que no sirve para el modelo predictivo generado en el proyecto.

Se construye un repositorio de información el cual permite hacer cruces de información que no había sido posible realizarlos por otros medios, transformándose la Unidad de Información para la Gestión Académica en la única responsable de entregar información referente a tasas de titulación y tiempos de permanencia.

Durante la elaboración del proyecto se han utilizado varias herramientas tecnológicas como Trello, Google Sites, Tableau y planillas de cálculo de Google, lo que ha permitido crear un modelo de trabajo el cual ha sido replicado en otras iniciativas por el autor, tales como análisis de encuestas de estudiantes PACE, seguimiento financiero y avance de actividades de proyectos FADOP²⁴. Todo se basa en una planilla de cálculo de Google en la cual se llena la información que posteriormente se quiere mostrar, se utiliza Tableau para leer el archivo y diagramarlo en reportes dinámicos los cuales se disponibilizan en un Google Sites, todos estos pasos son documentados en Checklist en Trello, para una posterior modificación o replicarlo en otra área.

8.3 Conclusiones sobre resultados del proyecto

Se ha diseñado una estrategia para evitar las deserciones, la cual en grandes rasgos es utilizar la información proveniente del proceso de postulación, becas y matrícula de estudiantes de primer año de programas de pregrado de la Universidad de Chile, aplicarles un modelo de redes neuronales que entrega un listado de los estudiantes que podrían desertar, los cuales son acompañados por Profesionales que son contratados en las unidades académicas con el fin de hacerles seguimiento y evitar que deserten.

Dentro de los 4 algoritmos que se utilizaron, Árboles de Decisión, Regresión Logística, Support Vector Machines y Redes Neuronales, estas últimas son las que tienen mejor comportamiento predictor.

En la evaluación económica se han realizado análisis de sensibilidad, donde se aprecia que el proyecto es muy rentable, ya que los beneficios sobre los costos son notablemente mayores. El punto donde el VAN se hace cero, es con 35 estudiantes que se puedan conseguir que no deserten, lo cual hace que el proyecto sea alcanzable.

El modelo obtenido predice el 74,32% de los estudiantes que van a desertar en su primer año de carrera, teniendo un gran impacto en cuanto a procesos y beneficios económicos.

Las variables que inciden en el modelo se pueden clasificar en 5 áreas:

- Educación Secundaria: Grupo Dependencia y NEM.
- PSU: Prueba de Lenguaje y Comunicación, Prueba de Matemáticas, Puntaje Ponderado y Preferencia.
- Socioeconómico: Ingreso Bruto y si tiene o no beca.

²⁴ FADOP: Fondo de Apoyo a la Docencia de Pregrado

- Personal: Género
- Grupo Familiar: Número de integrantes del grupo familiar y si están vivos sus padres.

El proyecto tiene una gran ventaja sobre otros de deserción estudiantil universitaria, ya que el objetivo principal es predecir tempranamente las deserciones de primer año, debe ejecutarse el modelo antes de que comiencen las clases, esto es a fines del mes de enero, en cambio existen en la universidad por lo menos dos proyectos de deserción, el primero se activa la alerta en agosto, una vez terminado el primer semestre y considerando a los estudiantes que han reprobado una asignatura, y el segundo en el segundo año, luego de tener la historia de un año para el estudiante.

8.4 Trabajo Futuros

Se ha detectado que existen muchos registros con errores de digitación y con falta de información, para mejorar aquello se debe mejorar la calidad de la información registrada en los sistemas de matrícula, para ello se ha hecho un piloto, pero debe ponerse en práctica, esto consiste en la elaboración de reportes que detecten esos errores y los publique en el Sistema de Información para la Gestión Académica. Lo anterior permitiría hacer análisis con otras variables y con mejor calidad.

Para armar la historia de datos de estudiantes que desertaron, se considera como desertor a aquellos que no se matriculan en segundo año, lo que contempla a quienes han renunciado o han sido eliminados, además existe un grupo de estudiantes que simplemente no se matriculan y no queda un registro de por qué no lo hacen, para ello se propone que al inicio del segundo año se ingrese la situación académica de todos los estudiantes, lo cual el día de hoy no se realiza. Lo anterior permitiría describir mejor al estudiante que deserta.

El proyecto ha sido desarrollado con tecnologías que han permitido la independencia con el área de tecnologías de la información, lo cual hace que la ejecución sea manual, como esto ocurre una vez en el año se ha presentado como una buena solución, sin embargo, desde a fines del año 2016, se está trabajando conjuntamente con la Unidad de Análisis Institucional.

Realizar una evaluación real del modelo, donde se pueda comprobar cómo efectivamente fue la predicción versus la realidad. Para realizar esto es necesario que pase un año, por lo que los estudiantes que ingresaron el año 2016 se podrán medir al 31 de diciembre del 2017.

Considerar variables curriculares, tales como carga académica, porcentaje de asignaturas aprobadas, promedio de notas, las cuales deben ser medidas a mediados de año. Esto permitiría hacer un contraste entre el modelo actual y el ampliado.

En el presente proyecto se ha desarrollado un modelo que predice la deserción en el primer año, sin embargo se puede ampliar y generar un modelo que pueda predecir la deserción en los años posteriores.

Generar alertas de estudiantes que se encuentran en el listado que contengan información de promedios de notas o alguna situación académica que aumente aún más la probabilidad de desertar.

CAPÍTULO 9: BIBLIOGRAFÍA

Barros V, O. (2008). Diseño Integrado de Negocios, Procesos y Aplicaciones TI. *Santiago, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, Departamento de Ingeniería Industrial.*

Bean, J. P. (1980). Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in higher education, 12(2)*, 155-187.

Betancourt, G. A. (2005). Las máquinas de soporte vectorial (svms). *Scientia et technica, 1(27)*.

Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2008). Análisis de los factores asociados a la deserción estudiantil en la Educación Superior: un estudio de caso
Analysis of the Factors Associated with the Drop-out Rate of Students in Higher Education: a Case Study. *Revista de Educación, 345*, 255-280.

Castaño, E., Gallón, S., Gómez, K., & Vásquez, J. (2009). Deserción estudiantil universitaria: una aplicación de modelos de duración. *Lecturas de economía, 60(60)*, 39-65.

Castro, J. A. F., & González, D. M. S. (1993). *Redes neuronales: algoritmos, aplicaciones y técnicas de programación.*

Díaz Peralta, C. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena. *Estudios pedagógicos (Valdivia), 34(2)*, 65-86.

Ethington, C. A. (1990). A psychological model of student persistence. *Research in higher Education, 31(3)*, 279-293.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine, 17(3)*, 37.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM, 39(11)*, 27-34.

Fernández, A. J. J. (1995). *Análisis de regresión logística.* Centro de Investigaciones Sociológicas (CIS).

Hax, A. C., & Wilde, D. L. (1999). The delta model: adaptive management for a changing world. *Sloan Management Review, 40(2)*, 11.

Hernández Orallo, J., Ramírez Quintana, M. J., & Ferri Ramírez, C. (2004). Introducción a la Minería de Datos. *Editorial Pearson Educación SA, Madrid*.

Himmel, E. (2002). Modelos de análisis de la deserción estudiantil en la educación superior. *Revista Calidad en la educación, 17(2)*, 91-108.

López, C. P. (2007). *Minería de datos: técnicas y herramientas*. Editorial Paraninfo.
Ministerio de Educación (2012). Deserción en la educación superior en Chile. *Serie Evidencias*, Año 1, N° 9.

Misión y visión. Recuperado de <http://www.uchile.cl/portal/presentacion/institucionalidad/39635/mision-y-vision>

Pascarella, E. T., & Terenzini, P. T. (1980). Predicting freshman persistence and voluntary dropout decisions from a theoretical model. *The Journal of Higher Education, 51(1)*, 60-75.

Rodríguez, A. A., & Vindas, M. A. S. (2011). La deserción estudiantil en la educación superior: el caso de la Universidad de Costa Rica. *Actualidades investigativas en educación, 5(4)*.

Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange, 1(1)*, 64-85.

Stage, F. K. (1989). Motivation, academic and social integration, and the early dropout. *American Educational Research Journal, 26(3)*, 385-402.

Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. *Neural processing letters, 9(3)*, 293-300.

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of educational research, 45(1)*, 89-125.

Tinto, V. (1989). Definir la deserción: una cuestión de perspectiva. *Revista de educación superior, 71(18)*, 1-9.

Universidad de Chile (2014). *Anuario Universidad de Chile, Síntesis 2006-2013*. Andros Impresores.

Universidades. Recuperado de <http://www.mifuturo.cl/index.php/donde-y-que-estudiar/universidades>

Velásquez, J. D., Olaya, Y., & Franco, C. J. (2010). Predicción de series temporales usando máquinas de vectores de soporte. *Ingeniare. Revista chilena de ingeniería*, 18(1), 64-75.

Vélez, E. C., Gómez, S. G., Portilla, K. G., & Velásquez, J. V. (2006). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. *Lecturas de economía*, 65, 9.

CAPÍTULO 10: ANEXOS

10.1 Anexo 1: Casos de Uso

Tabla 17: Caso de Uso 1

Nombre caso de Uso:	Carga de Datos		
Código Caso de Uso:	CU-01		
Actor Primario:	Profesional UIGA		
Actores Secundarios:	Jefe UIGA		
Caso de Éxito o Flujo Normal de Operación			
Descripción:	Profesional UIGA ejecuta la actualización de información para cargar el Datamart		
Precondiciones:			
#	Acción	Reglas Negocio	Requerimientos GUI
1	El PROFESIONAL UIGA carga "Archivo B" en "Base Pregrado" que contiene toda la información.	Se ejecuta una vez que llega el archivo desde el DEMRE, aprox.12 de enero.	
2	El PROFESIONAL UIGA carga "Archivo E" en "Base Pregrado" que contiene toda la información.	Se ejecuta una vez que llega el archivo desde el DEMRE, aprox.12 de enero	
3	El PROFESIONAL UIGA carga "Matriculados" en archivo "Base Pregrado" que contiene toda la información.	Se ejecuta una vez que todos los estudiantes de 1er año se matriculan, aprox.25 de enero	
4	El PROFESIONAL UIGA ingresa a Tableau y actualiza la información.		
5	El PROFESIONAL UIGA exporta los datos		
Excepciones o Flujos Alternativos			
#	Acción	Reglas Negocio	Requerimientos GUI
1	El PROFESIONAL UIGA cierra la ventana.		
2	El SISTEMA cierra la aplicación.		

Tabla 18: Caso de Uso 2

Nombre caso de Uso:		Generar Reporte	
Código Caso de Uso:		CU-02	
Actor Primario:		Profesional UIGA	
Actores Secundarios:		Jefe UIGA	
Caso de Éxito o Flujo Normal de Operación			
Descripción:		Profesional UIGA genera reporte.	
Precondiciones:			
#	Acción	Reglas Negocio	Requerimientos GUI
1	El PROFESIONAL UIGA ingresa a RapidMiner		
2	El PROFESIONAL UIGA abre el Proyecto "Deserción"		
3	El PROFESIONAL UIGA ejecuta el Proceso		
Excepciones o Flujos Alternativos			
#	Acción	Reglas Negocio	Requerimientos GUI
1	El PROFESIONAL UIGA cierra la ventana.		
2	El SISTEMA cierra la aplicación.		

Tabla 19: Caso de Uso 3

Nombre caso de Uso:		Actualizar Reporte	
Código Caso de Uso:		CU-03	
Actor Primario:		Profesional UIGA	
Actores Secundarios:		Jefe UIGA	
Caso de Éxito o Flujo Normal de Operación			
Descripción:		El USUARIO actualiza el listado de estudiantes que podrían desertar.	
Precondiciones:			
#	Acción	Reglas Negocio	Requerimientos GUI
1	El USUARIO abre hoja de cálculo de Google.		
2	El USUARIO actualiza la conexión a archivo Excel de resultado de RapidMiner.		
Excepciones o Flujos Alternativos			
#	Acción	Reglas Negocio	Requerimientos GUI
1			

Tabla 20: Caso de Uso 4

Nombre caso de Uso:	Ingreso a SIGA		
Código Caso de Uso:	CU-04		
Actor Primario:	Usuario		
Actores Secundarios:			
Caso de Éxito o Flujo Normal de Operación			
Descripción:	El USUARIO ingresa al SISTEMA con su correo institucional y su clave		
Precondiciones:			
#	Acción	Reglas Negocio	Requerimientos GUI
1	El USUARIO ingresa a Página Web "www.uchile.cl/vaa/siga"		
2	El SISTEMA muestra Pantalla de Autenticación		
3	El USUARIO ingresa su nombre de usuario		
4	El USUARIO ingresa su contraseña de Cuenta Pasaporte UChile		
5	El USUARIO selecciona "Listado de estudiantes que podrían desertar"		
Excepciones o Flujos Alternativos			
#	Acción	Reglas Negocio	Requerimientos GUI
1	Ingreso de usuario y/o contraseña erróneo	Arroja mensaje "Usuario y/o Contraseña mal ingresados", volver a paso 3	

10.2 Anexo 2: Creación de Datamart

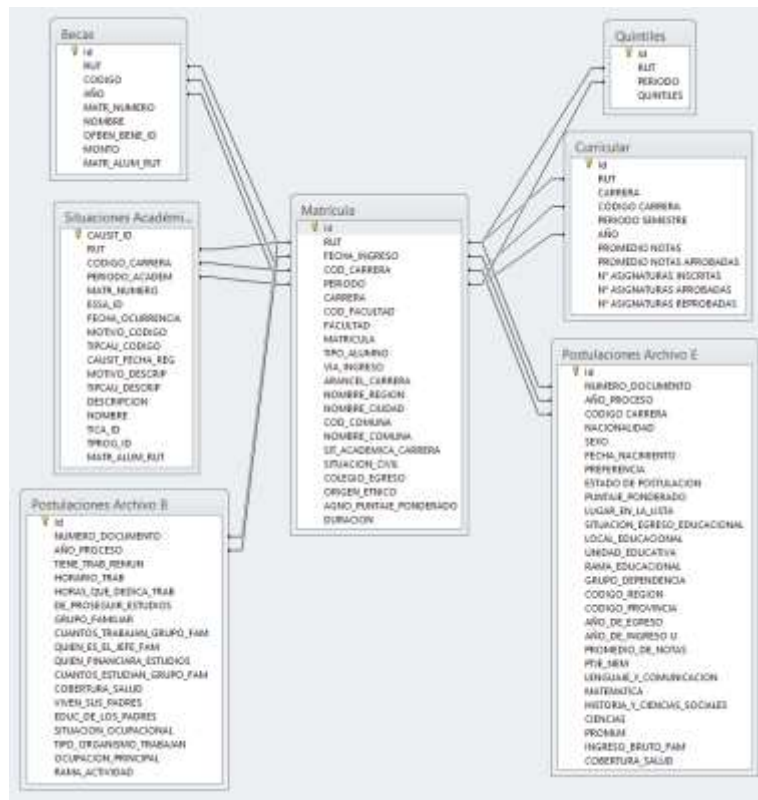


Ilustración 34: Datamart. Fuente: Elaboración propia.

10.2.1 Información de Becas

Se puede obtener la información actualizada al minuto de cada estudiante, para lo cual se debe ejecutar una consulta de base de datos que selecciona todos los campos que se listan a continuación:

Tabla 21: Información de becas

CAMPO	DESCRIPCIÓN
RUT	RUT del estudiante
MATRICULA	Número de matrícula del estudiante
OFBEN_BENE_ID	Código del beneficio
NOMBRE	Nombre del beneficio
FECHA_INICIO	Fecha de inicio del beneficio
MONTO	Monto del beneficio
CODIGO	Código del Programa de Formación
AÑO	Año del beneficio
Tablas utilizadas	AL_BENEFICIOS_ALUM
	AL_BENEFICIOS
	CU_CARRERAS

10.2.2 Información de Quintiles

A mediados de año la Dirección de Bienestar Estudiantil, entrego un archivo Excel con gran parte de los estudiantes nuevos con su respectiva marca de quintil, los que no se encuentran en el listado se clasifican como estudiantes sin quintil.

Los campos que posee, son:

Tabla 22: Información de Quintil

CAMPO	DESCRIPCIÓN
PERIODO	Año de información
RUT	RUT del estudiante
QUINTIL	Quintil socioeconómico del estudiante

10.2.3 Información de Matrícula

Se puede obtener la información actualizada al minuto de cada estudiante, para lo cual se debe ejecutar una consulta de base de datos que selecciona todos los campos que se listan a continuación:

Tabla 23: Información de Matrícula

CAMPO	DESCRIPCIÓN
PERIODO	Año de ocurrencia de la situación académica
RUT	RUT del estudiante
COD PROGRAMA	Código del Programa de Estudios
PROGRAMA	Nombre del Programa de Estudios
COD UNIDAD ACADEMICA	Código de Unidad Académica
UNIDAD ACADEMICA	Nombre de Unidad Académica
PATERNAL	Apellido paterno del estudiante
MATERNAL	Apellido materno del estudiante
NOMBRES	Nombres del estudiante
SEXO	Género del estudiante
MATRICULA	Número de matrícula del estudiante
FECHA INGRESO	Fecha de ingreso a la Universidad de Chile
TIPO ALUMNO	Indica la antigüedad del estudiante
VIA INGRESO	Vía de ingreso del estudiante las cuales son:
FECHA NACIMIENTO	Fecha de nacimiento del estudiante
NOMBRE REGION	Nombre de la Región donde vive el estudiante
NOMBRE COMUNA	Nombre de la Comuna donde vive el estudiante
NACIONALIDAD	Nacionalidad del estudiante
SIT ACADEMICA	Situación académica del estudiante
EGRESO EMEDIA	Año de egreso del colegio
PTJE PSU LENG Y COMU	Puntaje PSU Lenguaje y Comunicación
PTJE PSU MATEMATICA	Puntaje PSU Matemática
PTJE PSU HISTO Y CS	Puntaje PSU Historia y Ciencias
PTJE PSU CIENCIAS	Puntaje PSU Ciencias

PTJE PSU CIENCIAS BIOLOGIA	Puntaje PSU Ciencias Biología
PTJE PSU CIENCIAS FISICA	Puntaje PSU Ciencias Física
PTJE PSU CIENCIAS QUIMICA	Puntaje PSU Ciencias Química
PROMEDIO PAA PSU	Promedio PSU Lenguaje con PSU Matemática
PUNTAJE POND	Puntaje ponderado
PROCED EDUC	Procedencia educacional
PROMEDIO EM	Promedio de notas enseñanza media
PTJE NOTAS EM	Puntaje de notas enseñanza media
PREFERENCIA	Preferencia del estudiante
ORDEN LISTA	Número que quedó el estudiante en la lista de postulación
COLEGIO EGRESO	Nombre del colegio de egreso del estudiante
ORIGEN ETNICO	Origen étnico del estudiante
COMUNA ORIGEN	Comuna de donde proviene el estudiante
REGION ORIGEN	Región de donde proviene el estudiante
IDENTIFICADOR MATRICULA	Contiene RUT, código del programa y número de matrícula
IDENTIFICADOR GRADO	Contiene RUT y código del programa
ARANCEL	Arancel del programa del estudiante
E MAIL	E-mail del estudiante
SITUACION CIVIL	Situación civil del estudiante
REGION GFAM	Región del grupo familiar
Vistas utilizadas	SUG.V_ALUMNOS_HISTORICO
	SUG.V_ALUMNOS

10.2.4 Información de Situaciones Académicas

Esta información es vital para entender el motivo y/o causa de la deserción del estudiante, toda situación académica que tenga una causa y/o motivo ingresado aparecerá en esta consulta de base de datos, los campos extraídos son los siguientes:

Tabla 24: Información de Situaciones Académicas

CAMPO	DESCRIPCIÓN
CAUSIT_ID	Identificador Tipo de Causa
RUT	RUT del estudiante
MATRICULA	Número de matrícula del estudiante
COD_SIT_ACAD	Código situación académica
PERIODO	Año de ocurrencia de la situación académica
FECHA_OCURRENCIA	Fecha de ocurrencia de situación académica
COD_MOTIVO	Código del Motivo
COD_TIPO_CAUSA	Código de Tipo Causal
CAUSIT_FECHA_REG	Fecha de registro de Situación Académica
MOTIVO	Motivo de Situación Académica

TIPCAU_DESCRIP	Descripción Tipo de Causa
DESCRIPCION	Descripción de Situación Académica
COD_PROGRAMA	Código de Programa de Estudios
PROGRAMA	Nombre Programa de Estudios
TICA_ID	Tipo de Carrera
TPROG_ID	Tipo de Programa
Tablas utilizadas	CU_CARRERAS
	MC_CAUSAL_SITUACION
	MC_MOTIVO_CAU
	MC_TIPO_CAUSAL_SITUACION
	AL_EST_SITUACIONES_ACAD
CU_CARRERAS	

10.2.5 Archivo B: Inscritos, datos socioeconómicos y direcciones.

Tabla 25: Información de Archivo B

CAMPO	DESCRIPCIÓN
TIPO IDENTIFICACION	Tipo de documento
NUMERO DOCUMENTO	RUT del estudiante
AÑO PROCESO	Año de ingreso a la universidad
TIENE TRAB REMUN	¿Tiene trabajo remunerado?
HORARIO TRAB	¿Cuál es el horario de trabajo?
HORAS QUE DEDICA TRAB	Cantidad promedio de horas semanales dedicadas al trabajo.
DE PROSEGUIR ESTUDIOS	De proseguir estudios vivirá
GRUPO FAMILIAR	Cantidad de personas del grupo familiar
CUANTOS TRABAJAN GRUPO FAM	Cantidad de integrantes del grupo familiar que trabajan en forma remunerada:
QUIEN ES EL JEFE FAM	El jefe de familia es:
QUIEN FINANCIARA ESTUDIOS	Financiamiento principal y secundario
CUANTOS ESTUDIAN GRUPO FAM	Cantidad de integrantes del grupo familiar que estudian en la actualidad, en cada uno de los distintos niveles de enseñanza:
COBERTURA SALUD	Cobertura de salud
VIVEN SUS PADRES	Viven los padres del postulante:
EDUC DE LOS PADRES	Nivel educacional de los padres del postulante
SITUACION OCUPACIONAL	Situación ocupacional de los padres del postulante
TIPO ORGANISMO TRABAJAN	Tipo de organismo o empresa en que trabajan, o trabajaron por última vez, los padres del postulante:
OCUPACION PRINCIPAL	Ocupación principal actual, o última, de los padres del postulante:

RAMA ACTIVIDAD	Rama actividad económica en que trabajan los padres
CAMPO	DESCRIPCIÓN
TIPO IDENTIFICACION	Tipo de documento
NUMERO DOCUMENTO	RUT del estudiante
AÑO PROCESO	Año de ingreso a la universidad
TIENE TRAB REMUN	¿Tiene trabajo remunerado?
HORARIO TRAB	¿Cuál es el horario de trabajo?
HORAS QUE DEDICA TRAB	Cantidad promedio de horas semanales dedicadas al trabajo.
DE PROSEGUIR ESTUDIOS	De proseguir estudios vivirá
GRUPO FAMILIAR	Cantidad de personas del grupo familiar
CUANTOS TRABAJAN GRUPO FAM	Cantidad de integrantes del grupo familiar que trabajan en forma remunerada:
QUIEN ES EL JEFE FAM	El jefe de familia es:

10.2.6 Archivo E: Postulaciones efectivas

Tabla 26: Información de Archivo E

CAMPO	DESCRIPCIÓN
NUMERO_DOCUMENTO	RUT del estudiante
AÑO_PROCESO	Año de ingreso a la universidad
PATERNO	Apellido paterno del estudiante
MATERNO	Apellido materno del estudiante
NOMBRES	Nombres del estudiante
NACIONALIDAD	Nacionalidad del estudiante
SEXO	Género del estudiante
PREFERENCIA	Orden de postulación
CODIGO_CARRERA	Código del Programa de Estudios
ESTADO_POSTULACION	Estado que indica si estudiante fue seleccionado o quedó en lista de espera
PUNTAJE_PONDERADO	Puntaje Ponderado del Estudiante
LUGAR_LISTA	Lugar en la lista de postulación al programa de estudios
SITUACION_EGRESO_EDUC	Situación de egreso
LOCAL_EDUCACIONAL	Código numérico que identifica al local educacional (planta física) de egreso.
UNIDAD_EDUCATIVA	Código numérico que identifica a la unidad educativa de egreso.
RAMA_EDUCACIONAL	Código alfanumérico. Identifica la modalidad educacional.
GRUPO_DEPENDENCIA	Grupo dependencia del establecimiento educacional
CODIGO_REGION_ORIGEN	Código de Región de Origen
AÑO_EGRESO	Año de egreso
AÑO_INGRESO_U	Año de ingreso a la Universidad
NEM	Promedio de Notas
PTJE_NEM	Puntaje de Notas de Enseñanza Media
LENG_Y_COMUNICACION	Puntaje PSU Lenguaje y Comunicación
MATEMATICA	Puntaje PSU Matemática
HISTORIA_Y_CS_SOC	Puntaje PSU Historia y Ciencias Sociales
CIENCIAS	Puntaje PSU Ciencias
PROM_LM	Promedio Puntaje PSU Lenguaje y Comunicación y Matemática
FECHA_NAC	Fecha de Nacimiento
INGRESO_BRUTO_FAM	Ingreso Bruto Familiar

10.3 Anexo 3: Comparador de modelos

Se elaboró un comparador de modelos en una hoja de cálculo de Google, el cual permite ver las matrices de confusión de cada uno de los modelos en su etapa de testeo y de predicción. Se probaron más de 50 combinaciones, como por ejemplo probar con agrupaciones de 3 años para predecir el cuarto y promediar las matrices de confusión para ver el comportamiento general, cuando se encontró un modelo que respondía de mejor forma a los datos se cambiaron algunas variables como los ciclos de entrenamiento, tasa de aprendizaje y momentum de las redes neuronales.

The screenshot shows a Google Sheets spreadsheet titled "Comparador de Modelos". The spreadsheet is organized into several sections. The top section contains a grid of data with columns for different models and rows for various parameters and metrics. The data is organized into several sections, with some rows highlighted in green and others in red. The bottom section contains numerical values for various metrics, including accuracy and loss, for different model configurations.

Ilustración 35: Comparador de Modelos. Fuente: Elaboración propia.