



**PROPUESTA DE MODELO PREDICTIVO
PARA LA DESERCIÓN UNIVERSITARIA
APLICADO A CARRERAS DE INGENIERÍA**

JACK ELGUETA

**Tesis para optar al título de Ingeniero Civil en Informática y
Telecomunicaciones**

Profesor guía: Juan Duarte

**FACULTAD DE INGENIERÍA
ESCUELA DE INFORMÁTICA Y TELECOMUNICACIONES**

**Santiago, Chile
2016**



**PROPUESTA DE MODELO PREDICTIVO
PARA LA DESERCIÓN UNIVERSITARIA
APLICADO A CARRERAS DE INGENIERÍA**

JACK ELGUETA

**Tesis para optar al título de Ingeniero Civil en Informática y
Telecomunicaciones**

**Profesor guía: Juan Duarte
Comité: Andrea Nieto
Comité: Demian Schkolnik**

**FACULTAD DE INGENIERÍA
ESCUELA DE INFORMÁTICA Y TELECOMUNICACIONES**

**Santiago, Chile
2016**

 Jack Elgueta
✉ jack.elgueta@mail_udp.cl

Dedico esta tesis a mi familia, especialmente a mi padre, que de un modo u otro siempre ha dado su apoyo en mis peores momentos.

Agradecimientos

Agradezco a todos los profesores que de un modo u otro me han otorgado el conocimiento para la elaboración de este trabajo, especialmente a mi profesor guía Juan Duarte.

Resumen

Según datos estadísticos del Servicio de Información de Educación Superior (SIES, 2015) del Ministerio de Educación revelan que sobre un 30 % de los jóvenes se retira el primer año de la carrera (3 de cada 10). Al segundo año esta cifra asciende a un 43 %. Entre las 20 carreras con mayor deserción del país, al primer y segundo año, se encuentran al primer año, 7 carreras de ingeniería con un porcentaje de deserción sobre el 36 %, y para el segundo año, 12 carreras de ingenierías con un porcentaje de deserción sobre el 47 %.

De acuerdo a la problemática anteriormente mencionada, se expondrá un estudio basado en investigaciones, sobre factores que influyen en la deserción de los alumnos en las carreras de ingeniería. Con el fin de generar una propuesta de modelo predictivo de la deserción, basados en datos de alumnos de jornada vespertina y diurna.

El modelo a construir tendrá en cuenta factores utilizados en otras investigaciones que determinan deserción de un alumno, pero se basará principalmente de tres conceptos. El primero, Inteligencia del Cliente, en donde se determinarán constructos, sobre apreciación de los estudiantes del tema deserción, con ello se espera llegar a la identificación de variables para generar el modelo. El segundo, Inteligencia de Negocio, con el cual se procesarán los datos otorgados por una institucion privada, encontrando las variables identificadas con Inteligencia del Cliente. Y tercero, Minería de Datos, el cual otorgará el conocimiento de patrones de deserción en los datos.

Abstract

According to statistics of the Servicio de Información de Educación Superior (SIES, 2015) of the Ministry of Education reveal that over 30 % of young people retires in the first year Career (3 of 10). In the second year this figure rises to 43 %. Among the 20 career with the highest desertion in the country, the first and second year are the first year, 7 career engineering a dropout percentage of 36 %, and for the second year, 12 career of engineering with a dropout percentage on 47 %.

In agreement to the problematics previously mentioned, a study based on investigations will be exposed, on factors that influence the desertion of the students the careers of engineering. In order to generate a predictive model of the desertion, based on information of students of evening and diurnal day.

The model building will take into account factors used in other studies that determine a student's desertion, but is mainly based on three concepts. The first, Customer Intelligence, where constructs are determined on assessment of students desertion issue, thus it is expected to reach identifying variables to generate the model. The second, Business Intelligence, with which the data provided by the institutions will be processed, finding the variables identified with Customer Intelligence. And third, Data Mining, which grant the knowledge of attrition patterns in the data.

Contenido

Resumen	i
Abstract	iii
Lista de tablas	vii
Lista de figuras	ix
I Marco Teórico	1
Capítulo 1. Introducción	3
1.1. Antecedentes y Motivación	3
1.2. Contexto	4
1.3. Metodología de la solución	5
1.4. Objetivos	5
1.4.1. Objetivo general	5
1.4.2. Objetivos específicos	5
1.5. Metodología de trabajo	6
Capítulo 2. Deserción en Chile	9
2.1. Deserción en Chile	9
2.1.1. Estadísticas de deserción en Chile	10
Capítulo 3. Inteligencia del Cliente	15
3.1. Inteligencia del Cliente	15
3.1.1. ¿Qué es Inteligencia del Cliente?	15
3.1.2. Estructura de Relacionamiento y CRM	16
3.1.3. Metodología de la Metáfora	16
3.2. Inteligencia de Negocio	18
3.2.1. Bases Operacionales	21
3.2.2. Procesos ETL	21
3.2.3. Área Staging	22
3.2.4. Base Analítica (Data Warehouse)	23
3.2.5. Herramientas de Gestión	25

Capítulo 4. Minería de Datos	27
4.1. Minería de Datos	27
4.1.1. Fases del Descubrimiento del Conocimiento	28
4.2. Técnicas de Minería de Datos	31
 II Desarrollo	 33
 Capítulo 5. Inteligencia del Cliente al Tema Deserción	 35
5.1. Entrevistas con Metodología de Metáforas	35
5.1.1. Constructos	36
5.1.2. Tablou de los Constructos	47
 Capítulo 6. Aplicando Minería de Datos con RapidMiner	 51
6.1. Identificación de datos y variables	51
6.1.1. Recolección de Datos	51
6.1.2. Procesamiento de Datos	53
6.1.3. Muestra de Datos	54
6.1.4. Variables	54
6.1.5. Procesos en RapidMiner	56
6.2. Modelamiento	58
6.2.1. SVN	59
6.2.2. Regresión Lineal	60
6.2.3. Red Neuronal	61
6.2.4. Árbol de Decisión	62
 III Conclusiones	 65
 Capítulo 7. Conclusiones y recomendaciones	 67
 IV Información Adicional	 69
 Anexo A. Entrevistas con Metodología de la Metáfora	 75
 Anexo B. RapidMiner	 109

Lista de tablas

3.1. Diferencias OLTP vs OLAP. [1]	20
3.2. Diferencias BD Operacional vs BD Analítica.	23
5.1. Tablou de variables.	48
6.1. Variables.	55
A.1. Entrevista 1.	76
A.2. Entrevista 2.	78
A.3. Entrevista 3.	80
A.4. Entrevista 4.	81
A.5. Entrevista 5.	82
A.6. Entrevista 6.	84
A.7. Entrevista 7.	85
A.8. Entrevista 8.	87
A.9. Entrevista 9.	88
A.10. Entrevista 10.	89
A.11. Entrevista 11.	91
A.12. Entrevista 12.	93
A.13. Entrevista 13.	95
A.14. Entrevista 14.	96
A.15. Entrevista 15.	98
A.16. Entrevista 16.	99
A.17. Entrevista 17.	101
A.18. Entrevista 18.	102
A.19. Entrevista 19.	103
A.20. Entrevista 20.	105
A.21. Entrevista 21.	106
A.22. Entrevista 22.	108
B.1. Operadores RapidMiner	112

Lista de figuras

2.1.	Retención de 1er año por Tipo de Institución	11
2.2.	Retención de 1er año por Género	11
2.3.	Retención de 1er año por Jornada	12
2.4.	Retención de 1er año por Establecimiento de Origen	12
2.5.	Retención de 1er año por Acreditación Institucional	13
3.1.	Mapa de Consenso	17
3.2.	Esquema de Inteligencia de Negocio	21
3.3.	Esquema Estrella	24
3.4.	Esquema Copo de Nieve	25
4.1.	Proceso KDD	28
4.2.	Técnicas de minería de datos	30
5.1.	Mapa de Consenso Generado	36
6.1.	Tabla resultante	53
6.2.	Proceso de Predicción	56
6.3.	Importación de datos	57
6.4.	Remplazar de datos vacíos	57
6.5.	Validación	58
6.6.	Matriz de Confusión SVN Diurno	59
6.7.	Matriz de Confusión SVN Vespertino	60
6.8.	Matriz de Confusión Regresión Lineal Diurno	60
6.9.	Matriz de Confusión Regresión Lineal Vespertino	61
6.10.	Matriz de Confusión Red Neuronal Diurno	61
6.11.	Matriz de Confusión Red Neuronal Vespertino	62
6.12.	Matriz de Confusión	62
6.13.	Matriz de Confusión	63

Parte I

Marco Teórico

Capítulo 1

Introducción

1.1. Antecedentes y Motivación

En Chile el tema deserción se ha transformado en un problema que preocupa tanto para el Ministerio de Educación, instituciones de educación superior y para investigaciones, por este motivo que se han comenzado a examinar modelos y a utilizar sofisticados métodos cuantitativos para entender la deserción en el contexto de educación superior. Algunos trabajos han dado como resultado que el fenómeno deserción es multicausal [2], específicamente, en las características individuales de pre-ingreso, tales como el establecimiento de origen, el promedio de notas y ranking en la enseñanza media, los cuales son mejores predictores de la deserción que el puntaje PSU [3]. Otros resultados indican que el tipo de establecimiento de educación media impacta en la deserción. Estudiantes provenientes de colegios particulares tienen menores tasas de deserción en los últimos años de la carrera que aquellos provenientes de la educación pública [4]. También se encontró que a mayor puntaje PSU y a mayor ingreso familiar, menores son las probabilidades de deserción [5].

La motivación de esta propuesta es apoyar con información pertinente a las carreras con alto índice de deserción, principalmente carreras de ingeniería. El modelo, aunque en forma basal puede tener un patrón, no es el mismo para jornadas diurnas y vespertinas. Idealmente se usarán datos de ambos contextos. Con el resultado se espera dar información que ayude a eje-

INTRODUCCIÓN

cutar acciones de tipo preventivas de apoyo tanto al docente como al alumno.

Es importante destacar que el modelo predictivo se genera a partir de la data existente, basada en la historia de las carreras de ingeniería.

1.2. Contexto

La tasa de retención de los estudiantes en educación superior, en especial de primer año, es uno de los indicadores más utilizados a nivel internacional para medir la eficiencia interna de las instituciones de educación, considerando que la mayor parte de los estudiantes desertan al primer año. Por otro lado, la deserción, permanencia y titulación de los estudiantes universitarios se ha transformado en un elemento estratégico para las instituciones en la rendición de cuentas y también para el Ministerio de Educación, al integrar indicadores de eficiencia docente, como la tasa de retención y de titulación, con el fin de medir el uso de recursos del crédito con aval del estado, el cual está destinado al financiamiento estudiantil a partir del año 2006 este distribuye utilizando un modelo que fija un arancel de referencia para cada carrera de las universidades pertenecientes al Consejo de Rectores de Universidades Chilenas (CRUCH).

De acuerdo con el portal mifuturo.cl del Ministerio de Educación (diciembre, 2015), actualmente existen 157 instituciones reconocidas por el Ministerio de Educación, entre ellas 54 Centros de Formación Técnica (CFT), 43 Institutos Profesionales (IP), 25 Universidades Públicas y 35 Universidades Privadas [6].

En un estudio del SIES del Ministerio de Educación (2014) [7], sobre retención señala que:

- Las universidades tienen mayor tasa de retención que los Institutos Profesionales (IP) y los Centros de Formación Técnica (CFT). Alcanzando un 76,6 % para Universidades, 66,6 % para IP y 64,5 % para CFT.

- Las carreras profesionales (con y sin licenciatura) tienen tasas de retención significativamente mayores que las carreras técnicas. Con un 75 % para las carreras profesionales y un 65,1 % para carreras técnicas.
- Las instituciones de educación superior acreditadas tienen persistentemente mayor tasa de retención de primer año que aquellas instituciones que no lo están. Las instituciones acreditadas tuvieron una tasa de 72,2 %, mientras que las sin acreditación un 53,7 %.

1.3. Metodología de la solución

Proponer un modelo predictivo que permita identificar los alumnos de carreras de ingeniería con una mayor probabilidad de deserción temprana, de modo que la institución pueda generar mecanismos de apoyo.

La solución se basa en el trabajo de un conjunto de datos de carreras de jornada vespertina y diurna de una facultad de ingeniería.

Con estos datos, se realizará un proceso de investigación utilizando inteligencia del cliente obteniendo variables para un tablou con el que se busca aplicar técnicas de minería de datos con el fin de determinar la influencia de los datos en el análisis predictivo.

Con la información otorgada a través de la investigación y prueba de diferentes modelos, se espera generar una propuesta de modelo predictivo propio.

1.4. Objetivos

1.4.1. Objetivo general

Proponer un modelo predictivo para detectar de forma oportuna a alumnos con mayor probabilidad de desertar de la carrera de ingeniería a través de inteligencia del cliente y minería de datos.

1.4.2. Objetivos específicos

1. Describir conceptos básicos de inteligencia del cliente, inteligencia de negocio y minería de datos.

INTRODUCCIÓN

2. Identificar los diferentes factores que afectan la permanencia del estudiante en su carrera académica a través de inteligencia del cliente.
3. Identificar y clasificar técnicas de minería de datos.
4. Usar minería de datos para buscar patrones que predigan el comportamiento de deserción.
5. Desarrollar diferentes modelos que determinen patrones de deserción.
6. Analizar y validar los modelos desarrollados.

1.5. Metodología de trabajo

El trabajo será dividido en tres fases:

- Fase I

En esta fase se trabajará principalmente en la definición y contextualización del problema, analizando el ámbito de la deserción de manera global y a nivel país, este último separado en instituciones estatales e instituciones privadas. Se definirán conceptos básicos de inteligencia del cliente, inteligencia de negocio y de minería de datos, de manera general, para poder analizar las diferentes técnicas de minería de datos.

Se espera para esta fase tener conocimientos fuertes del tema deserción, minería de datos, inteligencia del cliente e inteligencia de negocio para dar paso a la siguiente fase.

- Fase II

Se trabajará en entrevistas, aplicando el concepto de inteligencia del cliente para determinar constructos que permitan identificar y determinar variables influyentes en la deserción universitaria.

Se espera para esta fase, definir los principales datos que influyen en la predicción de deserción.

■ Fase III

En esta fase se trabajará en la solución, generando diferentes modelos predictivos a través de la herramienta *RapidMiner*, software que permite aplicar varias técnicas de minería de datos. Los modelos se generarán teniendo en cuenta las investigaciones y conclusiones de las fases anteriores. Aplicando fases del proceso de minería de datos y validando los modelos con el conjunto de datos.

Se espera a partir de los modelos generados, realizar un análisis para determinar el modelo que mejor se adapte a la situación deserción.

Capítulo 2

Deserción en Chile

2.1. Deserción en Chile

En los últimos 30 años en Chile ha existido un crecimiento de manera significativo en el número de matrículas de pregrado para la educación superior, de 165 mil estudiantes matriculados a principios de los ochenta, a más de un millón de estudiantes para el año 2012.

El cual ha marcado un gran cambio en el sistema de educación superior, pasó de ser un sistema de élite a uno masivo. Sin embargo, este cambio trajo consigo el aumento de diferentes problemas, como es el caso de la deserción.

La deserción no es un fenómeno aislado, las cifras en porcentaje son muy elevadas, aproximadamente un 50 % de los alumnos que entran a universidades o centros de formación técnica no terminan su programa de estudio al cual se matricularon, situación que preocupa tanto al Estado como a los distintos actores del sistema. El Estado pierde recursos, los estudiantes y sus familias pierden la inversión en la formación de quienes no completan sus estudios, las instituciones se ven afectadas ya que dejan de recibir recursos asociados a estos estudiantes y deben adaptar su funcionamiento en cursos superiores a un menor número de alumnos. Además, para los jóvenes que desertan sus sueños de graduarse se estancan, lo que les genera frustración

y descontento.

2.1.1. Estadísticas de deserción en Chile

En Chile desde el año 2006 surge el Servicio de Información de Educación Superior (SIES), del Ministerio de Educación, a través del mandato establecido en la Ley 20.129, la que en su artículo 49º señala “corresponderá al Ministerio de Educación, a través de su División de Educación Superior, desarrollar y mantener un Sistema Nacional de Información de la Educación Superior, que contenga antecedentes necesarios para la adecuada aplicación de las políticas públicas destinadas al sector de educación superior, para la gestión institucional y la información pública de manera de lograr una amplia y completa transparencia académica, administrativa y contable de las instituciones de educación superior” [8].

El SIES recopila datos de las matriculas de prácticamente la totalidad de las instituciones de educación superior, con los cuales genera informes anuales de las tasas de retención. La retención se es la contraparte de la deserción, indica cuántos alumnos siguen sus estudios y logran continuar a la obtención de su título. El documento más reciente sobre las tasas de retención de las instituciones abarca los años 2010 al 2014 y se puede ubicar en [7]. Dicho documento, operacionalmente, calcula la tasa de retención al primer año como el cociente entre el número de estudiantes que ingresan como alumnos de primer año a un programa de estudios en un año determinado y el número de estos mismos estudiantes que se mantienen como estudiantes antiguos en la misma institución al año siguiente.

A continuación, se presentará una serie de estadísticas que describen el comportamiento de la retención entre los años 2010 y el año 2014.

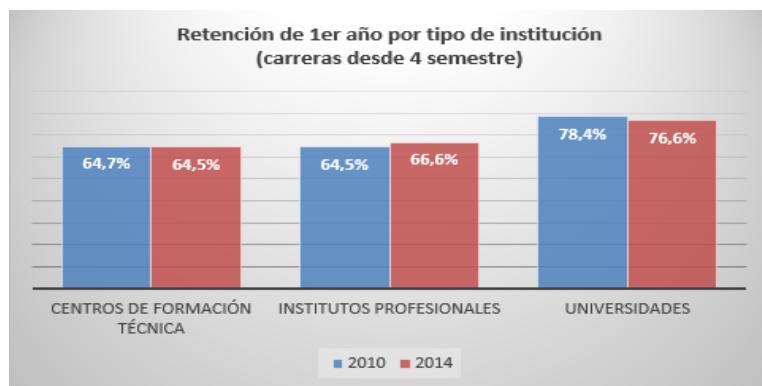


Figura 2.1. Retención de 1er año por Tipo de Institución

En la Figura 2.1 se puede apreciar la evolución de la retención en las instituciones chilenas entre los años 2010 y 2014. En el año 2014 la retención, tanto el CFT como en Universidades bajó con respecto al año 2010, la disminución se distribuye en un 0,2 % para CFT, y de un 1,8 % para Universidades, mientras que los Institutos registraron un alza de 2,1 %.

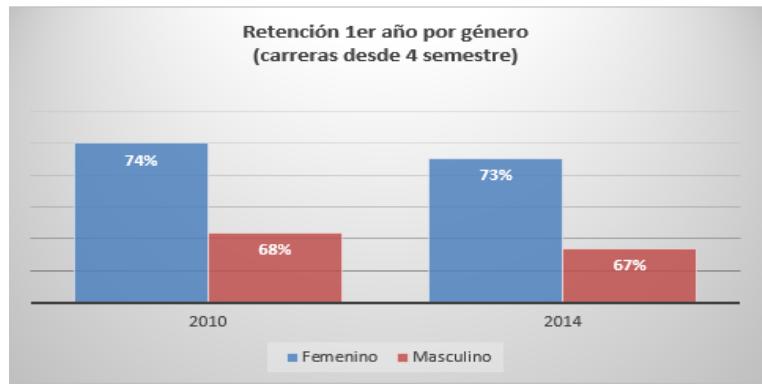


Figura 2.2. Retención de 1er año por Género

En cuanto a género la evolución de la tasa de deserción se puede apreciar en la Figura 2.2, donde para ambos género ha existido una baja en la tasa de retención, para el año 2010 se registra un 74 % para el género femenino y un 68 % para el género masculino, y para el año 2014 se registra un 73 % para el género femenino, y un 67 % para el género masculino, ambos con una baja de 1 %.

DESERCIÓN EN CHILE

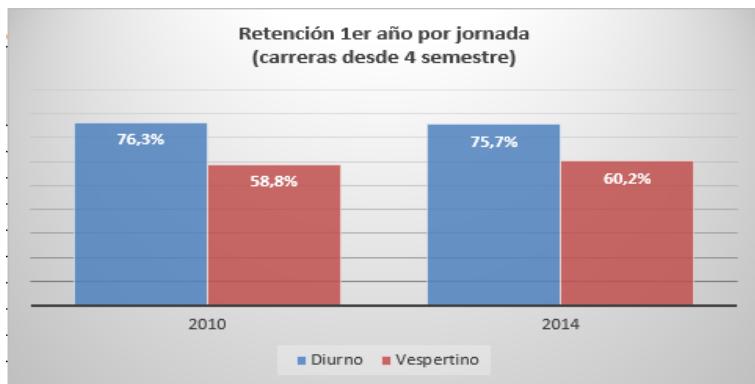


Figura 2.3. Retención de 1er año por Jornada

En lo que respecta a las jornadas de estudios, para el año 2010 la tasa de retención para carreras diurnas fue del 76,3 % y para carreras vespertinas fue de un 58,8 %, mientras que para el año 2014, para la jornada diurna fue de 75,7 % teniendo una baja de 0.6 %, y para la jornada vespertina, fue de 60,2 % obteniendo un incremento de 1,4 %, ver Figura 2.3.

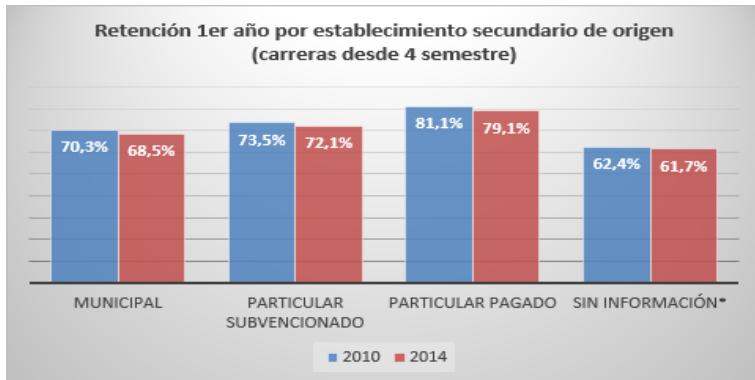


Figura 2.4. Retención de 1er año por Establecimiento de Origen

La Figura 2.4, genera la estadística a partir de los establecimientos de origen de los alumnos, mostrando para el año 2010 los siguientes resultados. Para establecimientos municipales un 70,3 %, para establecimientos particular subvencionado un 73,5 %, para establecimientos particular pagado un 81,1 % y para establecimientos sin información (matricula sin información de colegio de origen) un 62,4 %. Para el año 2014, para establecimientos municipales hubo una baja de 1,8 %, marcando 68,5 %, para establecimientos particular subvencionado hubo una baja de 1,4 %, marcando 72,1 %, para establecimientos particular pagado hubo una baja de 2 %, marcando 79,1 %,

para establecimientos sin información, también hubo una baja de 0,7 %, marcando 61.7 %.

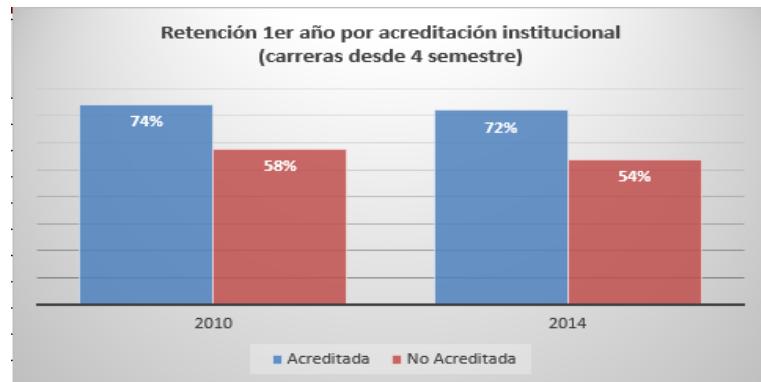


Figura 2.5. Retención de 1er año por Acreditación Institucional

La estadística para instituciones con y sin acreditación se puede visualizar en la Figura 2.5. La retención el 2010, para instituciones acreditadas fue de 74 % y para instituciones no acreditadas de 58 %. Para el año 2014, las instituciones acreditadas bajaron 2 %, alcanzando un 72 %, mientras que las no acreditadas llegaron al 54 %, registrando una baja de cuatro puntos porcentuales.

Capítulo 3

Inteligencia del Cliente

3.1. Inteligencia del Cliente

3.1.1. ¿Qué es Inteligencia del Cliente?

El proceso de Inteligencia del Cliente (Customer Intelligence, CI), gestiona, almacena y analiza datos e información relacionada a los clientes, buscando un mayor entendimiento de sus necesidades, para así mejorar el flujo de comunicación, el servicio prestado y diseñar y/o preparar decisiones estratégicas concretas. [9]

En otras palabras, consiste en convertir los datos en bruto con los que se cuentan, en información científica útil y analizable. [9]

Una gestión adecuada de CI, trae consigo beneficios para la empresa u organización. Mejorando la experiencia del servicio al cliente, toma estratégica de decisiones y predicción de hábitos o comportamiento del cliente. [9]

Inteligencia del Cliente esta fuertemente relacionado a la “Gestión de Relación con el Cliente” (Customer Relationship Management, CRM), CI es un componente clave para CRM, ya que CI capta datos e información del cliente que CRM no logra hacer. CI se basa en apreciaciones, opiniones del cliente, principalmente en aspectos de emociones que provoca el servicio y/o producto entregado por la empresa. En cambio un sistema tradicional

de CRM obtiene los datos e información del cliente basado en historiales, como por ejemplo historial de compras. Aplicando ambos conceptos de manera efectiva, CI otorga una importante fuente de información con respecto a comportamiento y experiencia de los clientes de una empresa. [10]

3.1.2. Estructura de Relacionamiento y CRM

Inteligencia del Cliente estudia, desde los ámbitos psicológicos y sociales al ser humano y su forma de relacionarse con el resto, ya que busca como objetivo relacionar los clientes con la empresa, tal como relacionarse con otro ser humano, generando un compromiso con una carga emocional.

La gestión de relación con el cliente CRM, es parte de una estrategia centrada en el cliente, una parte fundamental de su idea, es la recopilación de la mayor cantidad de información posible de los clientes, para poder entender sus necesidades y dar valor a la oferta. [11]

Los sistemas informáticos ayudan a gestionar la relación con el cliente, almacenando sus datos y aplicando Inteligencia de Negocio en conjunto con minería de datos.

3.1.3. Metodología de la Metáfora

La metodología de la metáfora es una herramienta del marketing que analiza el pensamiento consciente e inconsciente del consumidor, teniendo en cuenta puntos de neurociencia y psicología con el objetivo de comprender sus emociones y motivaciones, ya que están estrechamente relacionadas con los procesos del razonamiento y por ende las emociones contribuyen a una toma de decisiones sólidas.¹

En marketing, se piensa que el consumidor capta sus recuerdos como una cámara, un mecanismo que capta imágenes, pero con un toque de creatividad, lo que hace el recuerdo maleable, es decir, cambian constantemente de forma inconsciente, haciendo que los recuerdos sean metáforas.

¹“Viaje desde lo conocido”, Zalman Cap. 1

Cuando los consumidores usan las metáforas para describir un producto o servicio llevan sus pensamientos y sentimientos inconsciente a un nivel de conciencia que permite encontrar sentido a lo que descubren e influye en sus decisiones y acciones. Esto permite a las empresas y organizaciones diseñar ofertas más valiosas para los consumidores.²

En las ciencias cognitivas el captar o inferir el pensamiento de una persona a través de sus declaraciones o conductas es conocido como “constructos”. Los constructos no son el pensamiento ni la conducta real de la persona, más bien es una interpretación que realiza en este caso el mercadólogo del consumidor .³

Los constructos son más bien etiquetas que realiza una persona, a partir del análisis del comportamiento de otra. Los constructos pueden tener asociaciones entre si, generando un “mapa de consenso”. Un mapa de consenso nace a partir de los constructos de un grupo de personas, y se trata de un gráfico de pensamientos y sentimientos compartidos de un grupo sobre un tema en concreto. Figura 3.1

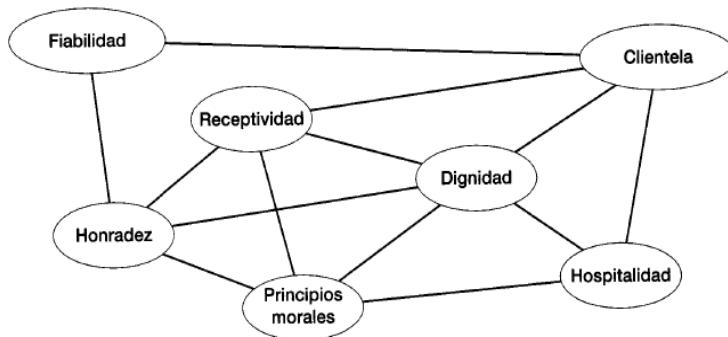


Figura 3.1. Mapa de Consenso ^a

^aFuente: Mind of the Market Laboratory/Harvard Business School

La metodología de la metáfora se aplica en entrevistas, las cuales pueden ir variando dependiendo del objetivo de la investigación, para el trabajo que se realizará en esta investigación, se seguirán los siguientes pasos:

²“Interrogar a la mente”, Zalman Cap. 4

³“Pensándolo bien”, Zalman Cap. 6

1. Recolección de revistas o periódicos con muchas imágenes.
2. Pedir al entrevistado que recorte imágenes que se relacionen o motiven sobre un tema en concreto.
3. Pedir que describan las imágenes recortadas y luego las agrupen en distintos grupos, señalando un tema por grupo.
4. Realizar una pregunta opuesta al tema en concreto y tomar apuntes respecto a las respuestas.
5. Definir constructos a partir de las apreciaciones que se obtuvieron con las descripciones de las imágenes y respuestas de la pregunta opuesta.
6. Identificar relaciones entre los constructos definidos y crear un mapa de consenso.
7. Generar métricas y variables a partir de los constructos y sus asociaciones.

3.2. Inteligencia de Negocio

Inteligencia de Negocio (Business intelligence, BI), es el proceso que permite transformar los datos en información y la información en conocimiento, permitiendo ayudar a la empresa u organización en la toma de decisiones. [12]

BI nace debido a la gran cantidad de datos que manejan las organizaciones, lo que a través de la combinación de tecnología, herramientas y procesos, logra la transformación de datos. Dentro de las tecnologías utilizadas en las empresas, se pueden identificar dos tipos de sistemas, los sistemas OLTP que proporcionan datos de orígenes y los sistemas OLAP que ayudan a analizar dichos datos: [13]

1. Sistemas OLTP (Online Transaccional Processing, Procesamiento Transaccional en Línea):

Son bases de datos orientadas al procesamiento de transacciones cortas, rápidas y en linea. Una transacción genera un proceso atómico (que debe ser validado con un commit, o invalidado con un rollback),

y que puede involucrar operaciones de inserción (INSERT), modificación (UPDATE) y borrado (DELETE) de datos.

2. Sistemas OLAP (Online Analytical Processing, Procesamiento Analítico en Línea):

Un sistema OLAP es un vector multidimensional, de N dimensiones, en el, la información se almacena en cada una de estas dimensiones, de forma ordenada y jerarquizada, lo cual ayuda a realizar un análisis rápido de su contenido. Las dimensiones de OLAP son diferentes perspectivas de análisis de información, en donde el usuario especifica diversos criterios que definen cuál y de qué forma será presentada, acumulada y ordenada la información, obteniendo resultados a una velocidad superior e la que se obtendría con un sistema de bases de datos relacional.

Las acciones básicas que tienen las tecnologías que implementan un sistema OLAP son: [12]

- Segmentar: Agrupar datos por condiciones especificadas.
- Filtrar: genera informes de datos acotados por algún parámetro.
- Profundizar (Drill down): Desglose de un dato, ejemplo, el desglose de un semestre, sería el mes.
- Sintetizar (Drill up): Deshace el desglose de “Profundizar”.
- Rotar (Drill anywhere): Realiza un desglose por una característica de una jerarquía distinta a la que se está visualizando actualmente.

Cada uno de estos sistemas están hechos para realizar una tarea específica, por lo cual, a pesar de que ambas son bases de datos, tienen diferencias. A continuación se detallan las principales diferencias entre los sistemas OLTP y OLAP, Tabla 3.1.

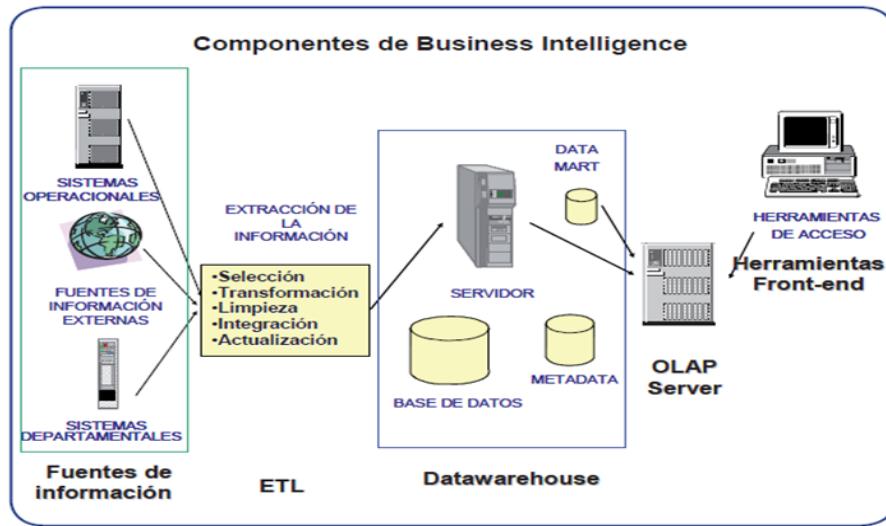
OLTP	OLAP
Son la fuente original de los datos	Los datos OLAP provienen de OLTP
Actualización en tiempo real	Solo realiza procesos de carga
Por lo general muy rápidos	Lento debido a la gran cantidad de datos a procesar
Consultas simples y estandarizadas	Consultas complejas
Bases de datos normalizadas	Bases de datos desnormalizadas
Utilizan poco espacio	Requiere un gran cantidad de espacio

Tabla 3.1. Diferencias OLTP vs OLAP. [1]

Los sistemas OLTP se pueden utilizar de manera independiente de OLAP, sin embargo cuando una empresa maneja grandes cantidades de datos, es necesario utilizar una arquitectura que involucra ambos sistemas. Arquitectura que maneja los siguientes componentes:

- Bases Operacionales (OLTP).
- Procesos ETL
- Área Staging
- Base Analítica (Data Warehouse)
- Herramientas de gestión

Estos componentes son los que conforman el proceso de Inteligencia de Negocio, y se puede visualizar en el siguiente esquema, Figura 3.2 :

**Figura 3.2.** Esquema de Inteligencia de Negocio ^a^aCurso Inteligencia de Negocios 2/2015, UDP

3.2.1. Bases Operacionales

Las bases operacionales son las encargadas de almacenar todos los datos de transacciones de una organización, datos que pueden ser de diferentes fuentes, por ejemplo, bases de datos Oracle, SQL Server, Excel entre otros.

3.2.2. Procesos ETL

Los procesos ETL (del inglés, Extract, Transform, Load) son un pieza fundamental de la Inteligencia de Negocio, ya que trata del proceso de mover datos de múltiples fuentes, transformar y luego cargar en otra base de datos.

Estos procesos consumen entre el 60 % y 80 % del tiempo de un proyecto de BI, y son proceso clave en la vida del proyecto. [14]

ETL se divide en 5 subprocesos:

1. Extracción: Recupera los datos necesarios físicamente de las distintas fuentes de información, dejando los datos en bruto. Su extracción puede ser manual o utilizando una herramienta. Sus principales problemas son las distintas fuentes de datos y distintas plataformas.

2. Limpieza: Recupera los datos en bruto y comprueba su calidad, elimina los duplicados y, cuando es posible, corrige los valores erróneos, si existen valores vacíos, los completa siempre y cuando sea posible.

Estos valores, generalmente se llaman valores sucios, y es muy probable encontrarlos, lo importante es saber el origen de dicha suciedad.

La limpieza se divide en distintas etapas:

- Depurar los valores: Localiza elementos individuales de las fuentes de datos y los aísla en estructura de destino, por ejemplo, Dirección → calle, número, comuna, ciudad.
 - Corregir: Corrige los datos individuales utilizando algoritmos y fuentes de datos externas.
 - Estandarizar: Se trabaja con formatos de datos definidos y se aplican rutinas para homogeneizar datos.
 - Relacionar: Busca y relaciona los valores de los registros, corrigiéndolos y estandarizándolos según reglas del negocio, eliminando datos duplicados.
 - Consolidar: Analiza e identifica relaciones entre registros relacionados y los junta en una sola representación.
3. Transformación: Recupera los datos limpios y de alta calidad, los estructura y sumariza en distintos modelos de análisis.
 4. Integración: Valida que los datos cargados son consistentes con las definiciones y formatos de la base analítica. Integra los datos en los distintos modelos de las distintas áreas de negocio que se han definido.
 5. Actualización: Se añaden los datos a la base analítica.

3.2.3. Área Staging

El área staging es una base de datos la cual es una copia de todas las bases operacionales, generada principalmente para realizar las siguientes actividades:

- Facilitar extracción de datos.

- Limpiar y mejorar la calidad de los datos.
- Operar y ejecutar procesos sin afectar el funcionamiento de las bases operacionales.

3.2.4. Base Analítica (Data Warehouse)

La base analítica o Data Warehouse es un repositorio o colección de datos orientada a un tema, integrada, variante en el tiempo y no volátil, que apoya el proceso de toma de decisiones de gestión.

En la base analítica se almacenan todos los datos procesados de Área Staging y ETL.

Un Data Warehouse puede tener varios DataMart, un DataMart son pequeños Data Warehouse que almacenan datos específicos de un área de negocio o departamento de una empresa.

A continuación se detallan las principales diferencias entre las bases operacionales y las bases analíticas, Tabla 3.2.

BD Operacional	BD Analítica
De transacciones	De consultas masivas
Predomina la actualización	Predomina la consulta
Actualizable	Carga, pero no actualiza
De tipo operativo	Ánálisis y toma de decisiones
Datos normalizadas	Datos en estrella o copo de nieve
Usuarios de perfiles medios o bajos	Usuarios de perfiles altos
Estructura normalmente relacional	Visión multidimensional

Tabla 3.2. Diferencias BD Operacional vs BD Analítica. ^a

^aCreación propia

Las bases analíticas están diseñados y desarrollados para realizar consultas complejas dentro de un volumen muy grande de datos, para lo cual se necesita una arquitectura para cumplir con este propósito, existen 3 arquitecturas, MOLAP (Multidimensional), ROLAP (Relacional) y HOLAP (Híbrido), la más usada es el modelo multidimensional.

Los modelos multidimensionales tienen 3 conceptos principales:

- Unidad de Medida: Atributo cuantificable asociado a los hechos. Ejemplos, Volumen de las ventas, número de transacciones efectuadas, porcentaje de beneficios.
- Tabla de Hecho: Tabla que almacena los valores detallados para medidas, permite realizar inserciones entre las dimensiones y las medidas.
- Dimensiones: Son tablas que permiten la agrupación de los hechos.

Las bases multidimensionales son utilizadas con dos esquemas:

1. Esquema Estrella: Es un esquema simple que utiliza una tabla de hecho, una o más tablas de dimensiones desnormalizadas, que están unidas con la tabla de hecho, Figura 3.3.

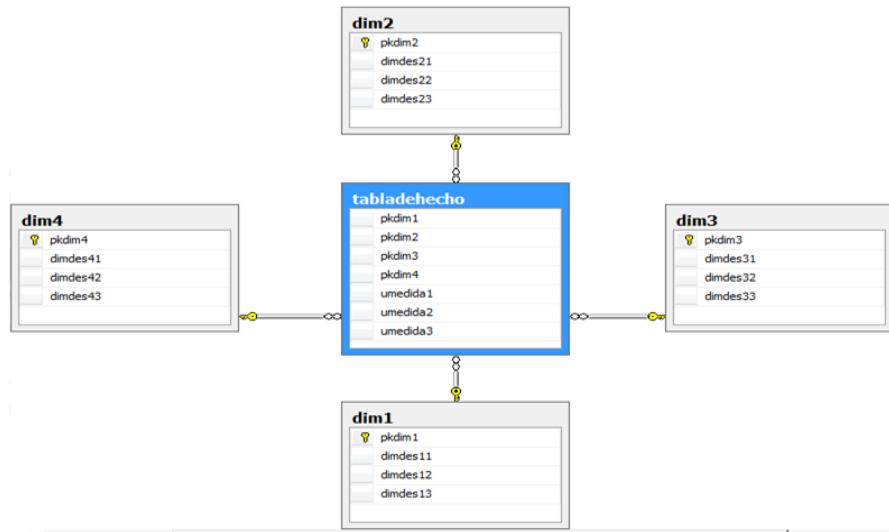
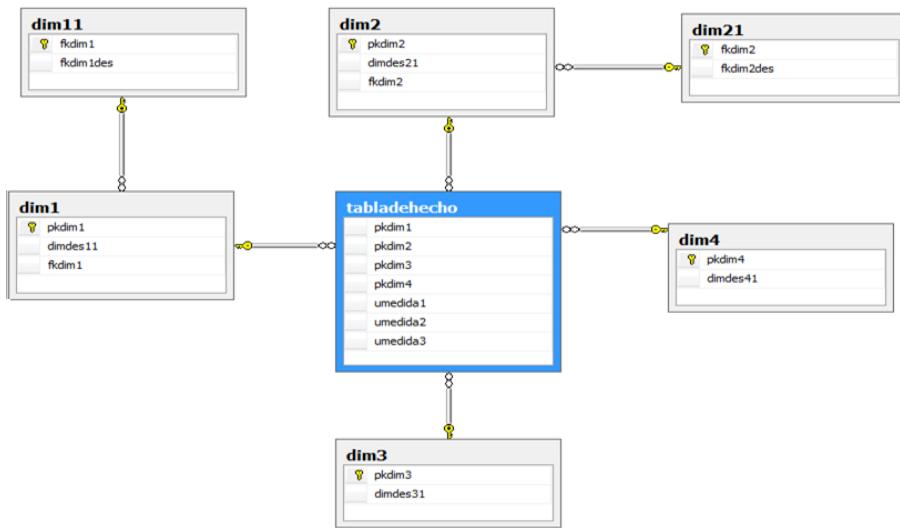


Figura 3.3. Esquema Estrella ^a

^aCurso Inteligencia de Negocios 2/2015, UDP

2. Esquema Copo de Nieve: Es un esquema complejo, derivado del esquema estrella, las tablas de dimensiones se encuentran normalizadas en múltiples tablas, y la tabla de hecho no es la única que establece uniones con las otras tablas, Figura 3.4.

**Figura 3.4.** Esquema Copo de Nieve ^a

^aCurso Inteligencia de Negocios 2/2015, UDP

3.2.5. Herramientas de Gestión

Después de que los datos son procesados y almacenados en la base analítica, estos se encuentran disponibles para ser visualizados y analizados a través de herramientas de gestión, que permiten a las personas de alto mando dentro de una organización tener la información de forma rápida y oportuna para la toma de decisiones. Existen una variedad de herramientas que cumplen diferentes objetivos, a continuación se identifican algunas de ellas.

3.2.5.1. Query & Reporting

Son herramientas que realizan consultas o informes que trabajan tanto sobre el detalle como sobre las agregaciones de la información. Manejan una gran cantidad de información y principalmente responden a la pregunta ¿Qué sucedió?.

3.2.5.2. Análisis OLAP

Las herramientas de análisis OLAP o “Analytics Reports” son utilizadas por los analistas que necesitan tener la información desde distintos puntos de vistas. Estas herramientas permiten navegar dentro de la información profundizando o ir al detalle, conocidos como navegación jerárquica (Drill

Down) y navegación al detalle (Drill Through). Estas herramientas responde a la pregunta ¿Por qué sucedio?.

3.2.5.3. Dashboard

Las herramientas de dashboard totaliza los datos, los cuales generalmente están representados en gráficas que representan el estado actual del negocio. La totalización de datos se representan como KPIs (Key Performance Indicator), en español, indicadores claves de desempeño, que son valores relacionados con un objetivo fijado de antemano por la organización y normalmente se expresa en porcentaje, sirven para visualizar el progreso en un aspecto concreto. Por ejemplo, una representación gráfica comúnmente usada es el semáforo, el cual puede representar las ventas del mes de una empresa fijando indicadores acordes a los objetivos de ventas, si el objetivo es alcanzar las 100 ventas en un mes, si se esta cumpliendo con dicho objetivo o más el semáforo estará en color verde, si es menor a 100 estará en amarillo y si es menor a una cantidad definida ejemplo 70, entonces se encontrará en color rojo.

Estas herramientas son las más usadas por los ejecutivos o personal de alto rango dentro de una organización, ya que permite monitorear tiempo real la situación actual de la organización, facilitando la toma de decisiones. Estas herramientas responden a la pregunta ¿Qué está sucediendo?.

3.2.5.4. Data Mining o Minería de Datos

La herramientas de data mining realizan análisis de datos históricos de la organización, con el objetivo de buscar patrones que predigan el comportamiento o acción de un suceso o persona. Para esto realiza acciones de clasificación, segmentación o predicción. Por ejemplo, el análisis de las ventas de los últimos 5 años permite estimar una aproximación de las ventas del año actual. Estas herramientas responden a la pregunta ¿Qué va a suceder?..

Capítulo 4

Minería de Datos

4.1. Minería de Datos

El proceso de minería de datos o exploración de datos es un proceso, considerado una etapa de un proceso mayor llamado “Descubrimiento de Conocimiento en Base de Datos” (KDD, del inglés, Knowledge Discovery in Databases), “proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y comprensible a partir de los datos”.¹

Usualmente ambos conceptos son intercambiables, pero se entiende por KDD al proceso de encontrar información y/o patrones útiles en los datos. En cambio minería de datos, es el uso de algoritmos para extraer información y/o patrones derivados dentro del proceso KDD.

Dentro de la definición del proceso KDD se pueden desprender los siguientes conceptos:

- Patrones válidos: Los patrones deben ser precisos y verdaderos para nuevos datos, similar al principio de inducción matemática.
- Novedosos: Los datos deben aportar algo nuevo al usuario, un conocimiento. Esto es la esencia de la minería de datos.

¹Fayyad et al. 1996

- Útil: Los datos deben conducir a acciones por parte de los usuarios receptores.
- Comprensible: Los datos deben ser posible de interpretar y por ende comunicar.

En la Figura 4.1, se puede apreciar el proceso de KDD.

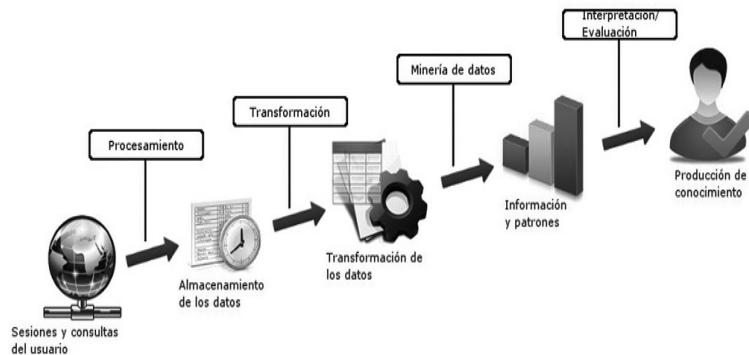


Figura 4.1. Proceso KDD ^a

^aImagen extraída de <http://www.scielo.org.co> articulo “Aplicación del proceso de KDD en el contexto de bibliomining: El caso Elogim”

4.1.1. Fases del Descubrimiento del Conocimiento

El proceso KDD puede ser representado en 7 objetivos, los cuales se logran en diferentes fases, estos objetivos son:

1. Determinar las fuentes de información que pueden ser útiles y dónde conseguirlas.
2. Diseñar un esquema de almacenamiento de datos, un Data Warehouse.
3. Implementación del almacén de datos que permita la navegación y visualización previa de los datos para el análisis.
4. Selección, limpieza y transformación de los datos a analizar.
5. Seleccionar y aplicar el método de minería de datos apropiado.
6. Evaluación, interpretación, transformación y representación de los patrones extraídos.
7. Difusión y uso del nuevo conocimiento.

4.1.1.1. Recogida de Datos

La primeras fases del KDD determinan que las fases sucesivas sean capaces de extraer conocimiento válido y útil a partir de la información original.

En la fase de recogida de datos se sigue los siguientes pasos:

- Selección de datos: Por lo general la información que se quiere investigar sobre un cierto dominio de la organización se encuentra en bases de datos y fuentes internas como externas.
- Pre-proceso o limpieza de datos: En este paso se debe eliminar el mayor número de datos erróneos o inconsistentes e irrelevantes, esto también se conoce como limpieza y selección.
- Transformación de datos: Una vez limpiados y seleccionados los datos, se realiza un proceso de transformación que permite homologar la información, dando como resultado un conjunto de filas y columnas denominado “Vista Minable”, con el fin de dejar los datos preparados para la modelización.

4.1.1.2. Modelamiento

El modelamiento o minería de datos es la etapa principal del KDD, aquí se determinan patrones y modelos para descubrir el conocimiento. Existen dos tipos de modelos y diferentes tareas para cada modelo.

1. Tareas del Modelo Predictivo

Son aquellas que buscan patrones que ayuden a predecir el comportamiento o tendencia de uno o varios valores.

- Clasificación: Los datos o registros son agrupados en clases, los cuales pueden tomar valores discretos. Su objetivo es determinar patrones en los registros, para identificar o predecir a que clase pertenecen los registros nuevos.
- Regresión: Se usa una regresión para predecir los valores ausentes de una variable basándose en su relación con otras variables del conjunto de datos.

2. Tareas del Modelo Descriptivo

Son aquellas que exploran las propiedades de los datos analizados para encontrar patrones entre ellos.

- Agrupamiento (Clasificación no supervisada): Es similar a la clasificación, excepto que los grupos no son predefinidos. El objetivo es segmentar un conjunto de datos en grupos que pueden ser disjuntos o no. Los grupos se forman basados en la similaridad de los datos en ciertas variables. Como los grupos no son dados a priori se debe dar una interpretación de los grupos que se forman.
- Correlaciones: Identifica el grado de similitud de dos variables numéricas.
- Asociación: Las asociaciones entre dos atributos ocurre cuando la frecuencia de que se den dos valores determinados de cada uno conjuntamente es relativamente alta.
- Reglas de asociación secuenciales: Es una variante de asociación que utiliza la variable tiempo para identificar correlación.

Para estos modelos y tareas existen diferentes técnicas de minería de datos las cuales se pueden aplicar, a continuación se muestra un cuadro de las técnicas y tareas que utilizan, Figura 4.2

Nombre	PREDICTIVO		DESCRIPTIVO		
	Clasificacion	Regresion	Agrupamiento	Reglas de Asociacion	Correlacion / Factorizaciones
Redes Neuronales	X	X	X		
Arboles de decision ID3, C4.5, C5.0	X				
Arboles de decision CART	X	X			
Otros Arboles de decision	X	X	X	X	
Redes de Kohonen			X		
Regresion lineal y logaritmica		X			X
Regresion logistica	X			X	
Kmeans			X		
Apriori				X	
Naive Bayes	X				
Vecinos mas proximos	X	X	X		
Analisis factorial y de comp. Ppales					X
TwoStep, Cobweb			X		
Algoritmos geneticos y evolutivos	X	X	X	X	X
Maquinas de vectores soporte	X	X	X		
CN2 rules (cobertura)	X			X	
Analisis discriminante multivariante	X				

Figura 4.2. Técnicas de minería de datos ^a

^aImagen extraída del libro “Introducción a la Minería de Datos”, pag. 148, cap. 6

4.1.1.3. Evaluación

La fase de evaluación tiene como objetivo probar y validar el modelo creado en las fases anteriores. Para realizar las pruebas del modelo, se divide el set de datos en dos grupos, un grupo de entrenamiento de los datos, que ayuda a identificar la predicción que se espera, y el otro grupo es de prueba se valida la predicción y se analiza el porcentaje de acierto de dicha predicción. Por lo general se obtienen varios modelos aplicando las diferentes técnicas de minería de datos, los cuales son comparados buscando aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

4.2. Técnicas de Minería de Datos

En el capítulo anterior se explico que existen diferentes técnicas y tareas en minería de datos para los distintos tipos de modelamiento.

El modelamiento acorde a este trabajo, corresponde a un modelo predictivo, en el cual se utilizarán cuatro técnicas en *RapidMiner*, para generar cuatro modelos, estas técnicas son:

- Regresión Lineal:

Es una técnica utilizada para la predicción numérica. Es una medida estadística que determina la relación entre una variable dependiente y una serie de variables independientes.

La regresión lineal intenta modelar la relación entre una variable escalar y una o más variables explicativas ajustando una ecuación lineal a los datos observados. Por ejemplo, uno podría querer relacionar los pesos de los individuos con sus alturas usando un modelo de regresión lineal [15].

- Árboles de decisión:

Es una representación de datos que tiene forma de árbol invertido, donde sus nodos raíces están en la parte superior, creciendo hacia abajo. Su objetivo es predecir el valor de un atributo destino (generalmente llamado clase o etiqueta) basándose en varios atributos de entrada,

cada nodo interior del árbol corresponde a uno de los atributos de entrada. El número de aristas de un nodo interior nominal es igual al número de valores posibles del atributo de entrada correspondiente. Los bordes salientes de los atributos numéricos están etiquetados con rangos disjuntos. Cada nodo de hoja representa un valor del atributo de etiqueta dado los valores de los atributos de entrada representados por el camino de la raíz a la hoja [16].

- Redes Neuronales:

Una red neuronal, es un modelo matemático o modelo computacional que se inspira en la estructura y aspectos funcionales de las redes neuronales biológicas. Una red neuronal consiste en un grupo interconectado de neuronas artificiales, y procesa la información utilizando un enfoque conexionista a la computación (el principio conexionista central es que los fenómenos mentales pueden ser descritos por redes interconectadas de unidades simples ya menudo uniformes). En la mayoría de los casos, una red neuronal es un sistema adaptativo que cambia su estructura basada en información externa o interna que fluye a través de la red durante la fase de aprendizaje. Las redes neuronales modernas suelen usarse para modelar relaciones complejas entre entradas y salidas o para encontrar patrones en los datos [17].

- Máquinas de Vectores de Soporte (SVN):

Las máquinas de vectores de soporte est醤dard toman un conjunto de datos de entrada y predice, para cada entrada dada, cuál de las dos clases posibles comprende la entrada, haciendo del SVM un clasificador binario no probabilístico lineal. Dado un conjunto de ejemplos de entrenamiento, cada uno marcado como pertenecientes a una de dos categorías, un algoritmo de entrenamiento de SVM construye un modelo que asigna nuevos ejemplos en una categoría u otra. Un modelo SVM es una representación de los ejemplos como puntos en el espacio, mapeados de modo que los ejemplos de las categorías separadas se dividan por una brecha clara que es lo más amplia posible. Nuevos ejemplos se asignan a ese mismo espacio y se predice que pertenecen a una categoría basada en qué lado de la brecha en la que caen [18].

Parte II

Desarrollo

Capítulo 5

Inteligencia del Cliente al Tema Deserción

Un aspecto importante del concepto de Inteligencia del Cliente, es entender las necesidades del consumidor. En el caso de este trabajo, se manejan los conceptos de IC aplicado a alumnos, para entender sus necesidades y motivaciones en terminar una carrera universitaria.

El siguiente desarrollo busca como objetivo ver las motivaciones de los alumnos, para convertirlos en constructos junto a sus asociaciones, los cuales entregarán métricas y variables para analizar el tema deserción. Las métricas y variables definidas en esta sección, ayudará a la selección de datos y construcción de los modelos, para luego aplicar las técnicas de minería de datos con la herramienta *RapidMiner* a la base de datos de la entidad privada.

5.1. Entrevistas con Metodología de Metáforas

Para captar las necesidades de los alumnos se utilizó la metodología de metáforas, en donde se le pide al entrevistado que recorte de 6 a 8 imágenes de revistas que representen para él, el concepto a evaluar. El concepto a evaluar en este caso, se representa con la pregunta principal ¿Qué te motiva a terminar la carrera?. (Ver Anexo A).

El marco muestral se define a partir de un segmento de cliente, para el caso de esta investigación, el segmento abarca a alumnos de educación su-

terior de instituciones universitarias.

Como el objetivo de las entrevistas es encontrar las necesidades y motivaciones de los alumnos, se utilizará un muestreo no probabilístico, del tipo intencional o por conveniencia, que consiste en seleccionar individuos que son más fáciles de entrevistar [19].

El resultado de las entrevistas, permitirá generar constructos a partir de sus necesidades y motivaciones.

5.1.1. Constructos

Los constructos son definiciones interpretadas a partir de las asociaciones de las imágenes escogidas por los entrevistados, estas definiciones derivan en variables a contemplar para armar un modelo.

Los constructos recogidos en esta investigación reflejan el sentir del alumno en la carrera, los cuales se pueden plasmar en su formación como profesional y en sus sueños futuros.

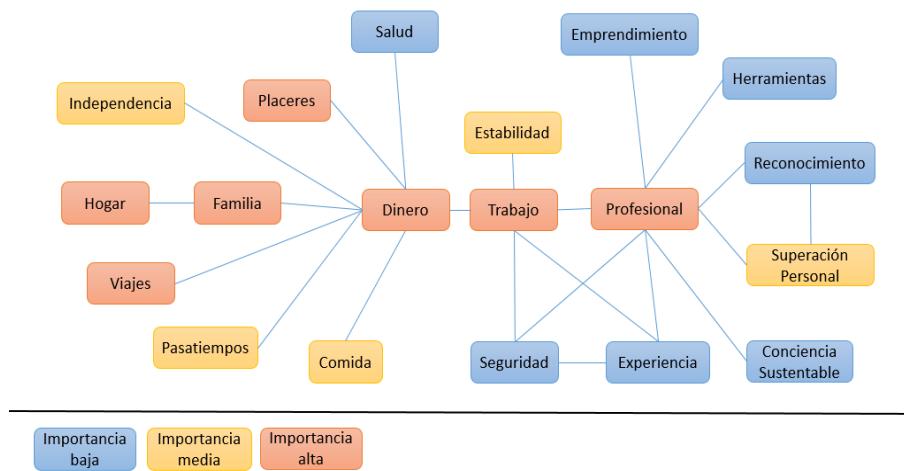


Figura 5.1. Mapa de Consenso Generado

La Figura 5.1 representa el *Mapa de Consenso* generado a partir de los constructos obtenidos de las entrevistas, los más importantes son los que se

repitieron en la mayoría de las entrevistas y son los principales insight de los alumnos al responder la pregunta ¿Qué te motiva a terminar la carrera?.

A continuación se detallan los constructos obtenidos, en donde las asociaciones son realizadas por los entrevistados, las definiciones son interpretaciones a las asociaciones, y las variables son los posibles datos ligados a las definiciones.

Constructo 1 - Familia

Asociaciones:

1. Representa el compartir con la familia una situación económica estable.
2. Tener una familia y tener dinero para mantener a la familia.
3. El terminar una carrera implica tener un buen trabajo y mucho dinero, con el cual se puede optar a mejores condiciones de vida.
4. Poder dar una buena educación a los hijos.
5. Poder formar mi familia alegre y saludable.

Definición:

Un alumno que valore la familia, también valorará el tiempo que tiene para dedicar a esta, mientras estudia.

Una alumno que vive en comunas apartadas tendrá menos tiempo con la familia, lo que redundaría en buscar lugares más cercanos.

Por otro lado es importante clasificar las carreras que otorgan un mayor bienestar para la familia.

Sumado a lo anterior de igual forma se debe analizar el impacto que tiene el deseo de estudiar la carrera.

Variables:

- Distancia a universidad (comuna).
- Prioridad de postulación a carrera
- Tiempo medio de encontrar trabajo estudiante egresado.

Constructo 2 - Independencia

Asociaciones:

INTELIGENCIA DEL CLIENTE AL TEMA DESERCIÓN

1. Tener un espacio personal grande y vanguardista.
2. Tener un auto para tener independencia.
3. Tener la autonomía de vivir solo y tener un hogar
4. Trabajar de forma independiente
5. Poder vivir sola y tener mi departamento.

Definición:

Un alumno que valora la independencia, se encuentra más dispuesto a conseguir beneficios que lo ayuden económicamente, así como también trabajar durante la etapa estudiantil para satisfacer sus necesidades.

Variables:

- Becas
- Créditos
- Beca alimenticia (Junaeb)
- Trabajador

Constructo 3 - Dinero

Asociaciones:

1. Representa el vehículo que podría adquirir teniendo recursos
2. Felicidad y estabilidad amorosa, gracias a una buena situación económica.
3. Poder acceder a muchas cosas, como bienes
4. Poder comprar tecnología, gracias al buen sueldo que tendrá
5. Ser un profesional significa poder tener mucho dinero

Definición:

Un alumno que tome en cuenta el dinero que ganará luego de egresar, debe tener conciencia en las estadísticas de su carrera, cual es la tasa de empleabilidad al 1er año de egreso y como aumentará su sueldo con los años de experiencia.

Variables:

- Empleabilidad al 1er año
- Ingreso promedio al 4to año

Constructo 4 - Placeres

Asociaciones:

1. Tener un título, se asocia a disfrutar placeres de la vida.
2. Representa las comodidades que uno podría tener una vez trabajando.
3. Poder obtener gustos exclusivos gracias al dinero.
4. Comprar cosas exclusivas

Definición:

Hace referencia a la libertad de poder gastar el dinero ganado en cosas o experiencias satisfactorias. Este constructo se encuentra asociado al constructo de Dinero, por lo que se aplican las mismas variables.

Variables:

- Empleabilidad al 1er año
- Ingreso promedio al 4to año

Constructo 5 - Trabajo

Asociaciones:

1. Trabajar de forma relajada, ya que con el título tienes una base para dedicarse a lo que te gusta.
2. Representa la adquisición de un bien raíz gracias al trabajo después de titulado
3. Es lo que quiero llegar hacer, y tener un trabajo luego de titularme.
4. Poder relacionarse con otras personas y poder tener un trabajo siner-gico

Definición:

Representa a lo que se dedicará y obtendrá el alumno una vez titulado. Se relaciona al constructo Dinero y se asocian las mismas variables.

Variables:

INTELIGENCIA DEL CLIENTE AL TEMA DESERCIÓN

- Empleabilidad al 1er año
- Ingreso promedio al 4to año

Constructo 6 - Profesional

Asociaciones:

1. Generar aportes al trabajo, ser un buen profesional
2. Tener un trabajo relacionado en redes informáticas y ser un buen profesional, para viajar por el mundo
3. Encontrar un trabajo en una empresa prestigiosa.
4. Es a lo que uno aspira ser como ingeniero, teniendo buen puesto y reputación.
5. Poder dedicarme al desarrollo de aplicaciones móviles.

Definición:

Un alumno que valora el ser un profesional, tomará en cuenta el prestigio y excelencia de la institución, como también su formación profesional, por lo que será interesante saber si la institución se encuentra acreditada y si el alumno participa o ha participado en actividades que ayuden a validar sus capacidades.

Variables:

- Acreditación Institucional
- Carrera Acreditada
- Talleres de Certificación
- Talleres de Desarrollo Profesional

Constructo 7 - Viajes

Asociaciones:

1. Tener la disponibilidad de tiempo y dinero, salir de la rutina, tener una vida más relajada
2. Después de la carrera hay tiempo para disfrutar y conocer el mundo.
3. Libertad en poder ir a lugares diferentes y con mucha naturaleza

4. Sin dinero no se puede viajar, y teniendo una carrera, uno puede ganar mucho dinero.
5. Poder viajar y hospedarse en buenos hoteles

Definición:

Representa la libertad y la posibilidad de poder conocer nuevos lugares, sin preocupaciones de dinero.

Los viajes para los entrevistados representa el conocer otros países, si bien hace referencia a algo más vacacional, se puede asociar con las diferentes oportunidades que puede otorgar la institución para realizar intercambios, más puntual si el alumno a realizado un intercambio por medio de la institución.

Variables:

- Alumno realizo intercambio

Constructo 8 - Pasatiempos

Asociaciones:

1. Tener una buena situación económica para disfrutar eventos culturales.
2. Tener muchos libros por el gusto de leer.
3. Poder tener la libertad de darse tiempos libres y hacer las cosas que uno quiere.
4. Dedicar tiempo para mi persona y realizar lo que me gusta, la cocina.
5. Representa al parte artística que podría dedicarme teniendo recursos.

Definición:

Hace referencia al tiempo disponible que tendrá, a poder cumplir y hacer lo que siempre le ha gustado hacer. Es importante para el alumno, que la institución a la cual pertenece disponga de talleres o facilidades que complementen sus estudios con los pasatiempos y si el alumno participa activamente de algún taller proporcionado por la institución.

Variables:

- Taller deportivo o cultural

INTELIGENCIA DEL CLIENTE AL TEMA DESERCIÓN

- Beca deportiva

Constructo 9 - Hogar

Asociaciones:

1. Representa el compartir con la familia una situación económica estable.
2. Tener una familia y tener dinero para mantener a la familia.
3. Poder acceder a una buena casa en un buen barrio
4. Tener una casa amplia y con bonita vista para recibir a la familia

Definición:

Representa independencia y unión familiar a través de la estabilidad económica. Se asocia a los constructos de Dinero y Familia

Variables:

- Empleabilidad al 1er año
- Ingreso promedio al 4to año
- Distancia a universidad (comuna).
- Estudiar 3 primeras prioridades.
- Tiempo medio de encontrar trabajo estudiante egresado.

Constructo 10 - Comida

Asociaciones:

1. Tener una buena situación económica.
2. Tener para comer y una vida sana.
3. Representa los gustos que me podría dar, teniendo recursos.
4. Tener gustos sin depender económicamente de otras personas
5. Tener un lugar agradable dentro de la casa

Definición:

Hace mención a satisfacer la necesidad de alimento y darse una buena vida. Se asocia al constructo de Dinero junto a sus variables.

Variables:

- Empleabilidad al 1er año
- Ingreso promedio al 4to año
- Beca alimenticia

Constructo 11 - Experiencia

Asociaciones:

1. Estar listo para enfrentar el mundo laboral.
2. Un profesional termina las cosas de manera más rápida, porque está preparado.

Definición:

Hace referencia a sentirse preparado, a estar seguro en lo que se trabaja y poder cumplir de manera excelente. Es importante para el alumno ver la vinculación con el medio que tiene la institución y las prácticas ofrecidas.

Variables:

- Prácticas realizadas
- Taller de Desarrollo Profesional

Constructo 12 - Superación personal

Asociaciones:

1. Cerrar una etapa de la vida, para pasar a una siguiente.
2. Significa el cierre y el comienzo de una nueva etapa.
3. Terminar una carrera transmite seguridad y sentirse útil para algo.
4. Representa crecimiento personal y laboral .
5. Es el logro de cumplir con el objetivo de ser ingeniero.

Definición:

Representa el alcanzar una meta, crecer, ser independiente y seguro. Esto se puede representar por los logros alcanzados por el alumno, como la cantidad de ramos aprobados en un semestre, promedio de notas o posición en ranking de alumno.

Variables:

INTELIGENCIA DEL CLIENTE AL TEMA DESERCIÓN

- Porcentaje ramos aprobados
- Promedio de notas
- Ranking del alumno

Constructo 13 - Salud

Asociaciones:

1. Poder tener el dinero para obtener buenos tratamientos médicos

Definición:

Tener un título implica poder acceder a una salud de calidad. Se asocia al constructo Dinero junto a sus variables.

Variables:

- Empleabilidad al 1er año
- Ingreso promedio al 4to año

Constructo 14 - Seguridad

Asociaciones:

1. Desarrollar seguridad personal
2. Tener una carrera te da seguridad de tener un buen trabajo
3. Salir de la carrera es un gran paso y te da libertad y seguridad en trabajar en lo que quieras

Definición:

Hace referencia a que un título certifica las capacidades de una persona, haciéndolas sentir más seguras. Un alumno que busca seguridad laboral, buscará mejorar sus capacidades con certificaciones, es importante entonces que la institución a la que pertenece certifique los conocimientos del alumno.

Variables:

- Talleres de certificación
- Acreditación institucional
- Carrera acreditada

Constructo 15 - Conciencia sustentable

Asociaciones:

1. Vida consciente con el planeta, a largo plazo
2. Crear una fundación para ayudar a niños con discapacidad
3. Me gustaría tener un desarrollo sustentable, poder ayudar a la naturaleza con mi carrera

Definición:

Se refiere a que una carrera entrega nociones y entendimiento sobre los efectos que tienen diferentes efectos sobre el planeta y la sociedad.

Variables:

- Participación en voluntariado social o ONG
- Ramos relacionados a conciencia sustentable

Constructo 16 - Reconocimiento

Asociaciones:

1. Representa el triunfo que puede llegar a tener como profesional.
2. Terminar la carrera, da una sensación de libertad y triunfo.
3. Estar graduado y ser reconocido como un ingeniero.
4. Como te observan las otras personas y como me proyectó como profesional
5. Refleja a mis padres, al terminar mi carrera ellos estarían orgullosos de mí

Definición:

Representa a la sensación de un buen estatus, de crecimiento y aceptación. Se asocia al constructo Superación Personal junto a sus variables.

Variables:

- Porcentaje ramos aprobados
- Promedio de notas

INTELIGENCIA DEL CLIENTE AL TEMA DESERCIÓN

- Ranking del alumno

Constructo 17 - Emprendimiento

Asociaciones:

1. Poder sacar un postgrado
2. Poder dar clases en un futuro.
3. Tener tiempo para poder realizar un curso de inglés y viajar al extranjero.
4. Poder emprender con un negocio gourmet
5. Tener mi propia oficina, trabajar con más gente y enseñar lo que sé

Definición:

Representa los sueños a futuros que se pueden cumplir gracias a un título. Para este constructo se relaciona la participación del alumno en distintas actividades que se relacionen a emprendimiento, como talleres o concursos a través de la institución a la que pertenece.

Variables:

- Talleres de emprendimiento

Constructo 18 - Estabilidad

Asociaciones:

1. Generar estabilidad y buenos lazos.
2. Trabajar con personas, tener vida social
3. Poder tener estabilidad económica y por ende tranquilidad

Definición:

Representa que un título otorga poder estar seguro tanto económicamente como emocionalmente. Un alumno con interés en tener una estabilidad económica y emocional, participará de actividades que lo ayuden a encontrar esa estabilidad, por ende las instituciones deben tener estas actividades, y el alumno participar de dichas actividades.

Variables:

- Talleres Bienestar estudiantil
- Talleres de Desarrollo Profesional

Constructo 19 - Herramientas

Asociaciones:

1. Sueños de ser papá, de tener las herramientas para dar un buen pasar
2. Los profesores motivan y te enseñan hacer trabajos de calidad
3. La universidad me da herramientas que me servirán para el futuro
4. Representa las habilidades obtenidas en la carrera

Definición:

Hace referencia al aprendizaje obtenido en la etapa universitaria. Es importante para el alumno tener las capacidades suficientes para poder ejercer sin problemas su título, por ende participara en actividades que le brinden estas herramientas.

Variables:

- Talleres de certificación
- Talleres de Desarrollo Profesional

5.1.2. Tablou de los Constructos

Variable	Descripción
Institución y Carrera	
Distancia casa - institución	Distancia en kilómetros
Empleabilidad al 1er año	Porcentaje de egresados con trabajo al 1er año
Ingreso promedio 4to año	Cantidad de dinero a ganar al 4to año de experiencia laboral
Acreditación Institucional	Indica si la institución se encuentra acreditada
Carrera acreditada	Indica si la carrera se encuentra acreditada
Desempeño Alumno	

INTELIGENCIA DEL CLIENTE AL TEMA DESERCIÓN

Variable	Descripción
Beca deportiva	Indica si el alumno tiene beca deportiva
Beca de estudio	Indica si el alumno tiene beca de estudio
Beca de alimento	Indica si el alumno tiene beca de alimento
Crédito	Indica si el alumno tiene crédito
Promedio notas acumuladas	Promedio total de todas las notas del alumno en la carrera
Ranking del alumno	Posición del alumno en la carrera
Prácticas	Cantidad de prácticas realizadas por el alumno
Porcentaje ramos aprobados	Indice de aprobación del alumno, sacado de todos los ramos cursados
Motivaciones del Alumno	
Prioridad de postulación a carrera	Lugar en el orden de postulación a su actual carrera
Trabajador	Indica si el alumno se encuentra trabajando
Taller emprendimiento	Indica si el alumno ha realizado talleres o cursos relacionado a emprendimientos
Taller bienestar estudiantil	Indica si el alumno ha participado en talleres que ayuden a su estabilidad emocional o de desempeño en la institución
Taller profesional	Indica si el alumno ha participado en talleres que ayuden a su futuro profesional
Intercambio estudiantil	Indica si el alumno ha realizado un intercambio
Taller certificación	Indica si el alumno ha participado en talleres que certifiquen sus conocimientos
Taller deportivo o cultural	Indica si el alumno ha participado en talleres o ramos de tipo deportivo o cultural
Voluntariado social	Indica si el alumno ha realizado un voluntariado social o participado en una ONG

Tabla 5.1. Tablou de variables.

Las variables derivadas del estudio de los 19 constructos generan el tablou de la Tabla 5.1 que se pueden ordenar en tres grupos, el primero “Institución y Carrera”, el cual tiene información general de la institución y carrera. El segundo “Desempeño Alumno”, el cual tiene información puntual del alumno sobre si posee beneficios y datos de su desempeño. El tercero “Motivaciones

del Alumno”, el cual tiene información puntual en la participación de diversas actividades que se asocian a sus motivaciones y a una orientación vocacional.

Capítulo 6

Aplicando Minería de Datos con RapidMiner

6.1. Identificación de datos y variables

6.1.1. Recolección de Datos

Los datos a utilizar en los modelos a presentar en este capítulo son extraídos de la base de datos de la institución privada. Estos datos corresponde a información de los años 2014 y 2015 de alumnos de las distintas carreras de ingeniería que imparte la institución. Además se extrae información a través del SIES sobre la institución y sus diferentes carreras de ingeniería con respecto a acreditación e indices de empleabilidad.

Los datos proporcionados por la institución se encuentra en un archivo Excel en el cual figuran los siguientes datos: Rut del estudiante, Apellido paterno del estudiante, Apellido materno del estudiante, Nombres del estudiante, Comuna de residencia del estudiante, Ciudad de residencia del estudiante, Facultad del estudiante, Carrera del estudiante, Jornada del estudiante, PSU de lenguaje PSU de matemáticas, PSU de ciencias, PSU de historia, Año de ingreso a la carrera, Código de la carrera, Semestre de inicio de la carrera, Notas de enseñanza media, Código RBD del colegio, Tipo de dependencia del colegio, Sexo del estudiante, Puntaje ranking del estudiante, Fecha de nacimiento del estudiante, Código de región del colegio, Región del

colegio, Comuna del colegio, Nombre del colegio, Código del plan del alumno, Código del plan de la carrera, Código del semestre de inicio de la carrera, Año de información, Semestre de información, Código del curso, Código del semestre de información, Semestre de ingreso a la carrera, Código del estado del curso inscrito, Código del estado del alumno al final del curso, Estado del alumno al final del curso, Sección del curso, Código del ramo, Nota final del curso, Ramo nemotecnico, Nombre del ramo, Nivel del ramo, Tipo de ramo, Código de semestre de inicio del plan (2014), Código de semestre de inicio de la carerra (2014), Sexo del estudiante, País gentilicio, Tipo de contrato, Correo personal, Teléfono fijo, Teléfono celular, Semestre de ingreso a la carrera, Código de la facultad, Beca fallecimiento (2014),Beca de cesantia (2014), Beca internado (2014), Beca Complementaria CAE (2014), Beca Mérito (2014), Beca de Almuerzo (2014), Beca de Fotocopia (2014), Beca de Plotter (2014), Beca Valech (2014), Beca Traspaso Valech (2014), Beca Indígena (2014), Beca Hijo de Profesor (2014), Beca Integración Territorial (2014), Beca Excelencia Mineduc (2014), Beca Juan Gómez Milla (2014), Beca Mun. de Las Condes (2014), Beca Puntaje Nacional (Interno) (2014), Beca Presidente de la Republica (2014), Beca Chaitén (2014), Beca Terremoto (2014), JUNAEB (2014), CAE (2014), Si tiene beca (2014), Si tiene beca interna (2014), Si tiene beca externa (2014), Beca Vocación de Profesor (2014), Mantención JUNAEB (2014), Beca Puntaje PSU MINEDUC (2014), Beca Deportista Destacado (2014), Beca Transporte (2014), Equidad (2014), Beca Complementaria Vocación de Profesor (2014), Beca de Honor (2014), Beca Balmaceda (2014), Beca Funcionario (2014), Beca Ranking NEM (2014),Beca Articulación (2014), Otra Beca MINEDUC (2014), Estado de plan alumno (2015), Código de semestre de inicio del plan (2015), Código de semestre de inicio de la carerra (2015), Beca fallecimiento (2015), Beca de cesantia (2015), Beca internado (2015), Beca Complementaria CAE (2015), Beca Mérito (2015), Beca de Almuerzo (2015), Beca de Fotocopia (2015), Beca de Plotter (2015), Beca Valech (2015), Beca Traspaso Valech (2015), Beca Indígena (2015), Beca Hijo de Profesor (2015), Beca Integración Territorial (2015), Beca Excelencia Mineduc (2015), Beca Juan Gómez Milla (2015), Beca Mun. de Las Condes (2015), Beca Puntaje Nacional (Interno) (2015), Beca Presidente de la Republica (2015), Beca Chaitén (2015), Beca Terremoto (2015), JUNAEB (2015), CAE (2015), Si tiene beca (2015), Si tiene beca interna (2015), Si tiene beca externa (2015), Beca Vocación de Profesor (2015), Mantención JUNAEB (2015), Beca Puntaje PSU MINE-

DUC (2015), Beca Deportista Destacado (2015), Beca Transporte (2015), Equidad (2015), Beca Complementaria Vocación de Profesor (2015), Beca de Honor (2015), Beca Balmaceda (2015), Beca Funcionario (2015), Beca Ranking NEM (2015), Beca Articulación (2015), Otra Beca MINEDUC (2015).

La información del SIES se encuentra en un sitio web, por lo que se creo una base de datos en MySQL llamada *Area Staging* con toda la información recopilada.

6.1.2. Procesamiento de Datos

La información recopilada de las dos fuentes de datos fue procesada, extra yendo y transformando varios de los datos que se presentaron anteriormente. Este procesamiento se realizo a partir de las variables consideradas en los constructos y del análisis de diferentes investigaciones del tema deserción en Chile que indican que son buenos predictores del Capítulo 2.

Figura 6.1. Tabla resultante

Los datos procesados se pueden apreciar en la Figura 6.1, la cual tiene principalmente variables de la “Institución y Carrera” y del “Desempeño del Alumno” definidas en el tablou de los constructos, Tabla 5.1, faltando las variables de “Motivaciones del Alumno”, por lo que para efecto de este es-

tudio, se propone un mecanismo para capturar estos datos.

El mecanismos a proponer es generar encuestas a final de cada semestre que capte y actualice las variables correspondientes a las motivaciones del alumno, de tal modo poder analizar la situación del alumno de manera semestral.

6.1.3. Muestra de Datos

La muestra de datos esta configurada para dos conjuntos de datos, el primero lo conforma datos de alumnos de jornada diurna y el segundo conjunto por alumnos de jornada vespertina.

El conjunto de alumnos de jornada diurna esta conformado por 709 observaciones, de los cuales 11 representan abandono.

El conjunto de alumnos de jornada vespertina esta conformado por 49 observaciones, de los cuales dos representan abandono.

6.1.4. Variables

A continuación se muestra una tabla con variable para modelar los datos de los dos conjuntos, el conjunto de jornada de vespertina no se consideran los puntajes PSU debido a que la institución tiene otro mecanismo de ingreso y la mayoría de las observaciones tiene esos campos vacíos.

Variable	Tipo	Descripción
Rut	N/A	Identificador
Comuna	N/A	Nombre de las comunas de residencia
Genero	Discreto	Toma los siguientes valores, Femenino = 1; Masculino=2; F=1
Edad	Continuo	Se considera edad del año 2014
Tipo establecimiento de origen	Discreto	Toma los siguientes valores, Municipal=1; Particular Subvencionado=2; Particular Pagado =3;

Variable	Tipo	Descripción
Notas enseñanza media	Continuo	Rango entre 10 y 70
PSU lenguaje matemáticas	Continuo	Rango de 150 a 850
PSU matemáticas	Continuo	Rango de 150 a 850
PSU ciencias	Continuo	Rango de 150 a 850
Acreditación Institucional	Discreto	Toma los siguientes valores, No Acreditada=0; Acreditada=1
Carrera	N/A	Nombre de carrera
Jornada	Discreto	Toma los siguientes valores, Diurna y Vespertina
Empleabilidad al 1er año	Continuo	Rango en porcentaje 0 a 100
Ingreso promedio 4to año	Continuo	Rango en dinero
Carrera acreditada	Discreto	Toma los siguientes valores, No Acreditada=0; Acreditado=1
Promedio notas acumuladas	Continuo	Rango entre 10 y 70
Beca14	Discreto	Toma los siguientes valores, No Beca=0; beca=1
Beca15	Discreto	Toma los siguientes valores, No Beca=0; beca=1
Junaeb14	Discreto	Toma los siguientes valores, No Junaeb=0; Junaeb=1
Junaeb15	Discreto	Toma los siguientes valores, No Junaeb=0; Junaeb=1
Cae14	Discreto	Toma los siguientes valores, No Cae=0; Cae=1
Cae15	Discreto	Toma los siguientes valores, No Cae=0; Cae=1
Estado	Discreto	Toma los siguientes valores, Activo y Abandono

Tabla 6.1. Variables.

En las tablas anteriormente presentadas, se definió dos tipos de variables [20]:

- Variables Discretas: Son variables que toma un conjunto de valores determinados, es decir, no acepta cualquier valor, solo los que existen en el en conjunto, por ejemplo, Genero toma dos valores, Masculino y Femenino.
- Variables Continuas: Son variables que pueden tomar un valor fijo dentro de un intervalo, por ejemplo, los puntajes PSU puede tomar cualquier valor dentro del rango 150 a 850.

6.1.5. Procesos en RapidMiner

La construcción de los modelos se implementará con la herramienta *RapidMiner* utilizando como entrada las tablas generadas anteriormente.

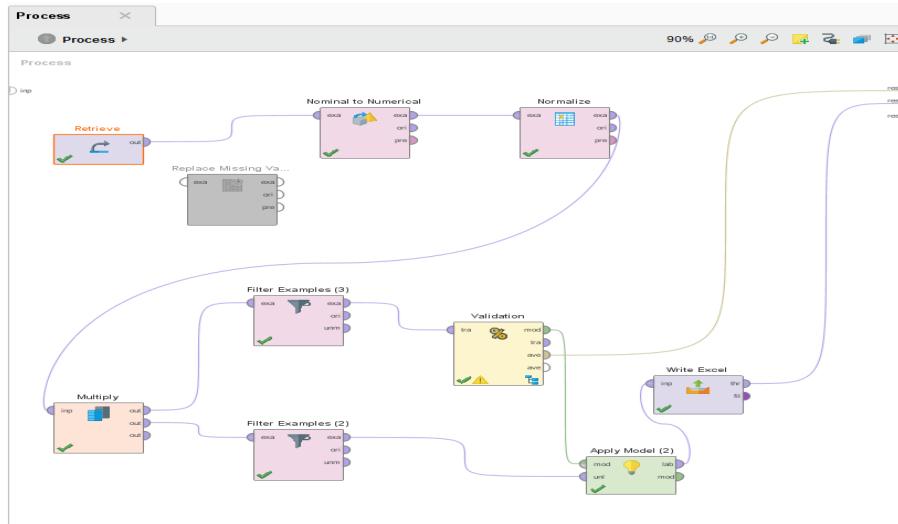


Figura 6.2. Proceso de Predicción

La herramienta *RapidMiner* permite a través de operadores, construir un proceso. Los procesos son una serie de pasos constituidos por operadores. La Figura 6.2 representa el proceso construido para generar la predicción de los datos. La descripción de cada operador se detalla en el Anexo B

Format your columns.										
	Date format	MMM d, yyyy h:mm:ss a z								<input type="checkbox"/> Replace errors with missing values
medio_... ger	junaeb14 integer	cae14 integer	beca14 integer	junaeb15 integer	cae15 integer	beca15 integer	estad... polynominal label			
1	0	0	1	0	0	0	ABANDONO			
2	0	0	0	0	1	0	ACTIVO			
3	0	0	0	1	1	1	ACTIVO			
4	0	0	0	0	1	0	ACTIVO			
5	0	1	0	0	1	1	ACTIVO			
6	0	0	0	0	0	0	ACTIVO			
7	1	1	1	1	1	0	ACTIVO			
8	0	0	0	0	0	1	ACTIVO			
9	0	1	1	0	0	0	ACTIVO			
10	0	0	0	0	0	0	ACTIVO			
11	0	0	0	0	0	0	ACTIVO			
12	0	1	0	0	0	0	ACTIVO			
13	0	1	1	0	1	1	ACTIVO			
14	0	1	1	0	0	1	ACTIVO			
15	0	0	0	0	0	0	ACTIVO			
16	0	0	0	0	0	0	ACTIVO			
17	0	0	0	0	0	0	ACTIVO			
18	0	1	0	0	1	0	ACTIVO			

Figura 6.3. Importación de datos

El primer paso antes de operar el proceso construido, es importar los datos para entrenar y predecir. Las variables se definen con el tipo de dato que las caracteriza junto al rol a cumplir dentro del modelo, para este caso, se tomo como rol identificador la variable “rut” y como variable a predecir “estado”, Figura 6.3.

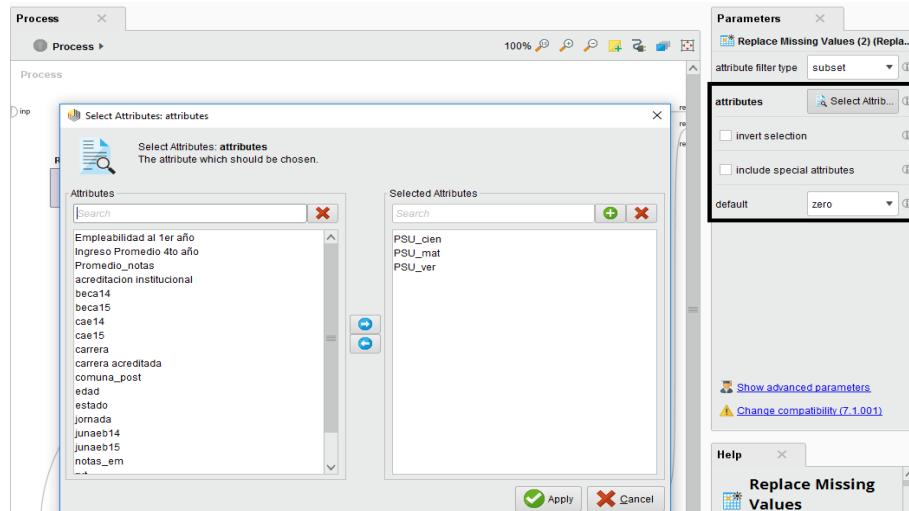


Figura 6.4. Reemplazar de datos vacíos

Los datos cargados son accedidos por el operador “Retrieve” y luego pasados hacia el operador “Replace Missing Values” el cual reemplaza los valores vacíos de las variables seleccionadas, en este caso se seleccionaron las varia-

bles de “PSU” los cuales tienen datos vacíos, estos datos serán reemplazados por ceros como se puede ver en la Figura 6.4.

Luego de que los datos son reemplazados, pasan al operador “Nominal to Numerical”, el cual transforma todas la variables a numéricas, para luego pasar al operador “Normalize”, normalizando los datos y luego al operador “Multiply” el cual multiplica los datos generando dos conjuntos, el primero para ser entrenado y el otro para validar, separados por los operadores “Filter Example”.

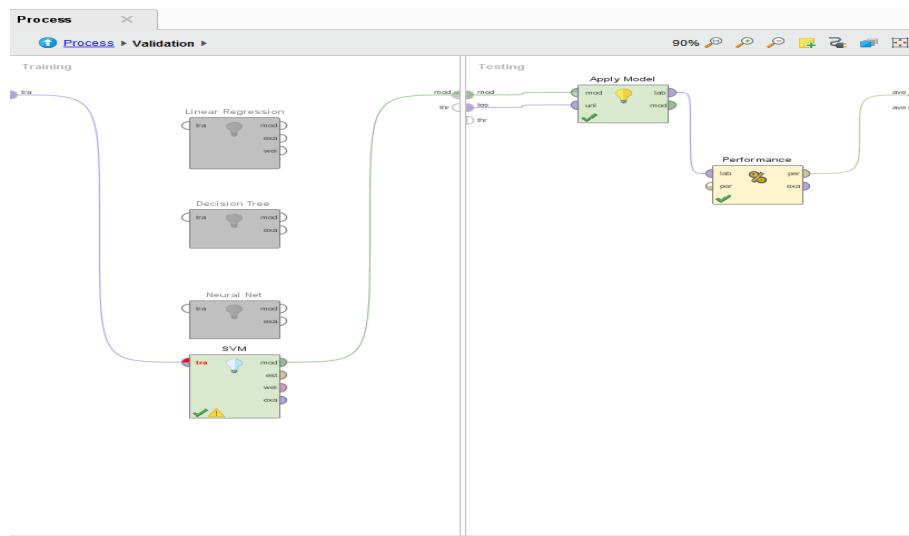


Figura 6.5. Validación

En el operador “Validation” se definen el porcentaje de datos a entrenar del total del conjunto de datos, en este caso se especifica un 70 %. En este operador, se puede seleccionar los operadores de predicción, que para este estudio serán los cuatro que se muestran en la Figura 6.5.

6.2. Modelamiento

El modelamiento se configuro de la siguiente manera, para el conjunto de datos de jornada diurna, se replicaron 53 observaciones dejando la variable a predecir vacía. Y para el conjunto de datos de jornada vespertina, se replicaron 21 observaciones dejando la variable a predecir vacía.

El detalle del resultado de las cuatro técnicas de predicción utilizadas en los conjuntos de datos, se especifica la matriz de confusión, las variables que más influyeron en cada predicción y el porcentaje real de predicción.

La matriz de confusión es una herramienta que permite visualizar el desempeño de un algoritmo empleado para aprendizaje supervisado. Las columnas representan el número de predicciones por clases y las filas representan las instancias de la clase. Esta herramienta permite ver si el sistema está confundiendo dos clases [21].

6.2.1. SVN

- Jornada Diurna

accuracy: 98.12%			
	true ABANDONO	true ACTIVO	class precision
pred. ABANDONO	0	3	0.00%
pred. ACTIVO	1	209	99.52%
class recall	0.00%	98.58%	

Figura 6.6. Matriz de Confusión SVN Diurno

La matriz de confusión que se muestra en la Figura 6.6 indica que SVN obtuvo una precisión del 98.12 %, especificando que la clase ACTIVO tuvo una precisión del 99.52 %, mientras que la clase ABANDONO tuvo 0 %. Sin embargo comparando las observaciones predichas con las observaciones reales, se obtuvo una precisión del 100 %.

Las variables con mayor peso en la predicción de la clase ACTIVO fueron: CAE15, PSU Ciencias, Beca15, Promedio notas, Carrera, Empleabilidad al 1er año, Junaeb15, NEM, Ingreso Promedio 4to año, Tipo colegio, Comuna y Genero.

- Jornada Vespertina

accuracy: 64.29%			
	true ACTIVO	true ABANDONO	class precision
pred. ACTIVO	9	0	100.00%
pred. ABANDONO	5	0	0.00%
class recall	64.29%	0.00%	

Figura 6.7. Matriz de Confusión SVN Vespertino

La matriz de confusión que se muestra en la Figura 6.7 indica que SVN obtuvo una precisión del 64.29 %, especificando que la clase ACTIVO tuvo una precisión del 100 %, mientras que la clase ABANDONO tuvo 0 %. Sin embargo comparando las observaciones predichas con las observaciones reales, se obtuvo una precisión del 100 %.

Las variables con mayor peso en la predicción de la clase ACTIVO fueron: Edad, NEM, Genero, Comuna, Junaeb14 y Junaeb15.

6.2.2. Regresión Lineal

- Jornada Diurna

accuracy: 72.74%			
	true ABANDONO	true ACTIVO	class precision
pred. ABANDONO	3	161	1.83%
pred. ACTIVO	5	440	98.88%
class recall	37.50%	73.21%	

Figura 6.8. Matriz de Confusión Regresión Lineal Diurno

La matriz de confusión que se muestra en la Figura 6.8 indica que Regresión Lineal obtuvo una precisión del 72.74 %, especificando que la clase ACTIVO tuvo una precisión del 98.88 %, mientras que la clase ABANDONO tuvo 1.83 %. Sin embargo comparando las observaciones predichas con las observaciones reales, se obtuvo una precisión del 81.13 %.

Las variables con mayor peso en la predicción de la clase ACTIVO fueron: NEM, CAE15, PSU Ciencias, Tipo colegio, Junaeb14, Promedio

notas, Comuna y Edad.

- Jornada Vespertina

accuracy: 100.00%			
	true ACTIVO	true ABANDONO	class precision
pred. ACTIVO	14	0	100.00%
pred. ABANDONO	0	0	0.00%
class recall	100.00%	0.00%	

Figura 6.9. Matriz de Confusión Regresión Lineal Vespertino

La matriz de confusión que se muestra en la Figura 6.7 indica que Regresión Lineal obtuvo una precisión del 100 %, especificando que la clase ACTIVO tuvo una precisión del 100 %, mientras que la clase ABANDONO tuvo 0 %. Sin embargo comparando las observaciones predichas con las observaciones reales, se obtuvo una precisión del 95.23 %.

Las variables con mayor peso en la predicción de la clase ACTIVO fueron: CAE14, Comuna y Beca15.

6.2.3. Red Neuronal

- Jornada Diurna

accuracy: 98.19%			
	true ABANDONO	true ACTIVO	class precision
pred. ABANDONO	0	3	0.00%
pred. ACTIVO	8	598	98.66%
class recall	0.00%	99.50%	

Figura 6.10. Matriz de Confusión Red Neuronal Diurno

La matriz de confusión que se muestra en la Figura 6.10 indica que Red Neuronal obtuvo una precisión del 98.19 %, especificando que la clase

ACTIVO tuvo una precisión del 98.68 %, mientras que la clase ABANDONO tuvo 0 %. Sin embargo comparando las observaciones predichas con las observaciones reales, se obtuvo una precisión del 94.34 %.

Se genero 127 nodos en donde las variables con mayor peso en la predicción de la clase ACTIVO fueron: NEM, Comuna y Edad.

- Jornada Vespertina

accuracy: 78.57%			
	true ACTIVO	true ABANDONO	class precision
pred. ACTIVO	11	0	100.00%
pred. ABANDONO	3	0	0.00%
class recall	78.57%	0.00%	

Figura 6.11. Matriz de Confusión Red Neuronal Vespertino

La matriz de confusión que se muestra en la Figura 6.11 indica que Red Neuronal obtuvo una precisión del 78.57 %, especificando que la clase ACTIVO tuvo una precisión del 100 %, mientras que la clase ABANDONO tuvo 0 %. Sin embargo comparando las observaciones predichas con las observaciones reales, se obtuvo una precisión del 100 %.

Se genero 22 nodos en donde las variables con mayor peso en la predicción de la clase ACTIVO fueron: tipo de colegio, Comuna, Promedio notas, Junaeb14, CAE14, Beca14, Junaeb15, CAE15 y Beca15.

6.2.4. Árbol de Decisión

- Jornada Diurna

accuracy: 95.57%			
	true ABANDONO	true ACTIVO	class precision
pred. ABANDONO	0	19	0.00%
pred. ACTIVO	8	582	98.64%
class recall	0.00%	96.84%	

Figura 6.12. Matriz de Confusión

La matriz de confusión que se muestra en la Figura 6.12 indica que Árbol de Decisión obtuvo una precisión del 95.57 %, especificando que la clase ACTIVO tuvo una precisión del 98.64 %, mientras que la clase ABANDONO tuvo 0 %. Sin embargo comparando las observaciones predichas con las observaciones reales, se obtuvo una precisión del 81.13 %.

La variable con mayor peso en la predicción de la clase ACTIVO fue: Comuna.

- Jornada Vespertina

accuracy: 100.00%			
	true ACTIVO	true ABANDONO	class precision
pred. ACTIVO	14	0	100.00%
pred. ABANDONO	0	0	0.00%
class recall	100.00%	0.00%	

Figura 6.13. Matriz de Confusión

La matriz de confusión que se muestra en la Figura 6.13 indica que Árbol de Decisión obtuvo una precisión del 100 %, especificando que la clase ACTIVO tuvo una precisión del 100 %, mientras que la clase ABANDONO tuvo 0 %. Sin embargo comparando las observaciones predichas con las observaciones reales, se obtuvo una precisión del 90.48 %.

La variable con mayor peso en la predicción de la clase ACTIVO fue: Promedio de notas.

En ambos casos de los conjuntos probados en los cuatro algoritmos utilizados, las matrices de confusión tienen un sesgo hacia la predicción ACTIVO, por lo que en todos los casos se realizó una comparación de las observaciones a predecir con las observaciones que tienen definido su estado.

6.3. Resultados

Parte III

Conclusiones

Capítulo 7

Conclusiones y recomendaciones

A partir del estudio de la deserción según estadísticas del SIES, el tema de deserción universitaria, no es un fenómeno aislado, aproximadamente el 50 % de los alumnos que ingresan a un programa de estudios universitario, no termina la carrera, preocupando a varios actores del sistema.

Los datos recopilados por el SIES, orientan en cierto sentido el panorama que se a presentado entre los años 2010 y 2014 sobre el tema deserción, tomando variables globales como institución, género, jornada, establecimiento de origen y acreditación.

El concepto de inteligencia del cliente, entrego herramientas para analizar las necesidades y relacionamiento que tiene el alumno con la universidad, esto ayudo a identificar a través de entrevistas, variables influyentes en continuar y/o desertar de la carrera. Las variables determinadas con inteligencia del cliente junto a las variables de información del SIES, ayudaron al proceso de inteligencia de negocio, las cuales fueron extraídas, transformadas y cargadas en una base analítica, para finalizar aplicando el concepto minería de datos y generar los modelos con la herramienta RapidMiner.

Parte IV

Información Adicional

Referencias bibliográficas

- [1] Olap vs oltp. [Online]. Available: <http://datawarehouse4u.info/OLTP-vs-OLAP.html>
- [2] C. A. Veliz., “*Acceso y deserción en la educación superior: caso aplicado a Chile,*” *Tesis de Magister. Universidad de Chile, Santiago, Chile,* 2012.
- [3] T. Larroucau., “*Estudio de los factores determinantes de la deserción en el sistema universitario chileno,*” *Tesis de Magister. Universidad de Chile, Santiago, Chile,* 2013.
- [4] S. Celis., “*Student attrition and student time-to-degree at a selective engineering school in Chile,*” *Documento interno, Escuela de Ingeniería y Ciencias, Universidad de Chile,* 2012.
- [5] C. Díaz., “*Factores de deserción estudiantil en ingeniería: Una aplicación de modelos de duración,*” *Información Tecnológica, 20(5):129–145,* 2000.
- [6] M. de Educación, “Instituciones de educación superior vigentes - 30 diciembre de 2015,” p. 1, Dec. 2015. [Online]. Available: <http://www.mifuturo.cl/index.php/servicio-de-informacion-de-educacion-superior/listado-de-instituciones-vigentes-2015>
- [7] SIES, “Principales resultados evolución retención de 1er año 2010 - 2014,” pp. 1–3, Nov. 2015. [Online]. Available: <http://www.mifuturo.cl/index.php/informes-sies/retencion-de-pregrado>
- [8] Sies. [Online]. Available: <http://www.mifuturo.cl/index.php/academicos-einvestigadores>
- [9] ¿qué es la inteligencia del cliente? [Online]. Available: <http://www.bewitt.es/que-es-la-inteligencia-del-cliente/>
- [10] Definición inteligencia del cliente. [Online]. Available: https://es.wikipedia.org/wiki/Inteligencia_del_cliente
- [11] Customer relationship management. [Online]. Available: <http://fse-todo.com/doc/5874/index.html?page=5>

REFERENCIAS BIBLIOGRÁFICAS

- [12] Inteligencia de negocios. [Online]. Available: <http://www.gestiopolis.com/inteligencia-de-negocios-business-intelligence/>
- [13] Bases de datos olap y oltp. [Online]. Available: http://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx
- [14] Herramientas etl. [Online]. Available: <https://blogs.deusto.es/bigdata/herramientas-etl-y-su-relevancia-en-la-cadena-de-valor-del-dato/>
- [15] Regresión lineal. [Online]. Available: http://docs.rapidminer.com/studio/operators/modeling/predictive/functions/linear_regression.html
- [16] Árboles de decisión. [Online]. Available: http://docs.rapidminer.com/studio/operators/modeling/predictive/trees/parallel_decision_tree.html
- [17] Redes neuronales. [Online]. Available: http://docs.rapidminer.com/studio/operators/modeling/predictive/neural_nets/neural_net.html
- [18] Svn. [Online]. Available: http://docs.rapidminer.com/studio/operators/modeling/predictive/support_vector_machines/support_vector_machine.html
- [19] Muestreo no probabilístico. [Online]. Available: <http://www.estadistica.mat.uson.mx/Material/elmuestreo.pdf>
- [20] Variables discretas y continuas. [Online]. Available: https://es.wikipedia.org/wiki/Variable_discreta_y_variable_continua
- [21] Matriz de confusión. [Online]. Available: https://es.wikipedia.org/wiki/Matriz_de_confusi%C3%B3n



ANEXOS

Anexo A

Entrevistas con Metodología de la Metáfora

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

ENTREVISTA 1

Nombre: Marcel

Edad: 25

Carrera: 4to Año de Ingeniería Civil Informática y Telecomunicaciones

Comuna: Santiago Centro

Número	Imagen	Descripción	Asociaciones
1		“Una casa en la playa, ideal para pasear con la familia”	“Representa el compartir con la familia con una situación económica estable”
2		“Una botella de champange representa celebración y diversión”	“Tener un título, se asocia a disfrutar placeres de la vida”
3		“Un plato de comida especial representa los gustos exclusivos que uno se puede dar”	“Tener una buena situación económica”
4		“Tener una casa grande para la familia”	“El terminar una carrera implica tener un buen trabajo y mucho dinero, con el cual se puede optar a mejores condiciones de vida”
5		“Participación activa en la industria”	“Generar aportes al trabajo, ser un buen profesional”
6		“Tener una buena familia”	“Generar estabilidad y buenos lazos”
7		“Dar la mano”	“Generar buenos lazos profesionales y hacer negocios”

Tabla A.1. Entrevista 1.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 2 y 3, mencionando que son los placeres y gustos que se puede dar uno en la vida.

Grupo 2: imágenes 1, 4 y 6, mencionando el lado familiar, de tener algo consolidado y estable.

Grupo 3: imágenes 5 y 7, mencionando lado profesional, de tener estabilidad.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Los grupos 1 y 3 se complicarían ya que no se podrían lograr sin un título, en cambio el grupo 2 siempre estará porque es la familia”

ENTREVISTA 2

Nombre: Fabián

Edad: 23

Carrera: 4to Año Ingeniería Civil Informática y Telecomunicaciones

Comuna: Puente Alto

Número	Imagen	Descripción	Asociaciones
1		“Casa grande”	“Casa acogedora y armónica tanto para él como para su familia”
2		“Diseño de espacio vanguardista y acogedor”	“Tener un espacio personal grande y vanguardista”
3		“Casa en el lago”	“Espacio compartido, acogedor”

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Número	Imagen	Descripción	Asociaciones
4		“Cultura y música”	“Tener una buena situación económica para disfrutar eventos culturales”
5		“Pareja feliz”	“Compañera de vida bajo la premisa de familia, poder ofrecer un buen vivir a través de ser un profesional”
6		“Redes profesionales”	“Generar proyectos de ingeniería”
7		“Redes informáticas”	“Tener un trabajo relacionado a redes informáticas y ser un buen profesional, para viajar por el mundo”
8		“Niños felices”	“Sueños de ser papá, de tener las herramientas para dar un buen pasar”

Tabla A.2. Entrevista 2.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 4, 6 y 7, mencionando los gustos personales, ligando la parte profesional y la consecuencia de tener el dinero para darse los gustos.

Grupo 2: imágenes 5 y 8, mencionando la parte familiar.

Grupo 3: imágenes 1, 2, 3 mencionando la parte del hogar, una parte de acogida y en donde compartir.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Generaría mucha frustración, ya que la carrera es una herramienta para poder tener todo lo de las imágenes”

ENTREVISTA 3

Nombre: Cristian

Edad: 25

Carrera: 5to Año Ingeniería Civil Informática y Telecomunicaciones

Comuna: La Florida

Número	Imagen	Descripción	Asociaciones
1		“Auto”	“Tener un auto para tener independencia”
2		“Guitarra”	“Tener sus gustos, la música”
3		“Seguridad informática”	“Poder dedicarse a la seguridad informática”
4		“Departamentos”	“Tener un lugar donde vivir”
5		“Mundo”	“Poder realizar viajes por el mundo”
6		“Cerveza”	“Poder darse sus gustos”
7		“La familia”	“Tener una familia y tener plata para mantener a la familia”

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Número	Imagen	Descripción	Asociaciones
8		“Decoración”	“Tener su casa decorada”

Tabla A.3. Entrevista 3.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 2 y 5, mencionando gustos personales, viajar a diferentes partes del mundo

Grupo 2: imágenes 3, 4 y 7, mencionando la parte familiar, trabajando en lo que le gusta, teniendo su casa y familia.

Grupo 3: imágenes 6 y 8, mencionando que son gustos exclusivos, no tan importantes como los del grupo 1.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Solamente podría tener la parte familiar, los otros grupos serían bastante complicados de satisfacer”

ENTREVISTA 4

Nombre: Manuel

Edad: 24

Carrera: 4to Año Ingeniería Civil Informática y Telecomunicaciones

Comuna: Providencia

Número	Imagen	Descripción	Asociaciones
1		“Linux”	“Tener sus trabajos realizados en Linux”

Número	Imagen	Descripción	Asociaciones
2		“Mujer”	“Tener una linda mujer”
3		“Pan”	“Tener para comer y una vida sana”
4		“Auto”	“Tener suficiente trabajo para comprar un auto familiar”
5		“Redes informáticas”	“Tener un trabajo en el área de redes informáticas”
6		“Seguridad informática”	“Poder desempeñarse en el área de seguridad informática”
7		“Madre e hija”	“Tener una familia con una hija sana”

Tabla A.4. Entrevista 4.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 2, 3, 4 y 7, mencionando que es la vida familiar.

Grupo 2: imágenes 1, 5 y 6 mencionando que es la vida trabajo.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“No habría suficiente dinero para sustentar todo”

ENTREVISTA 5

Nombre: Cristian

Edad: 23

Carrera: 4to Año Psicología

Comuna: Antofagasta

Número	Imagen	Descripción	Asociaciones
1		“Inexperto”	“Falta de experiencia al terminar la carrera, motiva a trabajar”
2		“Presión”	“Cerrar una etapa de la vida, para pasar a una siguiente”
3		“Profesional”	“Estar listo para enfrentar el mundo laboral”
4		“Biblioteca”	“Dejar atrás los libros”
5		“Superación”	“Sentir la sensación de éxito”
6		“Deudas”	“Tener mayores responsabilidades”
7		“Magister”	“Poder sacar un postgrado”

Tabla A.5. Entrevista 5.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 3, 4 y 5, mencionando que representa el éxito y el sentirse realizado por completar la carrera

Grupo 2: imágenes 1, 2, 6 y 7, mencionando la incertidumbre de enfrentar el mundo laboral.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Afectaría en la complicación de tener que pagar el crédito con aval del estado”

ENTREVISTA 6

Nombre: Gabriel

Edad: 24

Carrera: 5to Año Psicología

Comuna: Santiago Centro

Número	Imagen	Descripción	Asociaciones
1		“Librero”	“Tener muchos libros por el gusto de leer”
2		“Psicología clínica”	“Dedicarse a su profesión”
3		“Graduado”	“Significa el cierre y el comienzo de una nueva etapa”“
4		“Autonomía”	“Ser una persona independiente”

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Número	Imagen	Descripción	Asociaciones
5		“Deudas”	“Al titularse se tendrán muchas deudas”
6		“Profesor”	“Poder dar clases en un futuro”

Tabla A.6. Entrevista 6.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 2, 3 y 6, mencionando que representa la independencia

Grupo 2: imágenes 4 y 5, mencionando que refleja el capitalismo y que siempre habrán deudas.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Afectaría en la complicación de tener independencia”

ENTREVISTA 7

Nombre: Tania

Edad: 24

Carrera: 4to Año Ingeniería Civil Industrial

Comuna: Peñalolen

Número	Imagen	Descripción	Asociaciones
1		“Libertad laboral”	“Trabajar de forma independiente”

Número	Imagen	Descripción	Asociaciones
2		“Salud”	“Poder tener el dinero para obtener buenos tratamientos médicos”
3		“¿Qué hacer?”	“Necesidad de dinero y experiencia”
4		“Estilo de vida”	“Vida consciente con el planeta, a largo plazo”
5		“Seguridad”	“Desarrollar seguridad personal”
6		“Viajes”	“Tener la disponibilidad de tiempo y dinero, salir de la rutina, tener una vida más relajada”
7		“Trabajo estable”	“Trabajar con personas, tener vida social”

Tabla A.7. Entrevista 7.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 4 y 6, mencionando que representa tranquilidad y bienestar espiritual, relacionando el trabajo independiente que da libertad, complementándose con su estilo de vida y viajes.

Grupo 2: imágenes 2 y 5, mencionando que representa seguridad, tener seguridad con su salud y frente a la delincuencia.

Grupo 3: imágenes 3 y 7, mencionando que representa el ámbito laboral, el no saber en que trabajará, pero si tener una idea de que trabajará con un

grupo de gente y hacer vida social.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Solo se verían cumplidas las imágenes del grupo 1 y muy a largo plazo, los otros grupos no los podría cumplir”

ENTREVISTA 8

Nombre: Carlos

Edad: 22

Carrera: 3er Año Ingeniería Comercial

Comuna: Santiago Centro

Número	Imagen	Descripción	Asociaciones
1		“Tranquilidad”	“Terminar una carrera transmite seguridad y sentirse útil para algo”
2		“Presentación de Apple”	“Representa el triunfo que puede llegar a tener como profesional.”
3		“Triunfo”	“Reconocimiento por llegar a una meta”
4		“Tiempo libre”	“Expresa los pasatiempos que se pueden realizar al terminar la carrera”
5		“Pasatiempos”	“Poder tener la libertad de darse tiempos libres y hacer las cosas que uno quiere”
6		“Trabajo”	“Trabajar de forma relajada, ya que con el título tienes una base para dedicarse a lo que te gusta”

Número	Imagen	Descripción	Asociaciones
--------	--------	-------------	--------------

Tabla A.8. Entrevista 8.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 4 y 5, mencionando que representa el tiempo libre, que ya obtenido un título, ya tienes una base para poder tener dinero y dedicarse a gustos más caros.

Grupo 2: imágenes 2, 3 y 6, mencionando que representa el triunfo, el éxito de terminar una meta.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“En lo personal tener un título no es el único camino para cumplir mis sueños, la sociedad impone que debe ser así, pero yo no lo veo como la única opción.”

ENTREVISTA 9

Nombre: Ilana

Edad: 23

Carrera: 3er Año Ingeniería Comercial

Comuna: Las Condes

Número	Imagen	Descripción	Asociaciones
1		“Comida”	“Dedicar tiempo para mi persona y realizar lo que me gusta, la cocina.”
2		“Curso de inglés”	“Tener tiempo para poder realizar un curso de inglés y viajar al extranjero.”
3		“Libertad”	“Terminar la carrera, da una sensación de libertad y triunfo.”

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Número	Imagen	Descripción	Asociaciones
4		“Viajes”	“Después de la carrera hay tiempo para disfrutar y conocer el mundo.”
5		“Independencia”	“Poder vivir sola y tener mi departamento.”
6		“Trabajo”	“Encontrar un trabajo en una empresa prestigiosa.”

Tabla A.9. Entrevista 9.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1 y 3, mencionando que representa el relajo y el poder realizar sus gustos, como cocinar.

Grupo 2: imágenes 2 y 4, mencionando que representa los viajes, y que aprender un idioma nuevo abre las puertas a conocer otros lugares.

Grupo 3: imágenes 5 y 6, mencionado que representa su independencia, para poder tener su departamento, necesita de un trabajo.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Sería difícil cumplir el grupo 3, no podría hacer el grupo 2, solamente me podría dedicar a las imágenes del grupo 1.”

ENTREVISTA 10

Nombre: Oscar

Edad: 28

Carrera: 6to Año Ingeniería Civil en Informática y Telecomunicaciones

Comuna: Macul

Número	Imagen	Descripción	Asociaciones
1		“Scooter”	“Representa el vehículo que podría adquirir teniendo recursos”
2		“Departamento”	“Representa la adquisición de un bien raíz gracias al trabajo después de titulado”
3		“Aire acondicionado”	“Representa las comodidades que uno podría tener una vez trabajando”
4		“Plato de comida”	“Representa los gustos que me podría dar, teniendo recursos.”
5		“Buda”	“Representa crecimiento personal y laboral”
6		“Descanso”	“Representa los viajes, los gustos, el relajo que podría tener”
7		“Arte”	“Representa al parte artística que podría dedicarme teniendo recursos”
8		“Viajes”	“Representa los lugares que podría conocer”

Tabla A.10. Entrevista 10.

Los grupos de imágenes que escogió el entrevistado fueron:

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Grupo 1: imágenes 2, 3 y 6, mencionando que representa su parte hogar.

Grupo 2: imagen 4, mencionando que representa sus gustos.

Grupo 3: imágenes 1, 5, 7 y 8, mencionando que representa su parte personal, sus sueños como persona.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Para cumplir todo esto, se necesitan recursos, y en el fondo la carrera me da un pie laboral para tener una buena situación económica, sin recursos se vería todo frustrado.”

ENTREVISTA 11

Nombre: Diego

Edad: 20

Carrera: 2do Año Ingeniería Civil Industrial

Comuna: Quinta Normal

Número	Imagen	Descripción	Asociaciones
1		“Gustos”	“Poder obtener gustos exclusivos gracias al dinero.”
2		“Hogar”	“Poder conseguir un buen hogar a partir del dinero”
3		“Empresario”	“Es a lo que uno aspira ser como ingeniero, teniendo buen puesto y reputación.”
4		“Superación personal”	“Es el logro de cumplir con el objetivo de ser ingeniero”
5		“Pareja”	“Felicidad y estabilidad amorosa, gracias a una buena situación económica.”
6		“Música”	“Cumplir con el sueño de ser músico una vez teniendo estabilidad económica”
7		“Viajes”	“Sin dinero no se puede viajar, y teniendo una carrera, uno puede ganar mucho dinero.”

Tabla A.11. Entrevista 11.

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 3 y 4, mencionando que representa el trabajo donde señala que es su aspiración a ser y su logro personal

Grupo 2: imágenes 2 y 5, mencionando que representa la familia, donde espera una relación de pareja y el hogar

Grupo 3: imágenes 1, 6 y 7, mencionando que representa sus gustos personales.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“No se podría conseguir el grupo 2 por falta de dinero al igual que el grupo 3, y el grupo 1 se vería frustrado, por el punto de que quiero ser ingeniero.”

ENTREVISTA 12

Nombre: Matias

Edad: 21

Carrera: 2do Año Ingeniería Civil Industrial

Comuna: Rancagua

Número	Imagen	Descripción	Asociaciones
1		“Gustos”	“Tener un sueldo millonario para disfrutar de los gustos”
2		“Familia”	“Ser solvente”
3		“Departamento”	“Tener la autonomía de vivir solo y tener un hogar”

Número	Imagen	Descripción	Asociaciones
4		“Pasatiempos”	“Con un titulo se puede obtener mucha plata y con ella poder tener pasatiempos”
5		“Placeres”	“Comprar cosas exclusivas”
6		“Reconocimiento”	“Estar graduado y ser reconocido como un ingeniero.”
7		“Grupo de empresarios”	“Es lo que quiero llegar hacer, y tener un trabajo luego de titularme.”
8		“Comida”	“Tener gustos sin depender económicamente de otras personas”

Tabla A.12. Entrevista 12.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 6 y 7, mencionando que representa el futuro profesional, los logros y experiencia que va a lograr.

Grupo 2: imágenes 2 y 3, mencionando que representa la familia, ser solvente con la familia gracias a la carrera.

Grupo 3: imágenes 4, 5 y 8, mencionando que representa los gustos y pasatiempos que se pueda dar a través de su trabajo.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“No podría cumplir ninguno de estos sueños”

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

ENTREVISTA 13

Nombre: Maximiliano

Edad: 24

Carrera: 3er Año Ingeniería Civil Industrial

Comuna: Macul

Número	Imagen	Descripción	Asociaciones
1		“Profesores”	“Motivan y te enseñan hacer trabajos de calidad”
2		“Seguridad”	“Tener una carrera te da seguridad de tener un buen trabajo”
3		“Tiempo”	“Un profesional termina las cosas de manera más rápida, porque está preparado.”
4		“Ingeniería”	“Una disciplina que te permite hacer muchas cosas, no hay límites”
5		“Mundo”	“Como te observan las otras personas y como me proyectó como profesional”
6		“Hijos”	“Poder dar una buena educación a los hijos”
7		“Trabajo en grupo”	“Poder relacionarse con otras personas y poder tener un trabajo sinérgico”

Número	Imagen	Descripción	Asociaciones
8		“Tarjeta de credito”	“Poder acceder a muchas cosas, como bienes”

Tabla A.13. Entrevista 13.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 4, 5 y 6, mencionando que representa lo que él puede dar teniendo un título.

Grupo 2: imágenes 1, 2 y 7 mencionando que representa las herramientas que te puede dar la universidad.

Grupo 2: imágenes 3 y 8, mencionando que representa los gustos y lo que se puede acceder.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Los grupos 1 y 2 si son malos, ejemplo profesores malos, son motivos para desertar de la carrera, al igual que el no poder hacer lo que quiero hacer, que se representa en el grupo 1.”

ENTREVISTA 14

Nombre: Ariel

Edad: 23

Carrera: 3er Año Ingeniería Civil Industrial

Comuna: Vitacura

Número	Imagen	Descripción	Asociaciones
1		“Gustos”	“Poder darse gustos, sin importar lo barato o caro”

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Número	Imagen	Descripción	Asociaciones
2		“Telefonos”	“Poder comprar tecnología, gracias al buen sueldo que tendré”
3		“Bar”	“Poder tener gustos exclusivos”
4		“Ropa”	“Tener un buen sueldo, me permite comprar mucha ropa.”
5		“Vacaciones”	“Poder ir a buenos lugares y tener relajo”
6		“Comedor”	“Tener un lugar agradable dentro de la casa”
7		“Tranquilidad”	“Poder tener estabilidad económica y por ende tranquilidad”
8		“Casa”	“Poder acceder a una buena casa en un buen barrio”

Tabla A.14. Entrevista 14.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 2, 3 y 4, mencionando que representa las cosas que podría comprar día a día.

Grupo 2: imágenes 5 y 7, mencionando que representa el relajo y las vacaciones que se podría dar siendo ingeniero.

Grupo 3: imágenes 6 y 8, mencionando que representa el tener la casa propia.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Entre a estudiar más que por el gusto a la ingeniería como tal, trate de buscar algo que me diera el balance entre algo que soy bueno y que me diera estabilidad económica, y dejar la carrera no podría obtener los gustos que señale en las imágenes”

ENTREVISTA 15

Nombre: Maximiliano

Edad: 23

Carrera: 4to Año Ingeniería Civil Informática y Telecomunicaciones

Comuna: Maipú

Número	Imagen	Descripción	Asociaciones
1		“Lujo”	“Poder vivir con muchos lujos”
2		“Paisaje”	“Vivir en un paisaje fantástico, sin estrés”
3		“Biblioteca”	“Un buen vivir, pude cumplir uno de mis gustos que es leer y despreocuparme de la vida”
4		“Hotel”	“Poder viajar y hospedarse en buenos hoteles”

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Número	Imagen	Descripción	Asociaciones
5		“Paraiso”	“Poder vivir en un lugar así, tranquilo”
6		“Viajes”	“Disfrutar la vida”

Tabla A.15. Entrevista 15.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 3, 4 y 6, mencionando que representa los placeres y los gustos que se puede dar teniendo un título.

Grupo 2: imágenes 2 y 5 mencionando que representa el lugar perfecto donde vivir.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“No podría conseguir nada, porque no tendría una situación económica alta para poder darme los lujos y gustos que deseo.”

ENTREVISTA 16

Nombre: Ignacio

Edad: 25

Carrera: 5to Año Ingeniería Civil Informática y Telecomunicaciones

Comuna: Lampa

Número	Imagen	Descripción	Asociaciones
1		“Viajes”	“Poder tener una vida llena de viajes y poder descubrir el mundo”

Número	Imagen	Descripción	Asociaciones
2		“Dinero”	“Ser un profesional significa poder tener mucho dinero”
3		“Casa”	“Cumplir con mi sueño de poder tener una casa grande y lujosa”
4		“Objetos”	“Poder adquirir las cosas que siempre quise”
5		“Comida”	“Disfrutar de buenos platos”
6		“Vino”	“Poder disfrutar de buenos placeres”
7		“Plataformas móviles”	“Poder dedicarme al desarrollo de aplicaciones móviles”

Tabla A.16. Entrevista 16.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 4, 5 y 6, mencionando que representa los objetos y placeres que puede obtener.

Grupo 2: imágenes 2, 3 y 7, mencionando que representa su entorno familiar y profesional.

¿Cómo se ven afectadas tus motivaciones si desertas?.

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

“Claramente me costaría mucho más poder conseguir lo que sueño, tendría que buscar otro camino para poder ganar el dinero suficiente.”

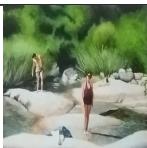
ENTREVISTA 17

Nombre:Carla

Edad: 20

Carrera: 1er Año Ingeniería Civil Industrial

Comuna: Santiago

Número	Imagen	Descripción	Asociaciones
1		“Casa”	“Tener una casa amplia y con bonita vista para recibir a la familia”
2		“Tranquilidad”	“Poder tener lo necesario para poder estar tranquila”
3		“Pasatiempo”	“Me gusta mucho nadar y muchas actividades similares, pero para eso necesito dinero”
4		“Vacaciones”	“Libertad en poder ir a lugares diferentes y con mucha naturaleza”
5		“Fundación”	“Crear una fundación para ayudar a niños con discapacidad”
6		“Negocio”	“Poder emprender con un negocio gourmet ”
7		“Tranquilidad”	“Tener una vida segura y feliz”

Número	Imagen	Descripción	Asociaciones
--------	--------	-------------	--------------

Tabla A.17. Entrevista 17.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 2 y 3, mencionando que representa su vida en la ciudad como tener un departamento, disfrutar sus lujos y hacer deportes.

Grupo 2: imágenes 4 y 7, mencionando que representa sus vacaciones soñadas, en donde ella prefiere lugares bien tranquilos con mucha naturaleza.

Grupo 3: imágenes 5 y 6, mencionando que representa sus metas y sueños a realizar en el futuro.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Si desertara por cualquier motivo creo trataría de buscar otro camino, otra carrera para poder cumplir mis objetivos, ya que no podría cumplir algunas, ya que requiero de dinero”

ENTREVISTA 18

Nombre: Gabriel G.

Edad: 19

Carrera: 2do año Ingeniería Civil Obras Civiles

Comuna: El Bosque

Número	Imagen	Descripción	Asociaciones
1		“Mi Presente”	“Me gusta estudiar, me gusta los números y ser una persona pensativa”
2		“Familia”	“A futuro poder formar una familia, teniendo ya mi carrera”

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Número	Imagen	Descripción	Asociaciones
3		“Viajes”	“Poder disfrutar viajes con mi familia”
4		“Hogar”	“Con el dinero que ganaré poder tener mi casa propia”
5		“Futuro”	“Me gustaría tener mis propias cosas con el dinero que ganaré”
6		“Trabajo”	“Me gustaría poder liderar obras y aplicar todo lo aprendido en la carrera”

Tabla A.18. Entrevista 18.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 2 y 3, mencionando que representa la familia.

Grupo 2: imágenes 4 y 5, mencionando que representa las cosas materiales que tendrá.

Grupo 3: imágenes 1 y 6, mencionando que representa el presente y futuro, aplicando los conocimientos obtenidos en la universidad.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Me sentiría mal frente a mi familia, ya que ellos están pagando por mis estudios, ademas se haría muy difícil de cumplir mis otros sueños ya que no ganaría el dinero que ganaría estudiando esta carrera.”

ENTREVISTA 19

Nombre: Constanza

Edad: 20

Carrera: 1er Año Ingeniería Civil Industrial

Comuna: Maipú

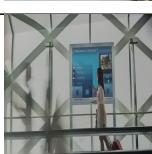
Número	Imagen	Descripción	Asociaciones
1		“Aprendizaje”	“La universidad me da herramientas que me servirán para el futuro”
2		“El Presente”	“Una red de aprendizaje que me llevará al futuro”
3		“Casa Propia”	“Tener mi casa propia y poder independizarme”
4		“Familia”	“Poder formar mi familia alegre y saludable”
5		“Viajes”	“Me gustaría viajar por muchas partes del mundo para poder aprender y disfrutar”
6		“El futuro”	“Tener mi propia oficina, trabajar con más gente y enseñar lo que sé”

Tabla A.19. Entrevista 19.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1 y 2, mencionando que representa “la llave al futuro”, ya que son las herramientas que la universidad le enseñan para poder ser ingeniero.

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Grupo 2: imágenes 5 y 6 mencionando que representa su futuro, relacionando su trabajo y los viajes que hará.

Grupo 3: imágenes 3 y 4 mencionando que representa su familia, constituir una buena familia y un hogar en donde vivan todos juntos.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Me afligiría bastante por las motivaciones ya que quiero ser una buena ingeniera, aun así podría cumplir lo de la familia, pero lo otro sería muy difícil.”

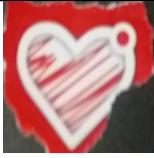
ENTREVISTA 20

Nombre: Renata

Edad: 19

Carrera: 1er Año Ingeniería Civil Industrial

Comuna: Providencia

Número	Imagen	Descripción	Asociaciones
1		“Auto”	“Poder comprar mi auto soñando con el dinero que ganaré”
2		“Corazón”	“Me haría sentir plena y bien conmigo misma”
3		“Teléfono y Reloj”	“Terminar un ciclo y armar la vida”
4		“Maleta”	“Refleja que terminando la carrera podré viajar mucho”

Número	Imagen	Descripción	Asociaciones
5		“Casa”	“Poder tener una casa amplia”
6		“Colores”	“Como vería mi alma”
7		“Buda”	“Representa la felicidad”

Tabla A.20. Entrevista 20.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 4 y 5, mencionando que representa su lado externo, con lo económico y superficial, que no es tan primordial.

Grupo 2: imágenes 2, 3, 6 y 7, mencionando que representa su lado interno, lo que le ayudará a crear su personalidad, su vida y felicidad.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“No afectaría en nada, simplemente estudiaría otra carrera”

ENTREVISTA 21

Nombre: Ramón

Edad: 19

Carrera: 1er Año Ingeniería Civil Industrial

Comuna: Santiago

Número	Imagen	Descripción	Asociaciones
1		“Familia”	“Poder formar una familia feliz”
2		“Padres”	“Refleja a mis padres, al terminar mi carrera ellos estarían orgullosos de mí”
3		“Seguridad”	“Salir de la carrera es un gran paso y te da libertad y seguridad en trabajar en lo que quieras”
4		“Viajes”	“Poder viajar con la familia y conocer más lugares”
5		“Auto”	“Tener un buen trabajo, ser éxito y tener lujo”
6		“Hijos”	“Tener hijos sanos y que no les falte nada”

Tabla A.21. Entrevista 21.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 1, 2 y 6, mencionando que representa su entorno familiar, que estarían orgullos de terminar un ciclo.

Grupo 2: imágenes 3, 4 y 5, mencionando que representa los lujos que podría obtener si terminara la carrera.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Podría cumplir el tema de la familia, el segundo grupo dependería de como me fuera a futuro”

ENTREVISTA 22

Nombre: Ian

Edad: 19

Carrera: 1er Año Ingeniería Civil en Obras Civiles

Comuna: Providencia

Número	Imagen	Descripción	Asociaciones
1		“Cosas”	“Las cosas que uno puede optar”
2		“Herramientas”	“Representa las habilidades obtenidas en la carrera”
3		“Naturaleza”	“Me gustaría tener un desarrollo sustentable, poder ayudar a la naturaleza con mi carrera”
4		“Viajar”	“Poder viajar junto a mis cercanos”
5		“Disfrutar”	“Dejar de estar preocupado y juntarse con su familia o amigos”
6		“Tarjeta Bancaria”	“Poder tener dinero para poder pagar con diferentes medios de pago”

ENTREVISTAS CON METODOLOGÍA DE LA METÁFORA

Número	Imagen	Descripción	Asociaciones
7		“Seguridad”	“Poder tener dinero para estar seguro”

Tabla A.22. Entrevista 22.

Los grupos de imágenes que escogió el entrevistado fueron:

Grupo 1: imágenes 4 y 5, mencionando que representa su lado social.

Grupo 2: imágenes 1, 2 y 3, mencionando que representa su lado intelectual y lo que podría hacer al terminar la carrera.

Grupo 3: imágenes 6 y 7 mencionando que representa la estabilidad que le daría el terminar la carrera.

¿Cómo se ven afectadas tus motivaciones si desertas?.

“Se verían muy afectado, ya que tendría que trabajar y no podría cumplir varias metas”

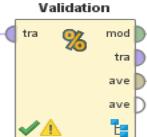
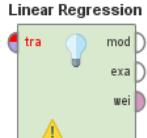
Anexo B

RapidMiner

RapidMiner es una herramienta de código abierto, usado para la minería de datos. Utiliza un ambiente gráfico que permite combinar distintos operadores y generar procesos de tratamiento y/o análisis de datos.

Operadores utilizados en la creación del proceso para la predicción de la deserción.

Operador	Descripción
Retrieve 	Este operador se utiliza para acceder a los repositorios. Transforma la mayor cantidad de datos del archivo de entrada en metadatos para un mejor procesamiento.
Nominal to Numerical 	Este operador cambia el tipo de atributos no numéricos seleccionados a un tipo numérico. También asigna todos los valores de estos atributos a valores numéricos.
Normalize 	Este operador normaliza los valores de atributos de los atributos seleccionados.
Replace Missing Values (2) 	Este operador reemplaza los valores faltantes en los ejemplos de atributos seleccionados por un reemplazo especificado.
Multiply 	Este operador copia su objeto de entrada a todos los puertos de salida conectados. No modifica el objeto de entrada.
Filter Examples (3) 	Este operador selecciona los datos de un conjunto de datos que se deben conservar y qué datos deben eliminarse. Se mantienen los datos que satisfacen la condición dada, se eliminan los datos restantes.

Operador	Descripción
Apply Model (2) 	Este operador aplica un modelo ya aprendido o entrenado en un conjunto de datos.
Write Excel 	Este operador escribe el conjunto de datos en un archivo de hoja de cálculo de Excel.
Validation 	<p>Este operador realiza una validación simple, es decir, divide aleatoriamente los datos en un conjunto de entrenamiento y un conjunto de pruebas y evalúa el modelo. Este operador realiza una validación de división para estimar el rendimiento de un operador de aprendizaje (normalmente en conjuntos de datos no vistos).</p> <p>Se utiliza principalmente para estimar la precisión con la que un modelo (aprendido por un operador de aprendizaje en particular) se llevará a cabo en la práctica.</p>
Linear Regression 	Este operador calcula un modelo de regresión lineal a partir de la entrada del conjunto de datos.
Decision Tree 	Genera un árbol de decisión para la clasificación de datos nominales y numéricos.

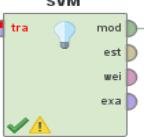
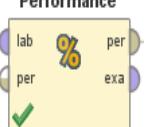
Operador	Descripción
	Este operador aprende un modelo por medio de una red neuronal feed-forward entrenada por un algoritmo de retropropagación (perceptrón multicapa).
	Este operador es un aprendiz de SVM (Support Vector Machine). Se basa en la implementación Java interna de mySVM por Stefan Rueping.
	Este operador se utiliza para la evaluación del desempeño. Proporciona una lista de valores de criterios de rendimiento. Estos criterios de rendimiento se determinan automáticamente para adaptarse al tipo de tarea de aprendizaje.

Tabla B.1. Operadores RapidMiner