

UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

Departamento de Ingeniería Comercial

MBA

**Modelo BI para Adaptar la Oferta de Programas de Apoyo Estudiantil a Perfiles de  
Riesgo de Deserción en la Universidad Masiva**

Guillermo Badillo Astudillo

MBA. Magister en Gestión Empresarial

Octubre de 2016



UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA

Departamento de Ingeniería Comercial

MBA

**MODELO BI PARA ADAPTAR LA OFERTA DE PROGRAMAS DE APOYO  
ESTUDIANTEL A PERFILES DE RIESGO DE DESERCIÓN EN LA  
UNIVERSIDAD MASIVA**

Tesis de Grado presentada por

**Guillermo Badillo Astudillo**

Como requisito para optar al grado de  
MBA. Magister en Gestión Empresarial

Director de Tesis: Hugo Osorio Zelada

**Octubre de 2016**

**TITULO DE TESIS:** Modelo BI para adaptar la oferta de programas de apoyo estudiantil a perfiles de riesgo de deserción en la Universidad Masiva

**AUTOR:**

**GUILLERMO ENRIQUE BADILLO ASTUDILLO**

**TRABAJO DE TESIS**, presentado en cumplimiento parcial de los requisitos para el Grado de MBA Magister en Gestión Empresarial de la Universidad Técnica Federico Santa María.

Observaciones: \_\_\_\_\_  
\_\_\_\_\_

|            |       |
|------------|-------|
| Profesor 1 | ..... |
| Profesor 2 | ..... |
| Profesor 3 | ..... |

Santiago, noviembre de 2016

Todo el contenido, análisis, conclusiones  
y opiniones vertidas en este estudio son  
de mi exclusiva responsabilidad

Nombre: Guillermo Enrique Badillo Astudillo

Firma.....

Fecha.....



## **RESUMEN EJECUTIVO**

La revolución tecnológica trajo consigo posibilidades y demandas, al igual que lo hizo la imprenta y la revolución industrial en su tiempo. Mayor complejidad en el uso de tecnologías y mayor flujo de información exigió una masa ciudadana y trabajadora más educada. Se desarrolló el proceso de masificación de la Universidad, una institución que no estaba preparada para recibir a un nuevo estudiante sin las cualidades intelectuales ni capital simbólico del estudiante universitario tradicional. La cultura Universitaria aún no se ajusta a la inclusión de este estudiante. Para hacerlo ha desarrollado diversas estrategias de retención, algunas con énfasis económico, académico, psicológico, social y/o médico.

Desde el ámbito de la gestión y la administración en la educación superior, altas tasas de abandono estudiantil son de gran impacto en el bienestar financiero de la Universidades, puesto que los retornos de los programas y proyectos de retención son muy bajos, y en algunos casos negativos. Además, las tasas de abandono o deserción, impactan directamente en los indicadores de calidad de la Universidades. Por lo anterior el modelo impactará directamente en la rentabilidad de los proyectos y programas asociados a la deserción o retención en la Instituciones de Educación Superior, y a través de esta rentabilidad exigida impacte en los indicadores de deserción y su contracara la retención.

La literatura científica da cuenta de las variables asociadas a la deserción, lo que permitiría construir perfiles de riesgo con necesidades distintas, que requieren estrategias de retención diferenciadas. Por otra parte, el sistema universitario chileno es diverso, por cuanto las poblaciones de cada institución tienen distintos elementos culturales dominantes, de inclusión social y riesgo académico. La inteligencia de negocios facilita la generación de sistemas de focalización de la atención mediante la estrategia de clasificación de los estudiantes en riesgo y la jerarquización de necesidades en el grupo que se integra cada año en una población particular. A la vez, la nueva información permitiría retroalimentar el sistema de información de cada institución. Este modelo es el que se desarrolla en la presente tesina.

Las potenciales trabajos futuros de este estudio incluyen: (i) extender los métodos y conjuntos de modelado predictivo con otras técnicas más recientes, (ii) mejorar las fuentes de información incluyendo los datos de estudios institucionales basados en encuestas intencionalmente elaborados y administrados cuidadosamente para fines de retención, y (iii) el despliegue del sistema informático como una ayuda a la toma de decisión para que los administradores evalúen su idoneidad y usabilidad en el mundo real.

## **ABSTRACT**

The technological revolution brought with it possibilities and demands, just as did the printing press and the industrial revolution in its time. Greater complexity in the use of technologies and greater flow of information demanded a more educated citizen and worker mass. The process of massification of the University was developed, an institution that was not prepared to receive a new student without the intellectual qualities nor symbolic capital of the traditional university student. The university culture still does not fit the inclusion of this student. In order to do so, it has developed various retention strategies, some with economic, academic, psychological, social and / or medical emphasis.

From the field of management and administration in higher education, high rates of student drop-out are of great impact on the financial well-being of universities, since the returns from retention programs and projects are very low, and in some cases Negative results. In addition, dropout or dropout rates have a direct impact on university quality indicators. Therefore, the model will directly impact on the profitability of the projects and programs associated with the dropout or withholding in the Higher Education Institutions, and through this required profitability impact on the indicators of desertion and its counterpart the retention.

The scientific literature accounts for the variables associated with desertion, which would allow the construction of risk profiles with different needs, which require different retention strategies. On the other hand, the Chilean university system is diverse, since the populations of each institution have different dominant cultural elements, of social inclusion and academic risk. Business intelligence facilitates the generation of attention targeting systems through the strategy of classifying at-risk students and the hierarchy of needs in the group that integrates each year into a particular population. At the same time, the new information would allow feedback from the information system of each institution. This model is developed in the present thesis.

Potential future work in this study includes: (i) Extending predictive modeling methods and assemblies with more recent techniques, (ii) improving sources of information including data from institutional surveys based on surveys that are intentionally designed and managed carefully And (iii) the deployment of the computer system as an aid to decision making for managers to assess their suitability and usability in the real world.

## CONTENIDO

|          |                                                                                                                 |           |
|----------|-----------------------------------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>INTRODUCCIÓN .....</b>                                                                                       | <b>5</b>  |
| <b>2</b> | <b>DEFINICIÓN Y JUSTIFICACIÓN DEL TEMA .....</b>                                                                | <b>7</b>  |
| <b>3</b> | <b>OBJETIVOS.....</b>                                                                                           | <b>9</b>  |
| 3.1      | OBJETIVO GENERAL: .....                                                                                         | 9         |
| 3.2      | OBJETIVOS ESPECÍFICOS: .....                                                                                    | 10        |
| <b>4</b> | <b>ALCANCE.....</b>                                                                                             | <b>10</b> |
| <b>5</b> | <b>ESTADO DEL ARTE.....</b>                                                                                     | <b>11</b> |
| 5.1      | ANTECEDENTES DEL NUEVO ESTUDIANTE EN LA UNIVERSIDAD MASIVA .....                                                | 11        |
| 5.1.1    | <i>La Universidad Masiva en Chile y su Nuevo Estudiante. ....</i>                                               | <i>11</i> |
| 5.1.2    | <i>Adaptación a la Universidad del Nuevo Estudiante: Estudio Cualitativo Nacional. ....</i>                     | <i>16</i> |
| 5.2      | MODELOS EMPÍRICOS DE ANÁLISIS DE LA DESERCIÓN EN EDUCACIÓN SUPERIOR .....                                       | 18        |
| 5.2.1    | <i>Factores Asociados a la Deserción en Estudios Chilenos.....</i>                                              | <i>18</i> |
| 5.2.2    | <i>Factores Asociados a la Deserción en una Propuesta Iberoamericana. ....</i>                                  | <i>21</i> |
| 5.2.3    | <i>Factores Asociados a la Deserción en Población Vulnerable. ....</i>                                          | <i>22</i> |
| 5.3      | DESERCIÓN UNIVERSITARIA EN CHILE: .....                                                                         | 25        |
| 5.3.1    | <i>Análisis de Casos Nacionales.....</i>                                                                        | <i>25</i> |
| 5.3.2    | <i>Estudios Empíricos con Población Nacional.....</i>                                                           | <i>29</i> |
| <b>6</b> | <b>MARCO TEÓRICO.....</b>                                                                                       | <b>35</b> |
| 6.1      | INTELIGENCIA DE NEGOCIOS.....                                                                                   | 35        |
| 6.1.1    | <i>Antecedentes del uso de Inteligencia de Negocios en Educación Superior.....</i>                              | <i>35</i> |
| 6.1.2    | <i>Herramientas para la Inteligencia de Negocios.....</i>                                                       | <i>36</i> |
| 6.1.3    | <i>Proceso de Descubrimiento de Conocimiento en Base de Datos, (Knowledge Discovery in Databases, KDD).....</i> | <i>38</i> |
| 6.1.4    | <i>El proceso KDD.....</i>                                                                                      | <i>40</i> |
| 6.1.5    | <i>Etapas de Minería de Datos del proceso KDD .....</i>                                                         | <i>40</i> |
| 6.1.6    | <i>Métodos de Minería de Datos.....</i>                                                                         | <i>40</i> |
| 6.1.7    | <i>Componentes de los algoritmos de minería de datos.....</i>                                                   | <i>41</i> |
| 6.1.8    | <i>Métodos de Minería de Datos más populares.....</i>                                                           | <i>42</i> |
| <b>7</b> | <b>METODOLOGÍA .....</b>                                                                                        | <b>44</b> |
| 7.1      | MÉTODO:.....                                                                                                    | 44        |
| 7.2      | DISEÑO: .....                                                                                                   | 44        |
| 7.3      | POBLACIÓN:.....                                                                                                 | 44        |
| 7.4      | MUESTREO:.....                                                                                                  | 44        |
| 7.5      | BASES DE DATOS: .....                                                                                           | 45        |
| 7.6      | PROCEDIMIENTO:.....                                                                                             | 45        |
| <b>8</b> | <b>DESARROLLO DE LA PROPUESTA .....</b>                                                                         | <b>47</b> |

|           |                                                                                                                                                                      |           |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| 8.1       | TEORÍA DE SISTEMAS APLICADO AL MODELO .....                                                                                                                          | 47        |
| 8.2       | ARQUITECTURA DEL SUBSISTEMA BI APLICADO AL MODELO .....                                                                                                              | 48        |
| 8.3       | DESARROLLO DEL MODELO DE CLASIFICACIÓN DE PERFILES DE RIESGOS DE LOS ESTUDIANTES .....                                                                               | 49        |
| 8.3.1     | <i>Introducción .....</i>                                                                                                                                            | 49        |
| 8.3.2     | <i>Pasos en la construcción del Modelo.....</i>                                                                                                                      | 51        |
| 8.3.3     | <i>Desarrollo Paso 1: Construcción del problema de negocios al que se dará solución con el Modelo. 53</i>                                                            |           |
| 8.3.4     | <i>Definición Conceptual de Perfiles de Riesgo de Deserción .....</i>                                                                                                | 55        |
| 8.3.5     | <i>Desarrollo Paso 2: Definición de las fuentes de datos internas y externas. Comprensión y conocimiento de los datos a usar en el modelo de clasificación. ....</i> | 56        |
| 8.3.6     | <i>Desarrollo Paso 3: Pre Procesamiento y Transformación de los Datos (Data Preprocessing and Data Transformation) .....</i>                                         | 60        |
| 8.3.7     | <i>Desarrollo Paso 4: Minería de Datos (Data Mining) .....</i>                                                                                                       | 62        |
| 8.3.8     | <i>Desarrollo paso 5: Evaluación e interpretación:.....</i>                                                                                                          | 70        |
| 8.4       | ALINEAMIENTO ESTRATÉGICO DE LA CARTERA DE PROYECTOS CON LOS PERFILES DE RIESGO .....                                                                                 | 73        |
| 8.5       | PROGRAMAS Y PROYECTOS PARA EVITAR EL ABANDONO ESTUDIANTEL .....                                                                                                      | 74        |
| 8.6       | INDICADORES O KPI ESTADÍSTICOS ASOCIADOS A LA DESERCIÓN .....                                                                                                        | 75        |
| <b>9</b>  | <b>CONCLUSIONES .....</b>                                                                                                                                            | <b>76</b> |
| <b>10</b> | <b>BIBLIOGRAFÍA.....</b>                                                                                                                                             | <b>78</b> |
| <b>11</b> | <b>ANEXOS.....</b>                                                                                                                                                   | <b>86</b> |

## INDICE DE TABLAS

|                                                                                                                 |     |
|-----------------------------------------------------------------------------------------------------------------|-----|
| Tabla 1: Ejemplo de variables de muestra .....                                                                  | 64  |
| Tabla 2: Tabla de Operacionalización de Variables para el Perfil Riesgo por Integración Social.....             | 98  |
| Tabla 3: Tabla de Operacionalización de Variables para el Perfil Riesgo por Trayectoria Académica.....          | 100 |
| Tabla 4: Tabla de Operacionalización de Variables para el Perfil Riesgo por Vulnerabilidad Socio-cultural ..... | 102 |

## INDICE DE GRÁFICOS

|                                                                                          |    |
|------------------------------------------------------------------------------------------|----|
| Grafica 1: Progresión de la Matrícula Universitaria por Tipo de Institución. ....        | 11 |
| Grafica 2: Tasas de Cobertura en Educación Superior (Libertad y Desarrollo, 2015) .....  | 13 |
| Grafica 3: Cobertura por Décil (Centro de Estudios Mineduc, 2012a).....                  | 13 |
| Grafica 4: Progresión por Quintil de la Cobertura Neta en Educación Superior .....       | 14 |
| Grafica 5: Distribución de la Matrícula del Quintil I y II por Tipo de Institución ..... | 15 |

## INDICE DE FIGURAS

|                                                                                          |    |
|------------------------------------------------------------------------------------------|----|
| Figura 1: Adaptación del Modelo de Tinto a Ingeniería Civil UCSC.....                    | 28 |
| Figura 2: Vista general de los pasos que componen el proceso KDD. ....                   | 39 |
| Figura 3: Sistema Referencial del Modelo Predictivo .....                                | 48 |
| Figura 4: Arquitectura típica de un sistema de inteligencia de negocios .....            | 48 |
| Figura 5: Componentes Modelo BI de Clasificación.....                                    | 50 |
| Figura 6: Formas de Pre-procesamiento de Datos .....                                     | 52 |
| Figura 7: Marco referencial de diseño y análisis de un Sistema de Información .....      | 54 |
| Figura 8: Fuentes de Datos Externas/Internas.....                                        | 57 |
| Figura 9: Árbol de decisión para perfil de riesgo académico.....                         | 65 |
| Figura 10: Arquitectura de Red Neuronal usada en este estudio .....                      | 69 |
| Figura 11: Ejemplo de un Árbol de decisión completo para perfil de riesgo académico .... | 72 |
| Figura 12: Alineamiento estratégico del portafolio .....                                 | 73 |
| Figura 13: Marco de gobierno para proyectos estratégicos .....                           | 74 |

## 1 INTRODUCCIÓN

La revolución tecnológica demandó a los ciudadanos mayor conocimiento para alcanzar el nivel instrumental de funcionamiento social, por lo tanto, la formación universitaria –aun cuando no se ejerza la profesión– otorga mayor probabilidad de encontrar empleo (Alegre & Benito 2010; Bernardi & Requena 2010; Mena et al. 2010; Poy 2010). En países como España, terminar una titulación entrega mayor capacidad de empleabilidad a los estudiantes, aun cuando durante la formación exista desilusión por la remuneración posterior a terminar y la escasa oferta laboral en la carrera (Alegre & Benito 2010; Bernardi & Requena 2010; Mena et al. 2010). Este elemento es central en la conformación de la Universidad Masiva.

La Universidad masiva, disminuyó las barreras de ingreso y abrió el ingreso a personas sin la experiencia y/o capacidad para estar en el modelo tradicional (Tinto 1989; Tinto 1993). Con ello, la deserción estudiantil es un problema que aborda la mayoría de las instituciones de educación superior en el mundo. Para apreciar el fenómeno en Chile se debe considerar que la OECD (2008) daba cuenta de una media de deserción en educación terciaria del 31%, con una máxima de 46% en Nueva Zelanda y una mínima de 10% en Japón. En Chile, los estudios tenían por consenso de que al menos la mitad de los alumnos que ingresaban a la educación superior abandonaban sus estudios antes de obtener el título (Acuña 2012). Aun cuando no se han realizado estudios de seguimiento que aborden al total de la población, la situación pareciera tener pocos cambios, si se considera que al 2000 González y Uribe (2002) describían una situación similar.

La evidencia muestra que el sistema de financiamiento y la matriz productiva nacional cobra importancia real en el análisis del fenómeno de la deserción en educación superior. Por ejemplo, en sociedades como la Uruguay, que poseen acceso universal junto a trabajo en el sector agrícola, las tasas de deserción son altas desde la secundaria y se concentran en el sector público, sin embargo, los ciudadanos se emplean (Fernández 2010). En Chile, la formación universitaria es importante para aumentar la oportunidad para acceder a empleo –al igual que en el caso español– pues la matriz productiva se centra en servicios y el Estado subsidia la demanda, sin distinguir oferentes, lo que implica un gasto importante (Cepe et al. 2013), conduciendo al análisis a la calidad de las Universidades y a la productividad que se logra con esos recursos.

Considerando lo antes descrito, el fenómeno de la deserción se configura como problema en Chile al considerar que: (a) la universidad ha modificado las barreras de acceso

transformándose en un sistema masivo en el que existe un nuevo perfil de estudiante, (c) en países con sistemas de educación superior masiva ha disminuido la empleabilidad y sueldos de las profesiones, (b) en Chile el financiamiento recae fundamentalmente en las familias lo que implica endeudamiento, (d) el estado asumió una función subsidiaria, con un alto gasto en educación y (e) las instituciones de educación superior no están preparadas para brindar educación de calidad a este nuevo estudiante. Estos aspectos son centrales en el análisis de la situación chilena.

El uso del conocimiento para producir valor –lo que demanda aprender- y la transformación del acceso, ha implicado la apertura del mercado a privados para aumentar la cobertura educativa, cuestión exitosa en Chile, introduciendo problemas como el endeudamiento a las familias y el gasto del Estado (Oliva 2008; Bellei 2013). Gracias al aumento en la cobertura, la (OECD 2009), como es la tendencia mundial. Según el Centro de Estudios del MINEDUC (2012), el crecimiento del sistema de educación superior en los últimos treinta años implicó pasar de una matrícula de 165 mil estudiantes de pregrado a principios de los años ochenta, a más de un millón el año 2012. En Chile se han disminuido de barreras de entrada, aun cuando las barreras de ingreso disminuyen la deserción (Custodio 2010; Fernández 2010). El actual estudiante ingresa con una preparación escolar más débil (Centro Interuniversitario de Desarrollo CINDA 2006; Elias 2008), configurando un nuevo perfil y nuevas necesidades para el sistema universitario (Centro Interuniversitario de Desarrollo CINDA 2006; Elias 2008). En este contexto:

- a. Las universidades que no han analizado exhaustivamente los nuevos perfiles de estudiantes y sus nuevas necesidades (preferencias profesionales, nivel de entrada en contenidos y habilidades), ni han destinado los recursos necesarios para atender a todos estos cambios (Elias 2008). Tampoco han adaptado sus prácticas docentes, lo que impacta en la “repitencia” y deserción (González & Uribe 2002).
- b. Para la gestión de los recursos del Estado, mayor cobertura no implicó más integración social: la cobertura en los niveles socioeconómicos más vulnerables siguió siendo baja (Acuña 2012) y se consolidó carácter subsidiario del estado en educación superior al privilegiar el aumento de subsidios a la demanda en instituciones públicas o privadas, aumentando tendencialmente los aportes absolutos a instituciones privadas (Cepe et al. 2013).
- c. El gasto privado en el financiamiento (principalmente los pagos de los estudiantes y el endeudamiento de las familias) fue en promedio cinco veces superior al gasto público durante el período comprendido entre 1990 y 2011 (Cepe et al. 2013). La OECD (2008)

reportó que en Chile la educación superior se financiara principalmente por los hogares, que absorben un 80% del gasto. Con ello el abandono afecta a las familias y nuevos profesionales.

Las cifras anteriores muestran que en Chile la deserción preocupa al Estado y los distintos actores del sistema, ya que se le asocian importantes pérdidas de eficiencia. Se pierde una parte de los recursos que el Estado, los estudiantes y sus familias han invertido en la formación. Además, las instituciones se ven afectadas, pues dejan de recibir los recursos asociados a estos estudiantes y deben adaptar su funcionamiento en cursos superiores a un menor número de alumnos. Finalmente, los jóvenes que desertan disminuyen su empleabilidad y ven truncados los sueños de graduarse. No es fácil estimar todos estos costos, pero para una noción de las pérdidas que genera este fenómeno González, Uribe y González (2005) estimaron un costo directo de \$ 47 mil millones de pesos anuales.

Considerando el costo objetivo y subjetivo para los diversos actores del sistema de educación universitaria y del mercado laboral, se configura como pregunta de investigación: ¿Es posible focalizar y adaptar la oferta de programas de apoyo estudiantil al perfil de riesgo de deserción en nuevos estudiantes de la universidad masiva?

## **2 DEFINICIÓN Y JUSTIFICACIÓN DEL TEMA**

La educación superior es considerada una inversión en capital humano con altas tasas de rentabilidad privada y social. El retorno a la inversión compensa costos de oportunidad, costos explícitos de la formación (matrículas, manutención, materiales, entre otros) y factores de riesgo. Este retorno incentiva a los individuos a invertir en educación superior, asumiendo costos privados, y contribuye a respaldar políticas públicas que incentiven la continuación de estudios superiores hasta lograr un nivel socialmente óptimo, a través de subsidios a la oferta (instituciones), o a la demanda (estudiantes).

El incremento de las tasas de cobertura en la educación superior conlleva un mayor acceso a ésta, especialmente de estudiantes pertenecientes a estratos de ingresos medio-bajo y bajos, los cuales alcanzan tasas de cobertura equivalentes a un tercio de la observada en los estratos de mayores ingresos, y muestran el más alto crecimiento (Libertad y Desarrollo 2015; Centro de Estudios Mineduc 2012a; Centro de Estudios Mineduc 2012b; Canales & De los Ríos 2009). Esto explica que, en la actualidad, un alto porcentaje de estudiantes sean primera generación en la educación superior en sus grupos familiares.

El mayor acceso a la educación superior, medido en términos absolutos o en tasas de cobertura, no garantiza, sin embargo, un éxito de las políticas públicas, ya que no todos los estudiantes que ingresan a ésta, efectivamente concluyen sus programas de estudios. No obstante, esta decisión conlleva costos privados y sociales, que se expresan en endeudamiento para una inversión inconclusa con beneficios netos relativamente bajos e incluso negativos en algunos casos (Cepe et al. 2013). Algunas universidades asumen pérdidas cuando han otorgado o avalado créditos que no necesariamente recuperan.

Un costo adicional, que puede llegar a ser muy importante, es el impacto en los ingresos que la universidad deja de percibir en el futuro al desertar el estudiante en los primeros años de su ingreso a la educación superior. Por otra parte, el Estado no recupera el retorno de su inversión por becas otorgadas o crédito subsidiado. En consecuencia, las tasas de deserción generan un problema de ineficiencia (aparte del impacto en la equidad) en el sistema universitario (Centro de Microdatos 2008).

Los estudios evidencian los costos que ameritan intervención. En lo personal implica fracaso, que afecta emocionalmente por la disonancia las aspiraciones e incide en la trayectoria ocupacional (González et al. 2005), hay pérdida de oportunidades laborales, disminuyen las posibilidades de conseguir empleos satisfactorios y los salarios son más bajos. El disgusto, la frustración y la sensación de fracaso tiene efectos en su salud física y mental (Centro Interuniversitario de Desarrollo CINDA 2006).

En lo institucional implica disminución del rendimiento académico de la universidad, menores ingresos, problemas de eficiencia y calidad por el incumplimiento de los planes establecidos, afectando la imagen de la universidad en el medio académico y la acreditación (González et al. 2005). En lo financiero hay menores ingresos por matrícula y costos adicionales (Centro Interuniversitario de Desarrollo CINDA 2006).

En lo social la deserción contribuye a generar inequidad y desequilibrios sociales (González et al. 2005), se retroalimenta el círculo de la pobreza y se gesta una “capa social” de frustrados profesionales, siendo posible la disminución del aporte intelectual y el aumento potencial del subempleo. También se incrementa el costo para el país por la sub-optimización de los recursos públicos (Centro Interuniversitario de Desarrollo CINDA 2006).

Dado lo anterior, y desde el ámbito de la gestión en instituciones de educación superior, se propone desarrollar el análisis basado en la evidencia de variables de ingreso asociadas al éxito y fracaso de los estudiantes en la adaptación en el sistema de educación universitaria, con el fin de diseñar un sistema que permita la clasificación de los estudiantes

en perfiles de riesgo de deserción mediante la aplicación de herramientas de inteligencia de negocios (BI). Esta clasificación permitiría a las instituciones de educación superior gestionar el riesgo para orientar las políticas y los recursos asignados a los programas e iniciativas de disminución de las tasas de deserción universitaria.

Este tipo de análisis es particularmente relevante si se considera que la interacción de variables deja a cada institución en una situación propia debido a la relación costo-ingreso asociado a: barreras de entrada, programas de formación y apoyo, forma de financiamiento y puntaje PSU. Estas variables, sensibles a la vida del proyecto de negocio son también variables con efecto en la deserción de la población hasta ahora estudiada: puntaje PSU del estudiante, Puntaje PSU y SIMCE de la escuela de procedencia, nivel de estudio de los padres, becas y créditos para estudiar, situación laboral, responsabilidad con el financiamiento familiar. Estudiar en instituciones que enseñan a estudiantes poco convencionales y poseedores del nuevo perfil de ingreso, coloca frente a poblaciones homogéneas en estas variables de riesgo y abre al hallazgo de variables fortalecedoras, es decir, abre espacio al diseño de modelos de gestión del riesgo como factor de eficacia y eficiencia, un nuevo abordaje en el debate de la calidad de las instituciones.

Por ello con esta propuesta se construye desde la evidencia perfiles de estudiante que permanecen y abandonan estudios superiores para focalizar el apoyo hacia los perfiles que la literatura ha logrado diferenciar. Para ello debe: (a) Identificar variables de riesgo para la deserción. (b) Construir perfiles que permitan agrupar estudiantes por tipo de riesgo primario. (c) Identificar variables asociadas a tipos de apoyo mediante programas de educación superior. (d) Asociar variables protectoras del estudiante que facilitan su adaptación al programa de apoyo requerido.

Estas variables permiten el análisis de los sistemas de integración a la educación superior –tradicionalmente pensados desde el perfil clásico de estudiante– a partir de las características de sus usuarios, que pueden corresponder de modo homogéneo al nuevo perfil de estudiante de la universidad masiva.

### **3 OBJETIVOS**

#### **3.1 Objetivo general:**

Diseñar un modelo BI que adapte estratégicamente la oferta de programas de apoyo a perfiles de estudiantes en riesgo de deserción de una Universidad masiva

### 3.2 Objetivos específicos:

- 1.2.2.1. Definir perfiles de riesgo a partir de las variables asociadas al riesgo de deserción.
- 1.2.2.2. Clasificar a los estudiantes mediante el apoyo de técnicas de Inteligencia de Negocios (BI) por perfiles de deserción.
- 1.2.2.3. Diseñar el Modelo BI para automatizar el proceso de clasificación.

## 4 ALCANCE

El alcance de la presente investigación está circunscrito al estudio de la deserción o abandono de los estudiantes de primer año de universidades chilenas clasificadas como masivas. Está orientado a proponer un modelo que permita adaptar y orientar estratégicamente los programas de apoyo estudiantil diseñados para disminuir la deserción o abandono de estudiantes en riesgo, según tres perfiles que se desprenden de la investigación sobre deserción que se ha aproximado al estudio del nuevo estudiante.

Se propone analizar la población de primer año de universidades específicas pues en la heterogeneidad del sistema universitario chileno, cada universidad se orienta a un grupo destinatario homogéneo, lo cual permite establecer rangos de clasificación. Junto a lo anterior, el análisis de la literatura muestra que: (a) la mayor parte de la deserción universitaria se produce en los dos o tres primeros años de carrera -según el área de conocimiento-, (b) las variables valoradas que permiten caracterizar y agrupar dan cuenta del tránsito entre la educación secundaria y terciaria (escolaridad de los padres, puntaje PSU, puntaje SIMCE, NEM, porcentaje de beca y crédito, etc.), (c) los indicadores al segundo año difieren y se vinculan a promedios, asignaturas aprobadas, asignaturas reprobadas, etc. (d) la diferencia en los indicadores puede implicar que el primer año es pertinente el diseño de programas de apoyo genéricos, mientras en segundo año puede ser pertinente el diseño de estrategias por sector de conocimiento o programa de carrera, situación fuera del análisis de este estudio.

El alcance de esta investigación es correlacional, pues pretende establecer relaciones entre las variables asociadas y predictoras de deserción con los sujetos estudiantes que ingresan a los programas de formación profesional, para clasificarlos en tres perfiles de riesgo según énfasis en causales de integración social, académica o ambas. En un segundo momento y dentro del clúster se realizará una segunda correlación con variables protectoras para entregar información que permita priorizar atención y adaptar la oferta junto a la estrategia al interior de los programas de apoyo.

## 5 ESTADO DEL ARTE

### 5.1 Antecedentes del Nuevo Estudiante en la Universidad Masiva

#### 5.1.1 *La Universidad Masiva en Chile y su Nuevo Estudiante.*

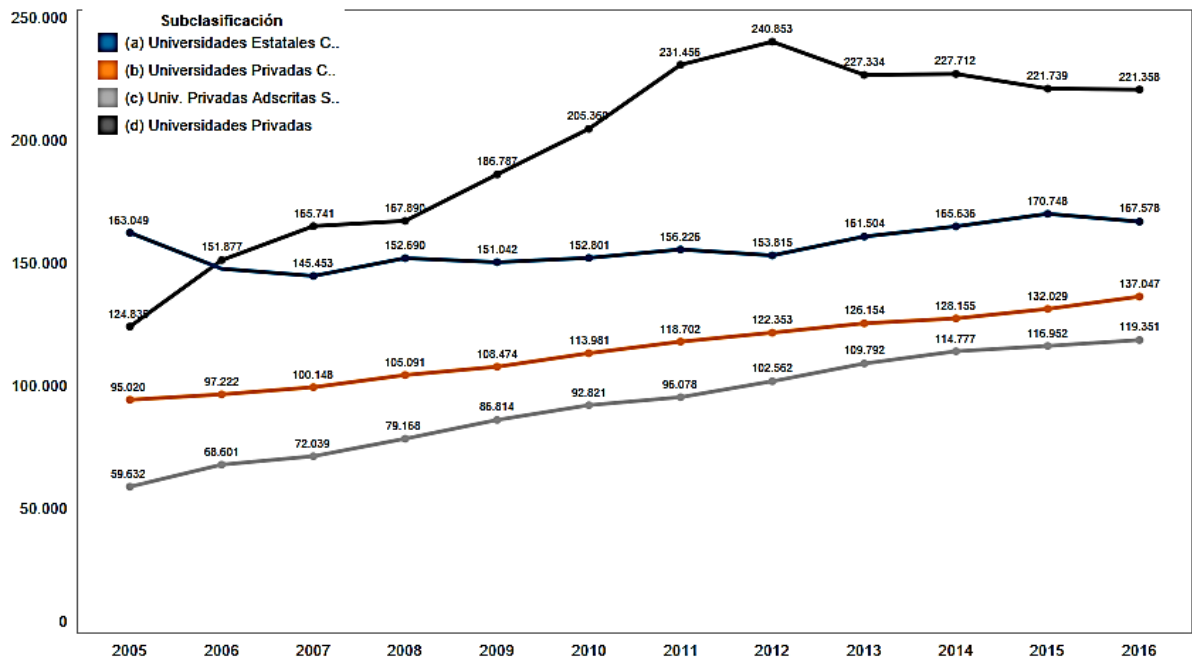
El sistema de educación chileno es masivo y diversificado, financiado mayoritariamente con recursos privados (Canales & De los Ríos 2009; Cepe et al. 2013; OECD 2009). Para analizarlo, un primer elemento a considerar es el aumento sostenido la matrícula.

En 1990 el sistema terciario contaba con cerca de 200.000 estudiantes de pregrado (Uribe 2004), al 2007 la matrícula de pregrado había alcanzado a 752.182 estudiantes (Canales & De los Ríos 2009). En 2011 había un total de 1.032.299 personas en la educación superior, de ellas 750.145 tenía entre 18 y 24 años de edad. Del total cerca de un 75% se encontraba matriculado en programas conducentes a títulos profesionales/ licenciaturas. Un 39% se concentró en universidades no pertenecientes al Consejo de Rectores (CRUCH), un 31% en universidades del CRUCH, un 21% en institutos profesionales y un 10% en centros de formación técnica (Centro de Estudios Mineduc 2012b).

Según la base de datos del Sistema de Investigación en Educación Superior (SIES), al 2016 la matrícula llega a 1.161.222 estudiantes, de ellos, 645.334 ingresaron a la Universidad. Un total de 221.358 ingresó a universidades privadas, 119.351 Universidades Privadas Adscritas al SUA, 167.578 a universidades estatales del CRUCH, 137.047 a universidades privadas del CRUCH (Departamento de Investigación e Información Pública - Consejo Nacional de Educación 2016).

Si bien hay años en que algunas instituciones no entregaron datos, la tendencia se presenta en el siguiente gráfico:

Grafica 1: Progresión de la Matrícula Universitaria por Tipo de Institución.

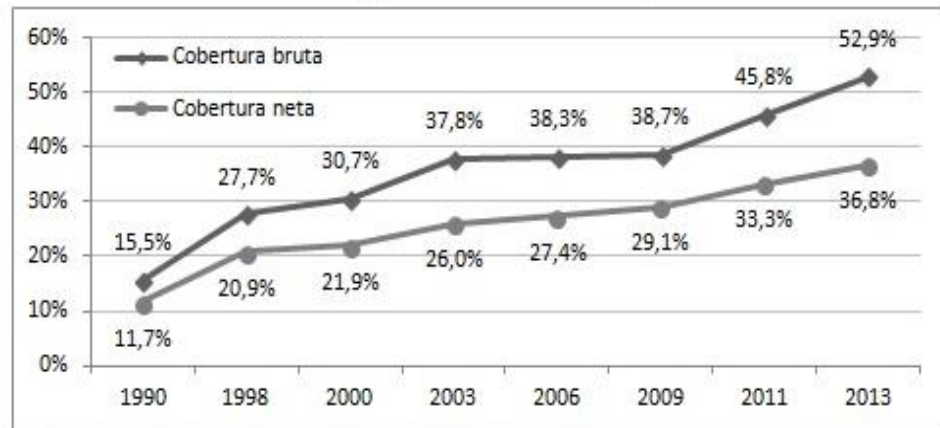


Fuente: Bases de Datos SIES 2016

Sin embargo, el aumento de la matrícula no da cuenta de la tasa de cobertura. Según los datos del SIES la tasa de cobertura bruta fue de 51,9% y la neta de 36,4% para el 2011 (Centro de Estudios Mineduc 2012a). De la encuesta Casen 2011, deriva una tasa de cobertura bruta de un 45.8% y cobertura neta de un 33.3% (Centro de Estudios Mineduc 2012b; Centro de Estudios Mineduc 2012a). La encuesta Casen del 2013 mostró tasas la tasa de cobertura neta de 36% y la de cobertura bruta alcanzó 52,9% (Ministerio de Desarrollo Social 2013). Ambas muestran crecimiento sostenido respecto al 2009, cuando la tasa de cobertura bruta fue de un 38.7% y la cobertura neta de un 29.1% (Centro de Estudios Mineduc 2012b; Centro de Estudios Mineduc 2012a; Libertad y Desarrollo 2015).

Las tasas calculadas a partir de los datos recolectados por el SIES del Ministerio de Educación no identifican datos por sujeto y un estudiante matriculado simultáneamente en dos carreras aparece dos veces. Las tasas calculadas a partir de la encuesta Casen si lo hacen. La encuesta Casen entrega cifras más válidas. Los resultados de la Encuesta CASEN muestran que cobertura en educación superior ha crecido de manera constante en los últimos 20 años (Centro de Estudios Mineduc 2012a; Libertad y Desarrollo 2015; Ministerio de Desarrollo Social 2013), cuestión representada en la siguiente gráfica.

Grafica 2: Tasas de Cobertura en Educación Superior (Libertad y Desarrollo, 2015)

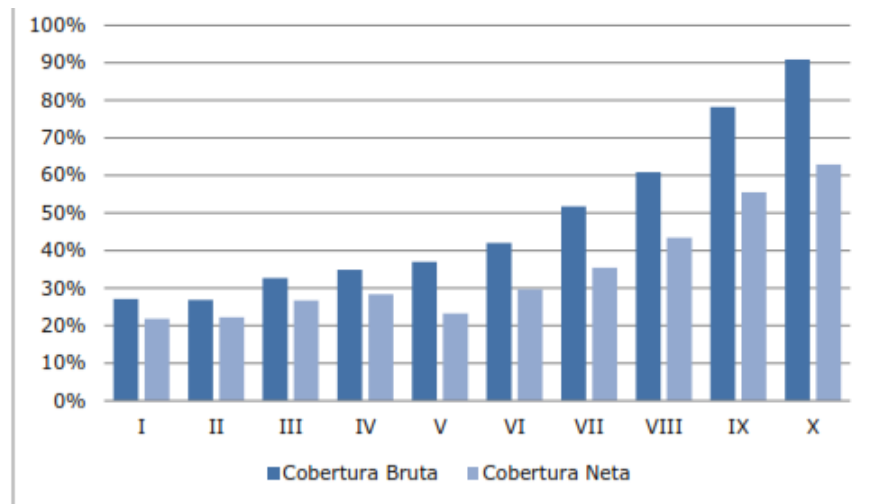


Fuente: Elaboración a partir de Encuestas CASEN, MDS.

Una fuerte expansión de expectativas educacionales de las familias chilenas, presente en todos los estratos de la sociedad (CIDE 2001; CIDE 2003) ha influido en el incremento de la matrícula terciaria. Esto invita a analizar un segundo elemento, la participación de los segmentos de menores ingresos.

Este grupo ha triplicado la participación en educación superior entre 1990 y 2006 – de acuerdo a cifras provistas por la encuesta CASEN– pasando de 4 a 14% en el primer quintil; de 6 a 17% en el segundo quintil y de 10 a 24% en el tercer quintil (Canales & De los Ríos 2009). La diferencia de cobertura entre el primer y el décimo décil disminuyó: en 1990 el 10% más rico tenía una cobertura 10 veces mayor que el 10% de menores ingresos, mientras que el 2011 era 3 veces mayor, aun así, se mantiene una diferencia de acceso importante (Centro de Estudios Mineduc 2012a; Centro de Estudios Mineduc 2012b). Esto se representa en el siguiente gráfico.

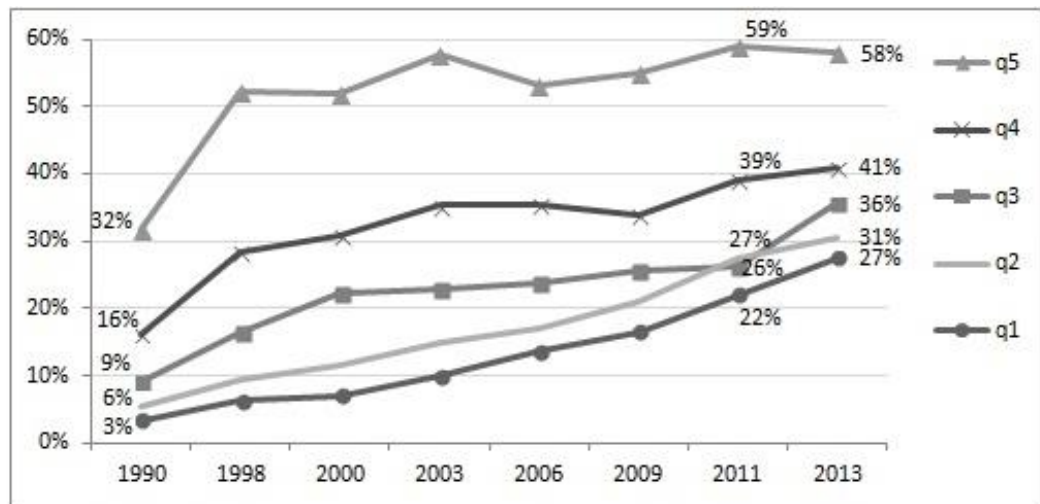
Grafica 3: Cobertura por Décil (Centro de Estudios Mineduc, 2012a)



Fuente: Encuesta CASEN

En la Casen 2013 la brecha se mantuvo, aunque la cobertura neta aumentó en los quintiles más bajos y se estancó que en el quintil más alto: 3 de cada 10 jóvenes del quintil de más bajo de ingresos asisten a educación superior, mientras 6 de cada 10 jóvenes del quintil más alto lo hacen (Ministerio de Desarrollo Social 2013). Esto se representa en la gráfica 4.

Grafica 4: Progresión por Quintil de la Cobertura Neta en Educación Superior

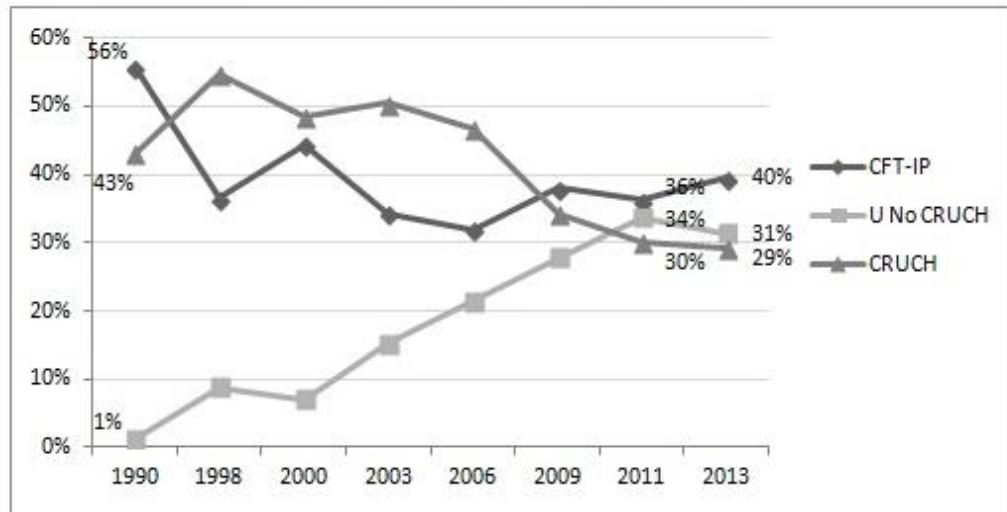


Fuente: Elaboración a partir de Encuestas CASEN, MDS.

Respecto a la distribución de alumnos de los dos primeros quintiles (40% de menores ingresos) el 40% asiste a educación superior técnico profesional, 31% a universidades fuera del CRUCH y 29% a universidades del CRUCH (Libertad y Desarrollo 2015; Ministerio de Desarrollo Social 2013). Si se observa la tendencia desde 1990, las universidades privadas

fuera del CRUCH han contribuido al incremento en el acceso de jóvenes de menores recursos (Libertad y Desarrollo 2015) .

Grafica 5: Distribución de la Matrícula del Quintil I y II por Tipo de Institución



Fuente: Elaboración a partir de Encuestas CASEN, MDS.

Las principales razones dadas por los jóvenes entre 18 y 24 años de los dos quintiles más bajos (q1 y q2) para no estudiar en educación superior fueron: 33% está trabajando o buscando trabajo, 22% evalúa que ya finalizó sus estudios, 12% por embarazo/maternidad y 10% por motivos económicos. Entre los jóvenes de los dos quintiles de ingreso más altos (q4 y q5), las principales razones fueron que trabajaba o busca trabajo (27%) y que ya finalizó sus estudios (17%) (Libertad y Desarrollo 2015; Ministerio de Desarrollo Social 2013).

La incorporación del sector más vulnerado de nuestra sociedad nos conduce a valorar un tercer elemento: el capital cultural de las familias. Antecedentes provistos por el Instituto Nacional de Estadísticas (2003) indican que cerca de un 70% de los estudiantes que actualmente están ingresando al sistema de educación superior, son primera generación en cursar estudios de nivel terciario. Para este grupo, el ingreso al sistema constituye un proceso de adaptación y aprendizaje de nuevos códigos prácticos y simbólicos, situación que los pone en desventaja respecto de otros estudiantes que cuenta con mayor capital cultural y, que por tanto, se encuentran más familiarizados con las demandas del sistema (Canales & De los Ríos 2009).

El ingreso a la universidad está mediado por la prueba de selección universitaria (PSU). Un estudiante que obtiene 475 puntos ha respondido 7 preguntas de un total de 75. Un estudiante que obtiene 600 puntos ha respondido alrededor de 30 preguntas de un total de

75 (Barahona & Aliaga 2013). Los planteles de calidad siempre tienen mayor demanda, por lo que serán éstos los que capten a los mejores alumnos (Donoso & Schiefelbein 2007). El nivel socioeconómico y el capital cultural de las familias son los principales factores que explican las diferencias de rendimiento, tanto entre los estudiantes como entre los establecimientos. Los estudiantes más pobres del sistema no ingresan a universidades más selectivas (del CRUCH) pues las habilidades intelectuales están estrechamente asociadas a la disponibilidad de un nivel importante de capital social, cultural, económico y educacional previo. Los alumnos más pobres, con menor capital social y probablemente con una educación de menor calidad, ingresan finalmente a planteles de baja selectividad y tienen un riesgo más alto de desertar del sistema (Donoso & Schiefelbein 2007).

A esto se debe agregar que los estudiantes suelen elegir en virtud del prestigio de la universidad (Arrau 2003; de los Ríos 2000) sin contar con la información suficiente que permita al alumno elegir adecuadamente (Arrau 2003). Esto, para de los Ríos (2000) da forma a asimetrías de información, que impiden contar con antecedentes suficientes sobre la calidad de las carreras en las diversas universidades.

#### ***5.1.2 Adaptación a la Universidad del Nuevo Estudiante: Estudio Cualitativo Nacional.***

Canales y De los Ríos (2009) abordaron el proceso de persistencia en estudiantes vulnerables ingresados a Universidades Tradicionales. Entrevistaron a 20 hombres y 15 mujeres pertenecientes a los dos primeros quintiles de ingreso de ocho universidades, que pertenecieran a grupos étnicos minoritarios, tuvieran responsabilidades familiares (ser jefe de hogar o tener hijos) y/o trabajaran más de 20 horas semanales, de distintas áreas de conocimiento e instituciones del territorio. Adicionalmente se realizaron dos entrevistas grupales para ampliar la validez de los principales hallazgos.

Las autoras encontraron que los riesgos que enfrentaron los estudiantes vulnerables son variados, destacando las dificultades académicas, emocionales/afectivas, sociales y económicas.

Dentro de las dificultades académicas agruparon la falta de hábitos de estudio y deficiencias en el manejo de contenidos. Las formas de aprendizaje adquiridas en la secundaria no se ajustaron a la demanda un trabajo sistemático, llegando a cuestionar las habilidades y competencias académicas que poseían. Este handicap se arrastró hasta etapas avanzadas del programa de estudios.

Las dificultades económicas se relacionaron a la situación laboral frágil de los adultos con los que convivían (bajos salarios, inestabilidad laboral). Los jóvenes estudiantes trabajaron a tiempo parcial para solventar parte de los gastos que no cubrían las ayudas económicas y los ingresos familiares (internet, computador, fotocopias, movilización, etc.).

En lo social dificultó su progresión formativa que vivieran situaciones como hacinamiento en sus hogares, delincuencia en sus barrios, entre otros. A la vez, padecían el alejamiento del hogar, el desconocimiento de la ciudad o zona donde estudiaban, los largos trayectos de traslado y junto a ello, la pérdida de grupos de referencia. Ellos iniciaron un proceso de adquisición y adaptación a nuevos códigos culturales y simbólicos (de los Ríos & Canales 2007). Con su ingreso a la Universidad surgió la necesidad de generar nuevas amistades y grupos de trabajo académico, cuestión que se contrapuso con sus problemas de integración educativa. En instituciones de elite, se sintieron distintos a sus compañeros y se reconocieron como minoría, por ello, el apoyo de sus pares (grupos minoritarios) resultó crucial durante sus primeros años como soporte “social” significativo y referente para sentirse partícipes e integrados a sus instituciones (Canales & De los Ríos 2009).

En relación a la familia, fueron la primera generación en acceder al sistema universitario. Desconocían la vida universitaria y sus requerimientos, lo que produjo un desajuste de expectativas. Sin embargo, las familias entregaron soporte afectivo y contención, su apoyo disminuyó el estrés asociado a los “fracasos” académicos, evitando valoraciones extremas a éxitos y fracasos. Los padres promovieron valores y actitudes como el esfuerzo, la perseverancia y el logro de metas de largo plazo. El conocimiento de experiencias directas en educación superior, a través de algún familiar, ayudó a adaptarse a los desafíos académicos y extra-académicos.

Las características personales/afectivas que identificaron las autoras en los estudiantes vulnerables que persisten fueron esfuerzo, perseverancia y confianza en sus capacidades o seguridad. Al detallar resultados, estas características se relacionaban a, motivación, autoconocimiento y racionalidad en la toma de decisiones. La motivación se producía porque tenían propósitos o metas claras: anticipan beneficios asociados a la obtención del título profesional, aspiran al bienestar y calidad de vida. En lo económico esperan seguridad, estabilidad económica y movilidad social ascendente. En lo social esperan realizar una contribución a la sociedad o comunidad, el reconocimiento de su familia, “devolver” o “llenar de orgullo” a sus familias (comunidades en indígenas). Dentro de la formación motiva el reconocimiento de otros alumnos que comparten características ‘sociales’ similares (Canales & De los Ríos 2009).

El autoconocimiento era el conocimiento de sus límites y posibilidades, conocimiento notable de capacidades personales, clara conciencia de su vulnerabilidad y de los riesgos a los que están expuestos. La racionalidad en la toma de decisiones implicó considerar sus habilidades y el contexto en el que se desenvuelven. Desarrollaron estrategias con pares para enfrentar sus dificultades académicas (estrategias de estudio colectivo, discusión, lectura y síntesis) y de integración social.

Canales y De los Ríos (2009) sostuvieron que los estudiantes vulnerables pueden llegar a experimentar soledad y angustia, sin que necesariamente existan los soportes institucionales para enfrentar estos desafíos. Los soportes que encontraron fueron: ayudas financieras (créditos y becas), servicios médicos, psicológicos y dentales, recursos docentes y de infraestructura (lugares de estudio, bibliotecas y recursos computacionales sin disponibilidad suficiente ni horarios adecuados a los requerimientos, especialmente en periodos de exámenes). Los sistemas de administración de recursos docentes son poco flexibles para las demandas de distintos grupos de estudiantes vulnerables (quienes trabajan, madres, etc.). Mediaciones humanas (tutores, consultorías docentes, etc.) son escasas y poco sistemáticas.

## **5.2 Modelos Empíricos de Análisis de la Deserción en Educación Superior**

### **5.2.1 Factores Asociados a la Deserción en Estudios Chilenos.**

La deserción de la educación superior se ha estudiado empíricamente identificando las variables asociadas a ella. Los primeros acercamientos fueron descriptivos y abordaron los factores descritos internacionalmente a través de encuestas y entrevistas. Los estudios se realizaron en estudiantes que desertaban de las universidades del CRUCH, donde –por efecto de las barreras de acceso– quien deserta, en general pertenece a sectores medios.

Díaz (2008) señala que los estudios frecuentemente agrupan las variables en cuatro categorías: personales, académicas, institucionales y económicas. Esto coincide con el análisis de González, Uribe y González (2005) y Sánchez, Navarro y García (2009). La construcción teórica describe los factores o dimensiones como:

- Personal: aborda las características del individuo, desde las demográficas como edad y género, psicológicas como la falta de actitud de logro en el crecimiento profesional, poco interés por los estudios en general y la carrera e institución en que estudia en particular, las expectativas respecto de la importancia de carrera que estudia. También se aborda la

poca acogida que brinda la Universidad y la incompatibilidad del tiempo dedicado al trabajo y a los estudios.

- Pedagógico o académico: incluye variables desde el tránsito a la universidad hasta la permanencia en el programa como una orientación profesional que no sustenta la decisión en información sólida antes de ingresar a la licenciatura, las características académicas previas como desarrollo intelectual, rendimiento académico, métodos de estudios, junto a los procesos de admisión y grado de satisfacción con la carrera; carga académica.
- Institucional: incluye la falta de preocupación institucional frente a resultados negativos de la organización universitaria en su conjunto. Aborda el efecto de las normativas académicas, el financiamiento estudiantil y los recursos universitarios como el nivel de formación y profesionalización de los profesores, la manera en que se organiza el trabajo académico, la pertinencia y la actualización del currículo, los apoyos materiales y administrativos, la calidad del programa o carrera, la carencia de una práctica temprana, la ausencia de asignaturas que aproximan al ejercicio profesional desde los primeros años, los contenidos y la forma de estudiar próximas a la enseñanza media y la relación con profesores y pares.
- Socioeconómico y laboral: incluye características socioeconómicas como el estrato socioeconómico, situación laboral del estudiante, situación laboral de los padres y nivel educacional de los padres hasta las condiciones económicas desfavorables del estudiante y la carencia de financiamiento, las bajas expectativas de encontrar trabajo estable y con una remuneración adecuada, la obligación de estar titulado para ejercer y la desarticulación familiar.

Estudios descriptivos realizados en Chile mediante metodología de encuesta y entrevistas –a directivos de organizaciones de educación superior y estudiantes que habían desertado– arrojaron los siguientes resultados:

El estudio de González y col (2005) entre los motivos para desertar –de carreras de ingeniería del CRUCH– detectó, en primer lugar, la disminución de la motivación pues las carreras no fueron lo que los estudiantes deseaban realizar y por tanto sus expectativas se incumplieron, expectativas poco realistas que los autores vincularon al capital cultural de las familias que orienta la elección. Para ellos es fundamental que el apoyo de las familias se relacione con las expectativas de los jóvenes y la realidad de sus carreras. La situación emocional derivada puede traer consecuencias psíquicas y de salud mental que no se produce si los afectados continúan otros estudios. El segundo lugar lo ocuparon: problemas de rendimiento, aprendizaje y para cumplir con las exigencias académicas; también el clima

institucional y la acogida de pares y docentes. El tercer motivo fue el sentimiento de desigualdad frente a los compañeros de carrera, en este sentido el tipo de establecimiento educacional se transformó en un importante factor. Las causas más consideradas por las autoridades fueron: falta de preparación de los estudiantes iniciales, problemas vocacionales (en las terceras, cuartas opciones de acuerdo a sus resultados PSU), dificultad de adaptación a exigencias y condiciones del medio universitario, que los profesores no están preparados para enfrentar a la nueva población estudiantil, factores económicos y familiares.

El Centro Interuniversitario de Desarrollo (CINDA) (2006) reconoció causas socioeconómicas externas al sistema de educación superior como: lugar de residencia, nivel de ingresos, nivel educativo de los padres, ambiente familiar y necesidad de trabajar para mantenerse o aportar a su familia. Causas institucionales fueron: incremento de la matrícula, carencia de mecanismos de financiamiento, desconocimiento de la profesión y metodología de las carreras, ambiente educativo e institucional y carencia de lazos afectivos con la universidad. Causales académicas fueron: formación académica previa, exámenes de ingreso, nivel de aprendizaje adquirido, excesiva orientación teórica, escasa vinculación de los estudios con el mercado laboral, falta orientación teórica, falta de apoyo y orientación de profesores, falta de información al elegir la carrera, carencia de preparación para el aprendizaje y reflexión autónoma, requisitos de los exámenes de grado; excesiva duración de los estudios, heterogeneidad del estudiantado e insuficiente preparación de los profesores para enfrentar la nueva población estudiantil. Entre las causas personales aparecieron las aspiraciones y motivaciones personales, disonancia con expectativas, insuficiente madurez emocional, grado de satisfacción de la carrera, expectativas en el mercado laboral, dificultades para la integración y adaptación, dedicación, falta de habilidades o interés por la carrera escogida.

Centro de Microdatos (2008) organizó el perfil de quien desertaba en tres factores: falta de claridad vocacional pues no queda en la carrera de su preferencia ni tuvo acceso a información y orientación dada habitualmente por familiares; situación económica que acareaba problemas de costo de oportunidad al destinar tiempo de estudio a trabajar (para complementar el ingreso familiar, financiar gastos de mantención, por disminución del ingreso familiar) y rendimiento académico deficiente por baja motivación, problemas vocacionales, debilidades académicas previas, dificultades en metodologías de enseñanza y aprendizaje, insatisfacción con la carrera.

### **5.2.2 Factores Asociados a la Deserción en una Propuesta Iberoamericana.**

En la actualidad es claro que el fenómeno del abandono es multidimensional y sus causas generadoras son diversas y múltiples. Por ello el proyecto Alpha Guía considera que los factores actúan de modo agregado y están explicados por las variables (Hernández 2013; Gómez 2013). Los factores que distinguieron se relacionaron a perspectivas a las que se inscriben estudios e intervenciones. Ellas fueron:

- **Perspectiva individual:** Considera que el abandono obedece a la pérdida de motivación del estudiante que afecta significativamente la persistencia (Fishbein & Ajzen 1975). La teoría sobre “conductas de logro” (Ethington 1990) plantea que el nivel de aspiraciones y expectativas de los estudiantes y sus familias decaen sustancialmente al no adaptarse fácilmente a la vida universitaria y, en algunos casos, variables como las creencias y actitudes de los estudiantes afectan positiva o negativamente su permanencia en la institución académica. Abandonar se entiende como una decisión del estudiante, basada en estas características personales. Variables asociadas a éste factor son: dependencia económica, motivación, hábitos de estudio, adaptación a la vida universitaria.
- **Perspectiva académica:** Hernández (2013) sostiene que el respaldo de la institución educativa de procedencia facilita o no enfrentar las exigencias de la vida universitaria (puntaje en las pruebas previas al ingreso a la universidad, hábitos de estudio) así como la dedicación y el número de créditos académicos cursados inciden en el rendimiento académico y éste en la posibilidad de abandonar la educación superior. El plan de estudios y el acompañamiento durante el proceso formativo tienen incidencia en el rendimiento y en el riesgo de abandonar los estudios.
- **Perspectiva económica:** Generalmente ha planteado el ingreso económico personal y familiar, la capacidad de financiamiento y la satisfacción de terminar un programa académico. En algunos casos se complementa con la valoración del conocimiento (Hernández 2013).
- **Perspectiva cultural:** Para la autora la sociedad actual tiene una crisis de valores implica desapegos, creencias y prácticas que pueden afectar la estabilidad emocional y motivación, y generar espacios propicios para el abandono. Igualmente las creencias y prácticas del contexto cultural del estudiante, pueden incidir positivamente en la búsqueda de estrategias para finalizar sus estudios.
- **Perspectiva institucional:** La institución educativa apoya a los estudiantes con beneficios estudiantiles, disponibilidad de recursos y calidad de la docencia, por ello diseña y pone en marcha programas de bienestar universitario (becas, créditos, auxilios, atención

sicosocial, entre otros), de calidad del ambiente universitario y enseñanza e integración del docente con los estudiantes (Hernández 2013).

### **5.2.3 Factores Asociados a la Deserción en Población Vulnerable.**

A la vez, la literatura internacional que se ha focalizado en estudiantes procedentes de sectores vulnerables –situación riesgosa en sí. Muestran la importancia de factores personales, familiares e institucionales. Estos abordajes son interesantes cuando existe acceso universal o subsidios que apoyan la inserción de este grupo en estudios superiores. Estos abordajes según Canales y De los Ríos (2009) identifican factores protectores que actúan reduciendo el riesgo de deserción y desplazan el foco de intervención desde el estudiante (su capacidad personal y perseverancia) hacia el contexto social e institucional en que se inserta. Centrar la atención en el contexto permite comprender el fenómeno como resultado de la interacción entre un estudiante y el medio en el cual se desenvuelve.

Para Pascarella y Terenzini (Terenzini & Pascarella 1991; Pascarella & Terenzini 2005) al igual que señala Corominas (2001) refuerza la persistencia: la calidad de la institución (pues el prestigio de la institución suele concretarse en el nivel de selectividad de entrada), el esfuerzo exigido en el proceso educativo, la seguridad o incertidumbre en la elección de la carrera, las becas o ayudas económicas en la financiación y la residencia o no en el campus universitario, que influirá en la integración social.

A continuación, se describirá el modo en que se abordan los factores que se asocian y/o causan la deserción desde esta mirada.

#### **5.2.3.1 Características Personales**

El estudio de las características individuales en la permanencia en la educación superior lleva una larga data. Canales y De los Ríos (2009) y Carrasco, Zúñiga, & Espinoza (2014) destacan la importancia de la autoestima, el optimismo, autocontrol, confianza en las capacidades y la perseverancia como características individuales que influyen en la persistencia y el éxito en educación superior. Ambos estudios las asocian al nivel socioeconómico bajo.

La confianza de los estudiantes en sus habilidades genéricas fue abordada por Goldfinch y Hughes (2007) quienes demostraron que la combinación de factores más significativa para explicar el éxito en el primer año fueron una puntuación baja en la escala de estilo de aprendizaje activo y alta confianza inicial en las habilidades de autoeficacia, gestión del tiempo y el trabajo en equipo, junto con una menor confianza inicial en

habilidades de comunicación escrita. Los estudiantes que se retiraron durante el año tuvieron significativamente menos confianza.

Catterall (1998) descubrió que los predictores de recuperación de un desempeño bajo fueron el apoyo de la familia, la capacidad de respuesta a los estudiantes de la escuela, la participación del estudiante en actividades escolares y comunitarias, todos ellos factores de la resiliencia individual. Floyd (Floyd 1996) analizó mediante entrevistas la resiliencia como un factor de éxito académico de 20 afroamericanos empobrecidos, identificó 3 mecanismos de protección: (1) crianza familiar y ambiente hogareño, (2) interacción con jóvenes y la participación comprometida de educadores interesados y otros adultos en sus vidas; (3) el desarrollo principal de dos rasgos de la personalidad: perseverancia y optimismo. La perspectiva de resiliencia ha ganado reconocimiento como marco para examinar por qué sólo algunos estudiantes permanecen, mientras que otros con características sociales y económicas similares, no lo hacen (Catterall 1998; Floyd 1996; Masten 1994; Johnson et al. 2015; Allan et al. 2014; Walker et al. 2006)

Mischel, Zeiss y Zeiss (1974) observaron que cuando los sujetos tenían la percepción de control sobre la tarea –pues su acción influía en el resultado– persistían, sin embargo cuando la actividad fundamental no afectaba el resultado, disminuyó la persistencia. Canales y de los Ríos (2009) sostienen que los estudiantes vulnerables que persistieron fueron aquellos que confiaron en que iban a lograr sus objetivos, más allá de las adversidades que encontraran en el camino.

Baker y Siryk (Baker & Siryk 1984; Baker & Siryk 1989; Baker et al. 1985) sostienen que la persistencia de estudiantes en la educación es un proceso que se relaciona a la diferencia entre expectativas de adaptación a la universidad –que se gestan desde el último año en secundaria– y la capacidad de los estudiantes de alcanzar estas expectativas durante el primer año. La diferencia entre ambas fue llamada desilusión. En relación a su capacidad de adaptarse a la universidad, los estudiantes de primer año esperan más de sí mismos de lo que posteriormente autoperciben que logran. Las diferencias individuales en esta desilusión tiene importantes correlatos conductuales (Baker & Schultz 1992). Baker, McNeil, & Siryk (1985) examinaron la relación entre las diferencias individuales en esta desilusión y el ajuste de comportamientos relevante. Encontraron que estudiantes desilusionados (a) tenían un desempeño académico más pobre, (b) eran más propensos a ser conocidos en el centro de servicios psicológicos, (c) tenían una tasa mayor de abandono de la universidad, y (d) tuvieron menos probabilidades de graduarse a tiempo que los estudiantes no desilusionados. Seis años después de matricularse la tasa de graduación fue de 55% para los estudiantes

desilusionados y 86% para los estudiantes no desilusionados. Ellos llamaron "mito de primer año" a la construcción previa que genera las expectativas que luego son contrastadas con las percepciones del entorno de la universidad y percepciones de ajuste al auto-evaluarse en la universidad. El proceso que estudian es el proceso de anticipación y en él, el ajuste con lo que posteriormente el entorno y ellos mismos esperan de su desempeño de ajuste al ambiente universitario.

Un meta-análisis indicó que el ajuste a la universidad es multidimensional, predictivo de los grados universitarios, y un inusual buen predictor de retención en la universidad. El ajuste se relacionó moderadamente a los rasgos individuales, el apoyo social y las relaciones de los estudiantes con sus padres. Se evidenciaron relaciones más débiles con variables demográficas, rendimiento previo, los enfoques de adaptación, y las variables que reflejan la independencia psicológica de los estudiantes de sus padres (Credé & Niehorster 2012). Lo que Canales y de los Ríos (2009) rescata de este modelo es que sugiere que los estudiantes desarrollan conciencia de la adversidad, visualizan activamente las posibilidades de éxito, es decir, los estudiantes que mejor ajustan sus expectativas de desempeño a las condiciones reales del mismo, son más exitosos.

#### *5.2.3.2 Características Familiares*

La familia ha sido objeto de especial atención en los estudios sobre persistencia y retención. Los padres son figuras relevantes en el desarrollo de habilidades, valores y apoyo social que ayuda a persistir. Jarret (1997) propone que los padres pueden actuar como apoyo, disminuyendo el riesgo de abandono: ellos pueden ayudar a los jóvenes a desarrollar una alta autoestima y un sentido de eficacia educativa a pesar de los obstáculos a los que se ven enfrentados.

El papel de las expectativas paternas es un factor relevante como fuerza motivadora para ingresar a la universidad (Ethington 1990; Canales & De los Ríos 2009). También juegan un rol en el desarrollo de "hábitos de estudio", ayudando a los estudiantes a desarrollar hábitos y prácticas educativas que facilitan el aprendizaje (Ethington 1990). Finalmente, éstas pueden contribuir al desarrollando valores educacionales relevantes como la puntualidad y la responsabilidad (Barnett 2004) lo cual facilita la progresión de los hijos en el sistema educativo.

Las familias juegan un rol importante en la "definición de prioridades". Por ejemplo, ellas pueden actuar, orientando a los estudiantes en las elecciones de asignaturas y programas educativos (Canales & De los Ríos 2009).

### 5.2.3.3 *Características Institucionales*

La persistencia de los estudiantes no puede ser entendida como una capacidad individual de determinados estudiantes, sino como el resultado de la interacción entre este y su contexto (Canales & De los Ríos 2009).

Una de las formas más tradicionales para promover la persistencia de alumnos vulnerables ha sido el uso de apoyos financieros o becas, perspectiva adoptada en el sistema nacional. Sin embargo, como plantean Gladieux y Swail (1998), las ayudas financieras no son suficientes para asegurar la progresión de estos alumnos. Se requiere del uso de estrategias complementarias, para mejorar la permanencia de estudiantes vulnerables en educación superior. Tinto (1987) sugiere los efectos positivos que tiene la interacción entre profesores y alumnos en la permanencia estudiantil, plantea que mientras mayor interacción entre alumnos y profesores, mayores son las probabilidades de que los alumnos persistan y finalicen sus estudios.

La persistencia de los estudiantes con sus instituciones está fuertemente determinada por las redes y relaciones sociales de los alumnos y por ello, Hays y Oxley (1986) plantean que dichas redes son fundamentales en la adaptación e integración estudiantil.

Pascarella y Chapman (1983), han señalado que el compromiso de los alumnos con sus instituciones tiene un efecto significativo en la persistencia de los estudiantes. Barr y Rasor (1999), por otra parte, sugieren que los estudiantes que utilizan los servicios estudiantiles y se afilian a grupos o asociaciones, persisten más que aquellos que no hacen uso de estos servicios o conexiones.

## 5.3 **Deserción Universitaria en Chile:**

### 5.3.1 *Análisis de Casos Nacionales.*

A continuación, se presentan casos, es decir, universidades y carreras en que se analizan las variables asociadas al abandono o buen rendimiento.

Aguirre (2012) estudió la validez predictiva de rendimiento de la PSU en la carrera de Ingeniería Civil en la Universidad de Chile, encontrando que la prueba que mejor explicó el rendimiento académico en el primer semestre fue la PSU de Ciencias, pues explicó en mayor medida la varianza en los datos y tuvo mayor significación en las regresiones de la mayoría de los casos. Los estudiantes que eligieron el módulo de física fueron los que obtuvieron mejores resultados académicos: mejores notas promedio y mayores tasas de

aprobación en los cursos. En menor medida, las notas de enseñanza media (NEM) y la PSU de Matemática tuvieron significancia en una cantidad limitada de casos. El género fue significativo a favor de los hombres. Al corregir los resultados por subestimación (si la predicción de la nota era menor a la nota real del alumno) o sobrestimación en el modelo, se observó que la PSU subestima al género femenino, sobreestima a los colegios municipales y a quienes optan por elegir el módulo de física en la prueba.

Díaz (2009) analizó la deserción en estudiantes de carreras de ingeniería de una cohorte desde año 2004 hasta enero del año 2007. Se evaluó la retención utilizando un modelo de Kaplan-Meier, y se determinó factores de deserción con un modelo de riesgos proporcionales de Cox. Los resultados mostraron que la mayor probabilidad de desertar se presentaba en los tres primeros semestres en estudiantes que provenían de colegios financiados por el estado. Fueron factores determinantes de la retención un alto promedio ponderado acumulado, disponer de crédito universitario, mayor puntaje PSU, provenir de un establecimiento educacional de enseñanza media científico-humanista y altos ingresos promedio familiares. Fueron factores de riesgo que el estudiante tuviera beca, mayor lugar de preferencia en la postulación y mayor edad de egreso de enseñanza media. El autor considera que el efecto negativo de la beca podría deberse a dificultades de integración social debido a una escasa red social, e ingresos familiares insuficientes para cubrir todos los costos universitarios.

Zúñiga, Mena, Oliva, Pedrals, Padilla y Bitran (2009) modelaron mediante análisis multivariante, el rendimiento académico de 272 estudiantes admitidos en la Facultad de Medicina de la Universidad Católica de Chile entre 2000 y 2003. Emplearon como VD variables demográficas, académicas, los métodos utilizados para evaluar el aprendizaje de los estudiantes y sus modos preferidos de procesamiento de la información. De un total de 29 variables modeladas, resultaron predictoras: la postulación previa a medicina, las notas de enseñanza media, el puntaje PSU en Matemática y Biología, el año en que se dictaron los cursos y los métodos de evaluación.

El estilo de procesamiento de la información influyó en el rendimiento en interacción con otras variables (el modo de percepción de información y los métodos de evaluación) en ciclo básico y formación preclínica. En el módulo Integrado I, la evaluación mediante test de elección múltiple correlacionó positivamente con el modo de percepción abstracta y negativamente con el modo activo. En el módulo Integrado II hubo correlación negativa entre Test de elección múltiple y el modo reflexivo, mientras la nota del Tutor correlacionó positivamente con la conceptualización abstracta y negativamente con el modo activo. El

procesamiento de la información influyó en el rendimiento al interactuar con el uso de la información y el sexo: hombres abstractos y activos rindieron mejor que las mujeres con esas características, y las mujeres abstractas reflexivas rindieron mejor que los hombres con similares condiciones (Zúñiga et al. 2009).

En la Universidad de la Frontera se realizó un estudio transversal de las cohortes 2006 a 2011 que comparó estudiantes mapuches con no mapuche analizando con modelo logit. Se determinó que la condición de ruralidad ponía en riesgo de deserción, mientras el nivel socioeconómico, proceder de un establecimiento científico-humanista, el promedio PSU y la vocación (medida por el número de preferencia de la carrera) tuvieron relación directa con la permanencia. Al agregar el rendimiento académico del primer año las variables protectoras dejaron de tener efecto. El avance curricular correlacionó con la permanencia, siendo la única relación que permaneció a partir del tercer año. Los estudiantes mapuche y no mapuche tienen menos probabilidades de desertar en la medida en que aumenta el nivel de avance curricular (Navarrete et al. 2013).

Barahona y Aliaga(2013) estudiaron las variables predictoras del rendimiento académico de los alumnos de primer año de las carreras de Humanidades de la Universidad de Atacama. Para medir rendimiento se empleó las notas, la que se relacionó a la PSU, escolaridad de la madre, promedio de notas de enseñanza media, trabajo, conformidad con la carrera y tipo de establecimiento (público o particular subvencionado). Se encontró que las variables trabajan, conformidad con la carrera, prueba de selección universitaria y promedio de las notas de enseñanza media se relacionaron significativamente y de modo directo, con excepción de trabajar.

Saldaña y Barriga (2010) adaptaron el modelo de Tinto para el análisis de la deserción universitaria en la carrera de Ingeniería Civil de la Universidad Católica de la Santísima Concepción entre 2004 y 2007. Se eliminó todos los casos que ingresaron por admisión especial o que se incorporaron a la carrera en semestres posteriores. Emplearon la metodología de análisis de eventos históricos, modelado a través de una regresión logística binaria, construyendo un modelo que predecía correctamente el 90,3% de los casos. Organizaron el siguiente modelo de análisis en 4 tiempos:

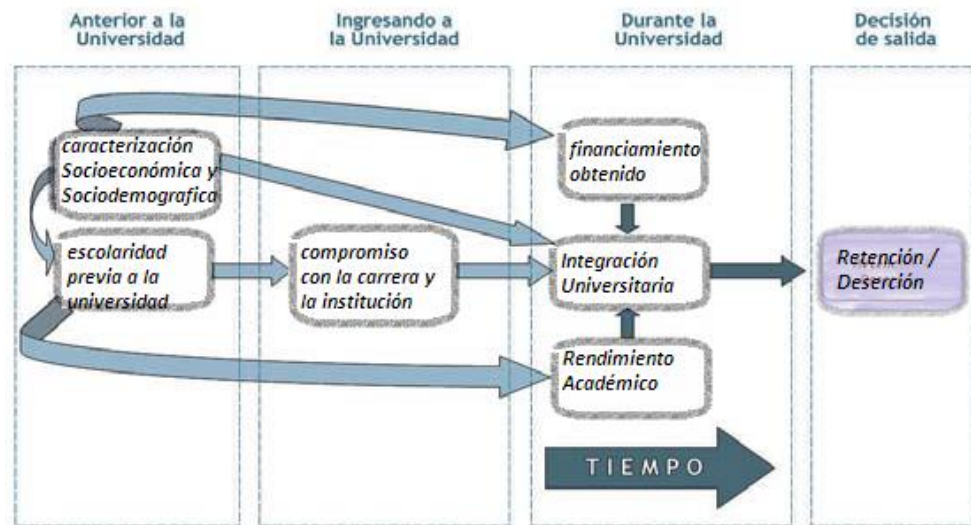


Figura 1: Adaptación del Modelo de Tinto a Ingeniería Civil UCSC

Los factores que más impactaron en la deserción fueron los ingresos familiares y el puntaje en las PSU de Lenguaje y Matemática. Una vez que el alumno ingresó a la universidad, las variables que se sumaron a las anteriores fueron el promedio semestral del alumno, la cantidad de créditos aprobados que el estudiante acumuló en cada semestre, y el porcentaje de financiamiento que haya obtenido para cubrir los costos de arancel anual de carrera. El modelado final de los datos dejó fuera la variable sexo, si bien existieron diferencias significativas que indican que las mujeres desertarían más que los hombres, estas diferencias se estarían manifestando a través de otras variables del modelo. Las variables que mejor explicaron la deserción fueron los ingresos familiares y el rendimiento académico de los estudiantes, con los efectos más fuertes y relación inversamente proporcional con la probabilidad de deserción.

El modelo controló las variables escolaridad previa mediante el puntaje PSU y el rendimiento dentro de la Universidad mediante créditos aprobados y notas. El efecto de los ingresos no pasó por la mala preparación académica producto del menor capital social y cultural de los alumnos, sino por otros aspectos que no han sido medidos en esta investigación, como podrían ser factores actitudinales y el compromiso con las propias metas. Las autoras consideran que en familias donde ambos padres son profesionales se dispone de mayor capital social y cultural que modele la actitud de estudio y finalizar una carrera (Saldaña & Barriga 2010).

### **5.3.2 Estudios Empíricos con Población Nacional**

En nuestro país, la primera década del 2000 los acercamientos al estudio de la deserción universitaria fueron de tipo descriptivo.

De los Ríos y Canales (2007) exploraron los factores explicativos de la deserción temporal y permanente mediante 12 entrevistas en profundidad a desertores permanentes y 12 entrevistas grupales a desertores temporales. En el estudio participaron 67 estudiantes de 4 universidades privadas y 6 universidades públicas (sólo 1 ubicada fuera de la Región Metropolitana). El abandono transitorio se relacionó a: baja integración a carreras o instituciones, baja calidad institucional (asociada a cambios de institución o programas), falta de claridad vocacional (había interés general el área de conocimiento), postulación a carreras no prioritarias para cambiarse internamente a otras carreras.

Los estudiantes que no correspondían con el perfil “tradicional” de alumnos universitarios –cosa manifiesta en su habitus o capital cultural– fueron quienes (a) procedían de colegios particulares subvencionados y municipales y (b) con padres y redes cercanas sin experiencia de estudios superiores. Los primeros tenían problemas de hábitos de estudio: perdían el interés por ir a clases, insistían a pruebas, tenían bajo rendimiento, tenían dificultad para adaptarse al ritmo universitario, desarrollar hábitos de trabajo y de estudio. En los segundos, la información y orientación familiar y social sobre las carreras era débil y las expectativas laborales y profesionales no se cumplían pues desconocían en qué áreas trabajaban los profesionales (de los Ríos & Canales 2007).

En quienes abandonaron de modo permanente, una proporción importante señaló causas económicas (asociadas a cambios en sus condiciones familiares) y familiares (como el nacimiento de un hijo o la muerte del padre o sostenedor), lo que dio cuenta de la vulnerabilidad de diversos grupos de estudiantes. La gran mayoría de ellos mencionaron no haber contado con ayuda institucional –a través de becas o préstamos– que permitiera aliviar su precaria situación financiera. Ellos también tomaron la decisión de estudiar sin demasiada certeza e información, tenían baja identificación con sus carreras y carecían de hábitos de estudios, disciplina y competencias (de los Ríos & Canales 2007).

El año 2002 con seguimiento en 2003, se entrevistó a 399 estudiantes que habían egresado de la educación media sobre la situación luego de egresar y las opiniones respecto del resultado en la PAA. El nivel socioeconómico se relacionó con la actividad que el alumno realizó luego de egresar de la enseñanza media, tres de cada cuatro alumnos del nivel socioeconómico más alto sólo estudiaba, comparado con uno de cada cuatro en el nivel más

bajo. El 31.4 % de los alumnos del estrato más bajo no se encontraban haciendo nada concreto (no estudiaba ni trabajaba), lo que disminuyó a 3.8% en alumnos del nivel más alto. En el nivel más alto, todos rindieron la PAA mientras que en el nivel más bajo sólo lo hizo el 64%. En este mismo sentido, se observó una relación positiva entre el nivel socioeconómico y el puntaje que el alumno obtuvo, logrando los mayores rendimientos el nivel socioeconómico más alto y los menores rendimientos los alumnos del nivel más bajo. Las atribuciones al puntaje inicialmente fueron internas, ya que la mayoría de los alumnos se auto inculpaba, pero la segunda mención fue al colegio como principal responsable del puntaje obtenido en la PAA (Gajardo 2006).

El Centro de Microdatos de la Universidad de Chile (2008) estimó que la tasas de deserción al término del primer año universitario era de 19% promedio en las universidades del Consejo de Rectores y 22% promedio en las universidades privadas sin Aporte Fiscal Directo. Los autores esperaban que al tercer año las tasas acumuladas fueran aproximadamente de 39% y 42% respectivamente. Las áreas con tasas de deserción más altas al primer año fueron: Administración y Comercio, y Derecho (29%), seguida de Humanidades con 28%. Las áreas con menores tasas fueron Salud (16%) y Educación (18%), manteniéndose esta tendencia de deserción en segundo y tercer año. A nivel de carreras, destacó Medicina de universidades del CRUCH en el primer año, con un 98,2% de retención, seguida de Odontología con 94,9% y Psicología con 94%, manteniéndose en altos niveles en años posteriores. En las universidades privadas autónomas también se observó mayor retención en Medicina y Odontología, en niveles de 85,8% y 82,1% respectivamente para el primer año. Sin embargo, en los dos años siguientes, disminuyó la retención hasta alcanzar en el tercer año 70,4% en Medicina y 70,8% en Odontología.

Entrevistas a quienes habían desertado mostraron que un 35% se cambió de carrera y de universidad, 15% se cambió de carrera en la misma universidad, 15% se retiró temporalmente para preparar la PSU nuevamente, 13% siguió la misma carrera, pero en otra universidad, 10% (principalmente de licenciaturas y pedagogías) estudió en CFT o IP. Quienes abandonaron Ingeniería Civil, Ingeniería Comercial y Derecho, continuaron estudios en otras carreras de nivel universitario. Alrededor de un 70% de los encuestados coincidió en que las causas más determinantes al primer año fueron: problemas vocacionales, situación económica de las familias y rendimiento académico. Entre los problemas vocacionales destacó: i) no quedar en la carrera de preferencia del alumno; ii) dificultades en acceso a información y orientación. Los hombres desertaron más que las mujeres, especialmente a partir del segundo año (Centro de Microdatos 2008).

Los estudios que marcaron la tendencia en el análisis de la deserción y permanencia en la educación superior desde el 2010 procedieron de la Escuela de Ingeniería de la Universidad Católica. Éstos tomaron los datos de diversas bases del Ministerio de educación (Departamento de Evaluación, Medición y Registro Educacional (DEMRE), el Servicio de Información de Educación Superior (SIES), el Departamento de Ayudas Estudiantiles del Ministerio de Educación y la Comisión Ingresos) accediendo a varias cohortes que permitieron la reconstrucción longitudinal de la trayectoria de estudiantes desde el egreso de educación media e inscripción en la PSU 2007 al 2010, pudiendo construir indicadores de abandono y persistencia.

En Chile se empiezan modelar empleando modelos lineales generalizados, inicialmente aplicados al efecto de las ayudas por parte del Estado. Se empleó el análisis de regresión logística (Logit) (Barrios et al. 2011; Barrios 2011a) y probit estándar (Meneses & Blanco 2010). Estos investigadores empezaron a corregir las mediciones y controlar la endogeneidad de los datos mediante procedimiento de dos o más pasos. Esta forma de análisis se replicó en otros estudios (Santelices et al. 2011; Centro de Estudios 2012; Navarrete et al. 2013). A continuación, se presentan los hallazgos de estos estudios.

El reporte de Meneses y Blanco (2010) abordó el efecto de las becas y créditos en la matrícula de educación superior. Determinó que en la educación universitaria y la ayuda financiera gubernamental (becas y créditos) aumentó en un 30% la probabilidad de inscripción. El modelo probit indicó que estar preseleccionado para la ayuda financiera aumentó en un 19% la probabilidad de inscripción en la universidad. El resultado no sesgado del procedimiento en dos etapas mostró que ser pre-seleccionado para la ayuda financiera aumentó la probabilidad en un 33%. Estas regresiones se realizaron por el control de puntajes PSU, SIMCE, los grados de secundaria y variables ficticias, para diferentes niveles de ingreso. La ayuda financiera orientó la elección a la Universidad dentro del sistema de educación superior y redujo la probabilidad de inscripción en los otros niveles del sistema, lo que los autores vincularon a las expectativas de rentabilidad y prestigio de la formación profesional.

En el estudio de Barrios (2011b) las variables que resultaron significativas en relación a la deserción fueron: que ningún padre llegara a la universidad, que los estudiantes vivieran solos, el aumento de la edad en hombres y en mujeres estar casadas. Señaló que un aumento en los ingresos generó disminución en las probabilidades de deserción sólo al superar los \$2.500.000. Los estudiantes de menores ingresos tienen más probabilidades de suspender al igual que los de más altos ingresos. El Crédito con Aval del Estado redujo en mayor medida

las probabilidades de deserción que el Fondo Solidario de Crédito Universitario y ambos créditos tuvieron un menor impacto a medida que aumentó el ingreso, de modo que ninguno de los dos créditos resultó significativo a partir de ingresos mayores a \$1.400.000. La variable beca no tuvo efecto sobre la deserción excepto en los estudiantes de educación, el autor relacionó lo primero a la ausencia de costo de salida y lo segundo a que los estudiantes de educación desertan muy poco. Sugiere profundizar este aspecto diferenciando tipo de beca. Para el autor no se justifica entregar becas ni subsidiar las tasas de los créditos, pues aumentar la permanencia en el sistema se cumple mejor sin ello.

Barrios (2011a) siguió el comportamiento de los estudiantes entre los años 2006 y 2008 para determinar abandono el primer o segundo año. Detectó que a mayor ingreso familiar menor deserción, a la vez, el efecto del ingreso decreció con su aumento. Haber recibido algún tipo de ayuda socioeconómica redujo las probabilidades de deserción, efecto que también decreció a medida que aumentó el ingreso. Por ello sostuvo que los instrumentos de financiamiento reducían las probabilidades de deserción con mayor efectividad a menores ingresos.

Tuvieron menor probabilidad de desertar las mujeres, los estudiantes solteros, los de mayor promedio de educación media y puntajes más altos en la PSU. El nivel educacional de los padres prácticamente no afectó las probabilidades de deserción y las aumentó la mayor edad del estudiante al momento de ingresar a la universidad y proceder de escuelas municipales. Se encontró también que y que las características del programa y de la casa de estudios afectaban las probabilidades de deserción (Barrios 2011a).

El Centro de Estudios del MINEDUC (2012), estudió la deserción del sistema universitario en estudiantes que rindieron la PSU 2007 y se matricularon en primer año en alguna de las universidades existentes en el país, replicando el modelo de Barrios, Meneses y Paredes (2011) sin establecer relaciones causales, abordaron la relación entre la permanencia en el sistema y las variables analizadas. Encontraron correlación positiva con ser mujer, el nivel de ingresos familiar, asistir a instituciones acreditadas, ingresar al programa en primera opción, un mejor puntaje NEM y una mejor PSU. Todas las ayudas estudiantiles resultaron significativas y positivas para permanecer en el sistema, siendo mayor la relación con el CAE, seguido por el Fondo Solidario, lo que dejó en el último lugar a las becas.

Se relacionó negativamente con la permanencia en el programa la edad de ingreso del estudiante, ser soltero, tener padres que no completaron la educación escolar, vivir de manera

independiente, la duración del programa y egresar de colegios municipales (Centro de Estudios 2012)

Desde el departamento de Ingeniería industrial de la Universidad de Chile y analizando las mismas bases de datos se realizaron dos estudios longitudinales siguiendo a una cohorte desde que rindió el SIMCE de segundo medio el año 2003. Ésta egresó de enseñanza media el 2005, rindió la PSU el año 2005, y entró al sistema de educación superior el año 2006, donde se entendió por deserción bruta a quienes no registraron matrícula a partir del año 2007 o 2008 y por deserción general, a aquellos que se cambiaron de carrera o universidad.

Acuña (2012) analizó mediante un modelo probit con selección en la muestra (bi-probit censurado en la probabilidad de deserción), teniendo como variable dependiente la probabilidad de desertar corregida por la probabilidad de acceso. Hizo un hallazgo muy interesante: al observar que los quintiles del alumno en el SIMCE tuvieron una correlación no lineal y en un efecto creciente con el acceso, del mismo modo, aunque con efecto inverso, sucedió con los quintiles PSU del alumno: en la medida que aumentaba el quintil menor era la probabilidad de desertar. Analizó por establecimiento y descubrió que los quintiles PSU del establecimiento tenían en la mayoría de los casos una correlación mayor con la deserción que los quintiles del alumno. De ello dedujo la importancia de la calidad los establecimientos de educación secundaria sobre el rendimiento posterior de los alumnos, pues al controlar los factores relacionados a los establecimientos de educación superior, las variables institucionales más fuertes se relacionaron al quintil del establecimiento en el SIMCE y la PSU.

En el caso de las variables asociadas a la institución de educación superior sólo presentó significancia que la institución estuviera acreditada. En financiamiento se encontró una correlación inversa y alta entre deserción y la proporción que financia el alumno vía beca o crédito, siendo más fuerte la relación con la proporción que se financia vía crédito. Con respecto a las variables del background familiar, se destacó la correlación y significancia que tuvo en la deserción tener hermanos en educación superior, enseñanza básica y pre-básica: mientras un hermano en educación superior disminuyó la probabilidad de desertar, tener 2 o más hermanos en educación prebásica y básica la aumentó. Por otra parte, padres con un nivel educacional inferior a la educación media aumentó significativamente la probabilidad de deserción (Acuña 2012).

En la misma cohorte Larroucau (2013) utilizó un modelo econométrico estático de tres ecuaciones para corregir, simultáneamente, la posible endogeneidad presente en la

selección de muestra y en la obtención de las ayudas estudiantiles, diferenciando la deserción temporal de la definitiva, analizando posibles efectos heterogéneos entre áreas de carrera y tipos de universidad. Sus hallazgos dieron cuenta de la heterogeneidad del sistema en el país.

Encontró tasas del 16% de deserción definitiva y 12% de deserción temporal (ligada al cambio de carrera y universidad), alcanzándose cerca de un 29% de deserción general, procesos ocurridos principalmente los dos primeros años de universidad, por lo tanto, las variables de acceso son determinantes. Concluyó que las características personales del estudiante y de su núcleo familiar fueron factores predictivos, cobrando un importante valor explicativo la calidad del establecimiento de origen, la calidad de la universidad y el tipo de carrera. Las ayudas estudiantiles disminuyeron la probabilidad de desertar definitivamente en más de un 60%.

Dentro de las características personales del estudiante tuvo valor explicativo el puntaje NEM y el ranking de notas de enseñanza media, teniendo un menor poder explicativo el puntaje PSU en matemática y lenguaje. Dentro del grupo familiar, el nivel ingreso del núcleo y el financiamiento con crédito, becas, o una combinación de ambos tuvo valor significativo, en menor grado afectó el nivel educacional alcanzado por la madre, la edad y el género del estudiante. Al igual que en el estudio de Acuña (2012) la calidad del establecimiento de origen afectó fuertemente la deserción: estudiantes provenientes de colegios con mejores puntajes en la PSU de matemática y menor grado de vulnerabilidad económica, presentaron menor probabilidad de deserción, incluso al controlar por la habilidad individual del estudiante. El autor hipotetizó que esto se podía deber a una mejor red de apoyo, o bien, el desarrollo de competencias no evaluadas en el puntaje PSU individual (Larroucau 2013).

En relación a la calidad de la Universidad los estudiantes que accedieron a universidades de mejor calidad presentaron menor probabilidad de desertar. El mayor valor explicativo lo alcanzó el puntaje promedio de PSU y la acreditación de la universidad sin que fuera relevante la selectividad, sin embargo, los estudiantes con menor probabilidad de desertar fueron los de mayores ingresos y mejor ranking de notas, que accedían a universidades selectivas del CRUCH. Al analizar por áreas de conocimiento, la relación con la deserción cambia drásticamente al comparar carreras de Educación y Humanidades no acreditadas con carreras acreditadas entre 6 y 7 años, lo que evidenció las diferencias de calidad entre las carreras de pedagogía (Larroucau 2013).

En el factor carrera, la mayor probabilidad de deserción se relacionó al sector “Ciencias” y la menor probabilidad a “Educación y Humanidades”. La PSU de matemática,

el nivel de ingreso familiar, el ranking de notas y la acreditación de la universidad interactuó con el área de conocimiento. La PSU de matemática afectó de mayor manera en carreras de “Ciencias” (hay mayor probabilidad de desertar definitivamente en los quintiles intermedios de la distribución), el ranking de notas del estudiante afectó en mayor medida a las carreras de “Salud” y “Educación y Humanidades” (menor deserción definitiva en el mayor quintil). El ingreso familiar afectó a las carreras de “Ciencias” (los estudiantes pertenecientes a los tramos de menor ingreso tuvieron mayor probabilidad de desertar definitivamente). Finalmente los estudiantes que accedieron a carreras con mejor empleabilidad tenían menor probabilidad de desertar (Larroucau 2013).

Los estudios, al igual que los modelos teóricos desarrollados por los autores revisados, dan cuenta que luego del segundo año, variables de rendimiento académico e inserción social en la Universidad, son mejores indicadores. Por lo tanto, las variables analizadas son funcionales para el análisis del primer año en un programa. Se presenta una tabla que resume el comportamiento de las variables presentes en los estudios nacionales documentados en este apartado (ver anexo1) para facilitar el análisis del valor asumido por las variables en los estudios empíricos nacionales, fundamental para planificar un modelo de clasificación de los estudiantes en riesgo de deserción durante el primer año en la educación superior.

## **6 MARCO TEÓRICO**

### **6.1 Inteligencia de Negocios**

#### ***6.1.1 Antecedentes del uso de Inteligencia de Negocios en Educación Superior***

La Inteligencia de Negocios (en adelante BI), es un término contemporáneo que se refiere a los datos y herramientas de software para organizar, analizar y proveer acceso a la información para ayudar a los gerentes y demás usuarios empresariales a tomar decisiones más documentadas (Laudon & Laudon 2012), y se compone de varias tecnologías, métodos y herramientas que ayudan principalmente a los niveles gerenciales altos y medios en el proceso de toma de decisión.

En la mayoría de los sectores industriales se ha incorporado y comenzado a utilizar herramientas, metodologías y tecnología de BI para obtener una ventaja competitiva con respecto a sus competidores, a través del uso de grandes volúmenes de datos históricos acumulados. Esta enorme cantidad de datos requiere un proceso de tratamiento especial para

ir en la mejora de los procesos de toma de decisión en las organizaciones. El enfoque de BI, ayuda en la construcción de múltiples vistas de todos los sistemas de la organización que comprenden datos de sus clientes, proveedores, competidores, y de recursos humanos (Hart et al. 2007).

La necesidad de BI para lograr una ventaja competitiva en las Universidades ha ganado un gran impulso en la última década, y el atractivo de implementarlo en las universidades es debido a varias razones: primero, actualmente las IES enfrentan una dura competencia y por lo tanto necesitan entender de mejor forma las fuerzas de los negocios para responder de manera efectiva a los dinámicos requerimientos de la industria. Segundo, las universidades siempre requieren de herramientas para predecir el rendimiento de los estudiantes, los patrones de empleabilidad, selección de cursos, deserción, etc. Tercero tal como todos los negocios, las universidades requieren de un análisis de costo beneficios, análisis de tendencias, análisis de su cadena de valor, etc. Todos estos requerimientos podrían ser soportados por aplicaciones de BI (Laudon & Laudon 2012).

### ***6.1.2 Herramientas para la Inteligencia de Negocios***

Las principales herramientas para la inteligencia de negocios incluyen el software para consultas e informes de bases de datos, herramientas para el análisis de datos multidimensional (procesamiento analítico en línea), y herramientas para la minería de datos (Laudon & Laudon 2012).

#### ***6.1.2.1 Procesamiento analítico en línea (OLAP)***

El Procesamiento Analítico en Línea (OLAP) soporta el análisis de datos multidimensional, el cual permite a los usuarios ver los mismos datos de distintas formas mediante el uso de varias dimensiones. Cada aspecto de información —producto, precios, costo, región o periodo de tiempo— representa una dimensión distinta. Así, un gerente de productos podría usar una herramienta de análisis de datos multidimensional para saber cuántas arandelas se vendieron en el Este en junio, cómo se compara esa cifra con la del mes anterior y con la de junio del año anterior, y cómo se compara con el pronóstico de ventas. OLAP permite a los usuarios obtener respuestas en línea a las preguntas ad hoc tales como éstas en un periodo de tiempo bastante corto, incluso cuando los datos se almacenan en bases de datos muy grandes, como las cifras de ventas de varios años (Laudon & Laudon 2012).

### 6.1.2.2 Minería de Datos

La Minería de Datos está compuesta por un conjunto de técnicas y algoritmos que se utilizan para analizar un conjunto de datos, en donde se pueden extraer ciertos patrones y relaciones entre ellos, entregando así información útil para los niveles de tomas de decisión en las organizaciones. En sí está más orientada al descubrimiento, y es considerada una herramienta avanzada en los temas de BI, y como una ayuda relevante en la aplicación de la estrategia de las organizaciones.

La Minería de Datos se utiliza en varios sectores industriales, tales como el financiero, telecomunicaciones, en las ventas por menor, y últimamente en la Educación (El Sheikh & Alnoukar 2011). Las IES han demostrado su interés en la Minería de Datos debido al alto potencial que tiene en su dominio (Alzoabi et al. 2011), por ejemplo Erdogan y Timor (2005) usaron la Minería de Datos para estudiar el efecto de los resultados de las pruebas de admisión y su correlación con el rendimiento de los estudiantes en la educación superior.

En las IES a nivel global existen varios estudios de Minería de Datos, por ejemplo Superby et al. (2006) analizó a los estudiantes de primer año para identificar las características que podrían tener mayores probabilidades de éxito académico. Kovačić (2010) trató de identificar predictores de alarma temprana de rendimiento de estudiantes en universidades, focalizándose en variables sociodemográficas (edad, sexo, etnia, educación, situación laboral, discapacidad) y ambiente de estudio (programa y semestre), que pueden influir en la persistencia y abandono (deserción) de los estudiantes de la universidad “Open Polytechnic” de Nueva Zelanda. Yu et al (2007) usó un enfoque de Minería de Datos para identificar los predictores de retención entre los estudiantes de primer año en la Universidad del Estado de Arizona.

Según Fayyad, et al. (1996) el uso de minería de datos se debe entender como un apoyo para los analistas, y no reemplaza al conocimiento que tienen los expertos del negocio, ni elimina la necesidad de entender los datos. Por consiguiente, no funciona por sí sólo, ya que los patrones que se encuentren en los datos deben ser interpretados y validados para ver si responden a las consultas del negocio, y si son aplicables en el mundo real.

Para estos autores los tipos de información que se pueden obtener de la minería de datos son: asociaciones, secuencias, clasificaciones, agrupamientos y pronósticos. Las *asociaciones* son ocurrencias vinculadas a un solo evento. Por ejemplo, un estudio de los patrones de compra en supermercados podría revelar que, cuando se compran frituras de maíz, el 65 por ciento del tiempo se compra un refresco de cola, pero cuando hay una

promoción, el 85 por ciento se compra un refresco de cola. Esta información ayuda a los gerentes a tomar mejores decisiones debido a que descubren la rentabilidad de una promoción.

En las *secuencias*, los eventos se vinculan en el transcurso del tiempo. Por ejemplo, podríamos descubrir que, si se compra una casa, el 65 por ciento del tiempo se compra un nuevo refrigerador dentro de las siguientes dos semanas, y el 45 por ciento se compra un horno dentro del mes posterior a la compra de la casa.

La *clasificación* reconoce los patrones que describen el grupo al que pertenece un elemento, para lo cual se examinan los elementos existentes que hayan sido clasificados y se infiere un conjunto de reglas. Por ejemplo, las empresas como las compañías de tarjetas de crédito o las telefónicas se preocupan por la pérdida de clientes estables. La clasificación ayuda a descubrir las características de los clientes con probabilidades de dejar de serlo y puede proveer un modelo para ayudar a los gerentes a predecir quiénes son esos clientes, de modo que puedan idear campañas especiales para retenerlos.

El *agrupamiento* funciona de una manera similar a la clasificación cuando aún no se han definido grupos. Una herramienta de minería de datos puede descubrir distintas agrupaciones dentro de los datos, como el hecho de encontrar grupos de afinidad para tarjetas bancarias o particionar una base de datos en grupos de clientes con base en la demografía y los tipos de inversiones personales.

Aunque estas aplicaciones implican predicciones, el pronóstico utiliza las predicciones de una manera distinta. Se basa en una serie de valores existentes para pronosticar cuáles serán los otros valores. Por ejemplo, el pronóstico podría encontrar patrones en los datos para ayudar a los gerentes a estimar el futuro valor de variables continuas, como las cifras de ventas. El análisis predictivo utiliza las técnicas de minería de datos, los datos históricos y las suposiciones sobre las condiciones futuras para predecir los resultados de los eventos, como la probabilidad de que un cliente responda a una oferta o que compre un producto específico (Fayyad, U. et al. 1996).

### **6.1.3 *Proceso de Descubrimiento de Conocimiento en Base de Datos, (Knowledge Discovery in Databases, KDD).***

KDD es un proceso genérico para encontrar información en un conjunto de datos, y con ello generar conocimiento. En un nivel abstracto, el campo KDD se refiere al desarrollo de métodos y técnicas para dar sentido a los datos. El problema básico del proceso KDD es el mapeo de datos de bajo nivel (que suelen ser demasiado voluminosos para comprenderlos

y procesarlos fácilmente) en otras formas que pueden ser más complejas (por ejemplo, un breve informe, una aproximación descriptiva o modelo del proceso que generó los datos), o más útil (por ejemplo, un modelo predictivo para estimar el valor de casos futuros). En el centro del proceso está la aplicación de métodos específicos de Minería de Datos (data-mining) para el descubrimiento y extracción de patrones (Ronald J. Brachman & Anand 1996).

*“en la literatura científica se puede encontrar una variedad de definiciones acerca del descubrimiento de conocimiento en base de datos, sin embargo una de las definiciones es la siguiente: “El descubrimiento de conocimiento en bases de datos es un campo de la inteligencia artificial de rápido crecimiento, que combina técnicas del aprendizaje de máquina, reconocimiento de patrones, estadística, bases de datos, y visualización para automáticamente extraer conocimiento (o información), de un nivel bajo de datos (bases de datos)” (Han et al. 2012).*

Entre las tareas más comunes que se desarrollan a través del proceso KDD están: Predicción, Clustering, Clasificación, Detección de anomalías (outliers), Asociaciones, Visualización, Comprensión de información.

Los tipos de Algoritmos usados en KDD son: Aprendizaje supervisado, Aprendizaje no supervisado, Aprendizaje reforzado, Aprendizaje activo o semi-supervisado

En la figura 2 se muestra el proceso KDD, compuesto por una etapa de selección, pre procesamiento, transformación, Minería de Datos, Interpretación/ Evaluación.

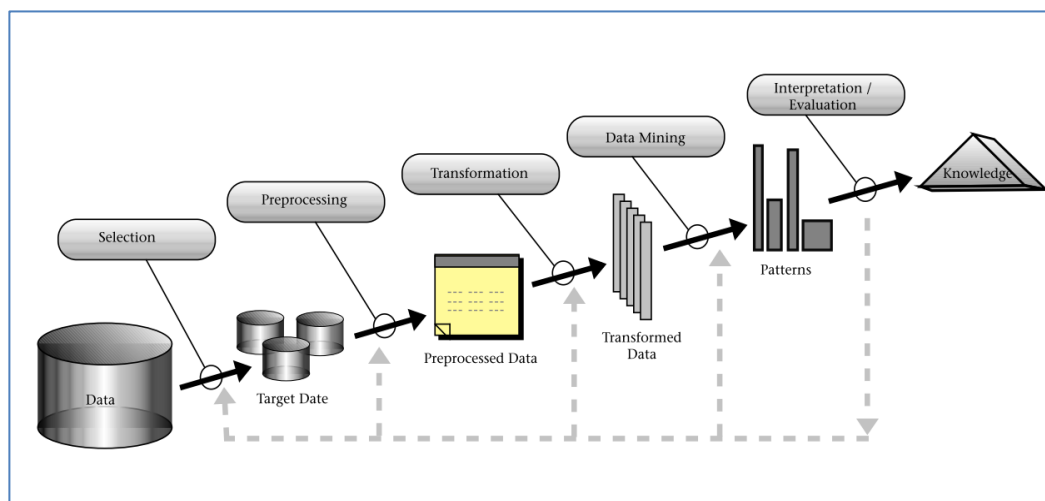


Figura 2: Vista general de los pasos que componen el proceso KDD.

Fuente: Fayyad et al. (1996)

#### **6.1.4 El proceso KDD**

El proceso KDD es un proceso centrado en el usuario, esto significa que tiene la propiedad de ser altamente interactivo e iterativo, que puede ser guiado por las decisiones que toma el usuario, o un agente inteligente (Fayyad, U. et al. 1996). Para los autores el proceso tiene nueve pasos (ver anexo 2).

El proceso KDD puede implicar una iteración significativa y puede contener bucles entre dos pasos cualesquiera, tal como se muestra en el flujo básico de pasos de la figura 2.

#### **6.1.5 Etapa de Minería de Datos del proceso KDD**

Fayyad et al (1996) plantean que el componente de minería de datos del proceso KDD a menudo implica la aplicación iterativa repetida de métodos particulares de minería de datos. Por otra parte, las metas de descubrimiento de conocimiento se definen por el uso previsto del sistema. Podemos distinguir dos tipos de metas: verificación y descubrimiento. Con la verificación, el sistema se limita a verificar la hipótesis del usuario. Con el descubrimiento, el sistema encuentra de forma autónoma nuevos patrones. Además, subdividimos el objetivo de descubrimiento en predicción, donde el sistema encuentra patrones para predecir el comportamiento futuro de algunas entidades y descripción, donde el sistema encuentra patrones para presentarlos a un usuario en una forma comprensible por el hombre.

Para los autores la minería de datos implica la adaptación de modelos a, o la determinación de patrones a partir de datos observados. Los modelos ajustados desempeñan el papel de conocimiento inferido: si los modelos reflejan conocimientos útiles o interesantes, forman parte del proceso KDD global e interactivo en el que suele requerirse el juicio humano subjetivo. Dos formalismos matemáticos primarios se utilizan en el ajuste del modelo: la estadística y la lógica.

El enfoque estadístico permite efectos no determinísticos en el modelo, mientras que un modelo lógico es puramente determinista. Nos centraremos principalmente en el enfoque estadístico de la minería de datos, que tiende a ser la base más utilizada para aplicaciones prácticas de minería de datos dada la presencia típica de incertidumbre en los procesos de generación de datos del mundo real (Fayyad, U. et al. 1996).

#### **6.1.6 Métodos de Minería de Datos**

Los métodos en la Minería de Datos pueden ser vistos como tres componentes algorítmicos primarios: representación del modelo, evaluación del modelo y búsqueda.

Los dos objetivos principales de alto nivel de la minería de datos en la práctica tienden a ser la predicción y la descripción. Como se dijo anteriormente, la predicción implica el uso de algunas variables o campos en la base de datos para predecir valores desconocidos o futuros de otras variables de interés, y la descripción se centra en encontrar patrones interpretables por humanos que describen los datos. Aunque los límites entre la predicción y la descripción no son agudos (algunos de los modelos predictivos pueden ser descriptivos, en la medida en que son comprensibles, y viceversa), la distinción es útil para comprender la meta general de descubrimiento. La importancia relativa de la predicción y la descripción para aplicaciones específicas de minería de datos puede variar considerablemente. Los objetivos de predicción y descripción se pueden lograr usando una variedad de métodos de minería de datos particulares, entre los cuales están los siguientes (Han et al. 2012):

La *clasificación* es el aprendizaje de una función que mapea (clasifica) un elemento de datos en una de varias clases predefinidas. La *regresión* es el aprendizaje de una función que asigna un elemento de datos a una variable de predictora de valor real. *Clustering* es una tarea descriptiva común en la que se busca identificar un conjunto finito de categorías o clusters para describir los datos.

#### **6.1.7 Componentes de los algoritmos de minería de datos**

El siguiente paso es construir algoritmos específicos para implementar los métodos generales descritos anteriormente. Según Han et al. (2012) se puede identificar tres componentes primarios en cualquier algoritmo de minería de datos: representación del modelo, evaluación del modelo y búsqueda.

Esta visión reduccionista no es necesariamente completa ni totalmente abarcadora; más bien, es una forma conveniente de expresar los conceptos claves de los algoritmos de minería de datos de una forma relativamente unificada y compacta.

##### **6.1.7.1 Representación del modelo:**

Es el lenguaje utilizado para describir los patrones descubiertos. Si la representación es demasiado limitada, entonces ninguna cantidad de tiempo de entrenamiento o ejemplos puede producir un modelo preciso para los datos. Es importante que un analista de datos entienda completamente las suposiciones representacionales que podrían ser inherentes a un método en particular. Es igualmente importante que un diseñador de algoritmos indique claramente qué suposiciones representacionales están siendo hechas por un algoritmo particular. Se debe tener en cuenta que el aumento de la potencia de representación de los

modelos aumenta el peligro de sobreentrenamiento de los datos de entrenamiento, lo que resulta en una reducción de la exactitud de predicción en los datos no vistos.

#### *6.1.7.2 Los criterios de evaluación de modelos*

Son declaraciones cuantitativas (o funciones de ajuste) de lo bien que un patrón en particular (un modelo y sus parámetros) satisface las metas del proceso KDD. Por ejemplo, los modelos predictivos a menudo se juzgan por la exactitud de la predicción empírica en algún conjunto de pruebas. Los modelos descriptivos pueden ser evaluados a lo largo de las dimensiones de precisión predictiva, novedad, utilidad y comprensibilidad del modelo ajustado.

#### *6.1.7.3 El método de búsqueda*

Consta de dos componentes búsqueda de parámetros y búsqueda de modelos. Una vez fijada la representación del modelo (o familia de representaciones) y los criterios de evaluación de modelos, el problema de minería de datos se ha reducido a puramente una tarea de optimización: buscar los parámetros y modelos de la familia seleccionada que optimicen los criterios de evaluación. En la búsqueda de parámetros, el algoritmo debe buscar los parámetros que optimizan los criterios de evaluación de modelos dados los datos observados y una representación fija del modelo. La búsqueda de modelos se produce como un bucle sobre el método de búsqueda de parámetros: la representación del modelo se cambia para que se considere una familia de modelos.

### **6.1.8 Métodos de Minería de Datos más populares**

Según Han et al. (2012) existe una amplia variedad de métodos de minería de datos, pero aquí, sólo nos enfocamos en dos técnicas de clasificación a usar en el presente estudio. Cada método se discute en el contexto de la representación del modelo, la evaluación del modelo y la búsqueda.

#### *6.1.8.1 Árboles de Decisión*

Los árboles de decisión y las reglas que utilizan divisiones univariadas tienen una forma representativa simple, haciendo que el modelo inferido sea relativamente fácil de comprender para el usuario. Sin embargo, la restricción a una representación particular de árbol o regla puede restringir significativamente la forma funcional del modelo y, por tanto, el poder de aproximación.

Los árboles de decisión son técnicas de clasificación, que contiene nodos y uniones dirigidas entre los nodos (links). cada nodo interno representa un atributo, y en donde se realiza un test basado en los valores de los atributos, y los links representan los resultados del test. los nodos hojas representan el resultado de la clasificación, luego para para clasificar un registro se debe recorrer desde el nodo raíz a la hoja resultante, y el camino recorrido dependerá de los valores del registro. los árboles de decisión son una técnica de aprendizaje supervisado, por lo tanto, necesitan de un set de registros pre-clasificados, conocidos como set de entrenamiento, que permite ajustar el modelo (encontrar una estructura apropiada para el árbol, explorar el espacio de hipótesis buscando un buen clasificador.

#### *6.1.8.2 Redes Neuronales Artificiales.*

Las redes neuronales artificiales (ANN) son comúnmente conocidas como técnicas matemáticas biológicamente inspiradas, capaces de modelar funciones no lineales extremadamente complejas (Haykin, 2008). Las redes neuronales artificiales pueden utilizarse tanto en la explicación como en diversos objetivos que cubren los modelos prospectivos en los estudios de minería de datos.

Las redes neuronales artificiales ponen en relación no lineal la variable dependiente con las variables independientes basándose en datos sin ningún supuesto estadístico. El Perceptrón multicapa es un método que utiliza un método de aprendizaje monitoreado y generalmente se usa para objetivos de clasificación académicos o experimentales. En este estudio, utilizamos una popular arquitectura de ANN denominada Perceptrón multicapa (MLP) con algoritmo de aprendizaje supervisado tipo back-propagation. MLP es capaz de producir modelos de predicción de tipos de clasificación y regresión, donde la única diferencia es que la variable de salida es nominal o numérica para las estimaciones de clasificación o regresión. MLP ha demostrado ser un aproximador fuerte de la función para los problemas de la predicción, es decir, dado el tamaño correcto y la estructura, MLP demuestra ser capaz de aprender las relaciones no lineales altamente complejas entre las variables de entrada y de salida (Şen et al. 2012).

## **7 METODOLOGÍA**

### **7.1 Método:**

El método empleado para abordar este problema de investigación es cuantitativo pues busca clasificar a estudiantes que responden al nuevo perfil de estudiante de nivel terciario, recién ingresados a las carreras de una universidad, según características asociadas al riesgo de deserción. Las características permitieron diseñar perfiles de riesgo de deserción a partir de los factores y variables evidenciados empíricamente en el país en las últimas dos décadas. Según muestran Hernández Fernández y Baptista (2006) una característica de los estudios cuantitativos es aumentar la complejidad de sus diseños en función de la evidencia empírica acumulada, empleando ésta como fuente válida.

### **7.2 Diseño:**

El estudio es de tipo no experimental, analítico de tipo ecológico, correlacional y corte transversal, empleando la clasificación de Schoenbach (1999). Al abordar la situación de un grupo puntual en una institución particular, el estudio es de tipo ecológico. Ya que los datos son de un grupo configurado en un momento puntual -inicio de un año académico-, es de corte transversal. En tanto se realizará un análisis secundario de datos contenidos en diversas bases: bases externas pertenecientes al ministerio de educación y ministerio previsión social y bases de datos internas con los datos de los estudiantes ingresados para relacionar estas variables en la configuración de perfiles, el estudio es correlacional.

### **7.3 Población:**

La población corresponde a todos los ingresos al primer año de una institución Universitaria. Considerar todos los ingresos implicó dos consideraciones: antigüedad del dato y traslado. Traslado incluyó los traslados de carrera e institución pues al contrastar los datos con un nuevo grupo normativo es pertinente reevaluar el riesgo en relación al nuevo grupo para priorizar. Al igual se ingresarán los datos con independencia de la antigüedad de los mismos, sin embargo, en función de la literatura, datos académicos de una antigüedad mayor a 5 años serán un indicador de riesgo para dos de los perfiles: académico y social, debido a que mayor edad se asocia a deserción.

### **7.4 Muestreo:**

El estudio puede considerarse un estudio no muestral pues requiere analizar todos los elementos de la población, definida por el grupo de ingreso en un año determinado a una institución universitaria específica. Por lo antes señalado el acercamiento es censal.

### **7.5 Bases de Datos:**

Las bases de datos externas utilizada en el presente estudio serán la base SIES, Simce y DEMRE dependientes del Ministerio de Educación y la base de la Encuesta Casen, dependiente del Ministerio de Desarrollo Social. La base de datos interna empleada será la de cada institución, con los antecedentes de cada estudiante. Su compatibilidad se producirá a partir de la identificación de cada sujeto por su número de cédula de identidad.

### **7.6 Procedimiento:**

Existirán tres perfiles de estudiantes definidos teóricamente tras el análisis de la literatura y la revisión del comportamiento de las variables involucradas en el abandono estudiantil en los estudios chilenos. Se definieran así tres grupos que faciliten la focalización de apoyo en primer año. Se realizará exploración de datos para especificar propiedades, características, y asignar los perfiles a los sujetos que se sometan al análisis. Tras esto, los indicadores asociados a los perfiles permitirán agrupar a los sujetos para describirlos en términos de porcentaje de la población de ingreso, porcentajes de indicadores alcanzados dentro del clúster, porcentaje que han alcanzado en los indicadores más determinantes del riesgo académico y/o social.

El modelo integrará un sistema de Inteligencia de Negocios básico, bajo un marco de referencia estándar, que incluirá un bloque de Minería de Datos (Data Mining), que mediante procesos, técnicas y algoritmos permitirá clasificar y agrupar los sujetos bajo estudio para identificar patrones comunes de comportamiento que entreguen información para la toma de decisiones. Debido a que el análisis se restringe al manejo de la deserción, este estudio no pretende diseñar un sistema de Inteligencia de Negocios con su arquitectura completa, por lo tanto, no es parte del modelo contar con una bodega de datos (Data Warehouse) para el almacenamiento de los datos.

En el ámbito de la Gestión Empresarial, se centrará la toma de decisión a niveles estratégicos, para la correcta orientación y priorización de los programas y proyectos relacionados con el apoyo a estudiantes, con el fin de disminuir la deserción o abandono en el primer año de las universidades que reciben a el “nuevo estudiante universitario”.

Gestionar la deserción implica focalizar, priorizar y adaptar los portafolios, programas y proyectos derivados de la planificación estratégica de la Universidad, con los perfiles de riesgo compuestos por estudiantes previamente clasificados y agrupados. Este alineamiento estratégico permitirá una asignación de recursos eficiente y eficaz, logrando indicadores claves de alto desempeño en las diferentes etapas de su ejecución.

Con la definición de perfiles de riesgo se podría derivar una tasa de riesgo asociada, para -luego de ello- determinar la rentabilidad exigida a la cartera de programas o de proyectos relacionados. Una limitación al estudio, será la ausencia de definición de esta tasa de descuento, que permitiría la evaluación de proyectos de inversión social, desde la perspectiva de evaluación de la calidad de negocios vinculados al desarrollo social y/o con orientación pública. Así, en esta propuesta, se concibe a la Universidad como una organización que integra la gestión empresarial a su cadena de valor, y a sus procesos de negocios claves y de apoyo a nivel estratégico, táctico y operativo, con la finalidad de obtener rentabilidades tanto social como financiera en sus portafolios de programas y proyectos, derivados de la planificación estratégica centrada en el estudiante o cliente.

## **8 DESARROLLO DE LA PROPUESTA**

En este capítulo se llevará a cabo el desarrollo de la propuesta de solución a modo de lograr un modelo que permita clasificar a los estudiantes por perfiles de riesgo, y con esta información se puedan ajustar los programas de apoyo para disminuir la deserción o abandono de los estudiantes de una Universidad masiva.

### **8.1 Teoría de sistemas aplicado al Modelo**

Como primera etapa en el desarrollo de la propuesta, es necesario establecer un modelo de referencia general para el modelo, y que estará basado en la Teoría General de Sistemas. En este modelo general se visualiza a la Universidad como un Sistema con una etapa de entrada, una etapa de procesamiento o transformación, una etapa de salida, y una etapa de realimentación, tal como se muestra en la figura 4. En este tipo de Sistema, la etapa de Entrada (Input) corresponde a los datos brutos de los estudiantes, familiares etc., recolectados y obtenidos a través de las diferentes bases de datos sugeridas, como los son: MINEDUC, SIES, SIMCE. etc. La etapa de Procesamiento (Processing) corresponde al proceso llevado a cabo por el subsistema de inteligencia de negocios (BI). La etapa de Salida (Output) entrega la información acerca de la clasificación de los estudiantes por perfiles de riesgo. La etapa de realimentación (feedback) está compuesta principalmente por los resultados obtenidos en la evaluación social de los proyectos que conforman los programas de apoyo a los estudiantes, así como también la evaluación de los KPI asociados a la deserción.

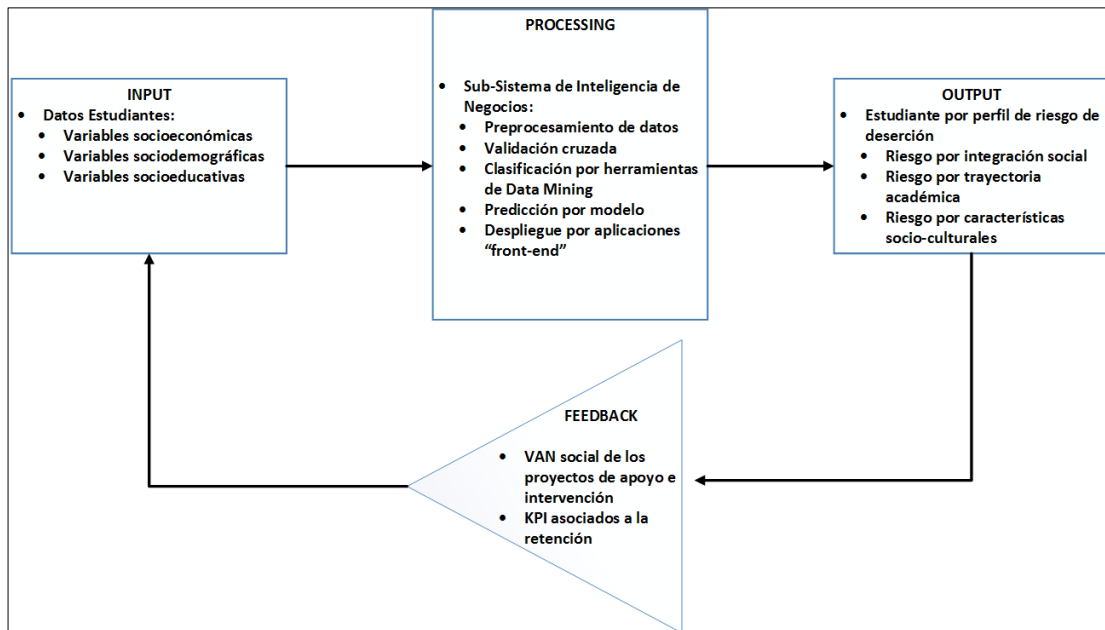


Figura 3: Sistema Referencial del Modelo Predictivo  
Fuente: Elaboración propia

## 8.2 Arquitectura del subsistema BI aplicado al Modelo

En esta sección se introducirá la arquitectura típica para el subsistema de BI aplicado al Modelo Predictivo, principalmente en las etapas de Entrada, Proceso y Salida. La figura 4 muestra las diferentes etapas para estos sistemas BI, con las áreas segmentadas, que se trabajarán en el presente estudio.

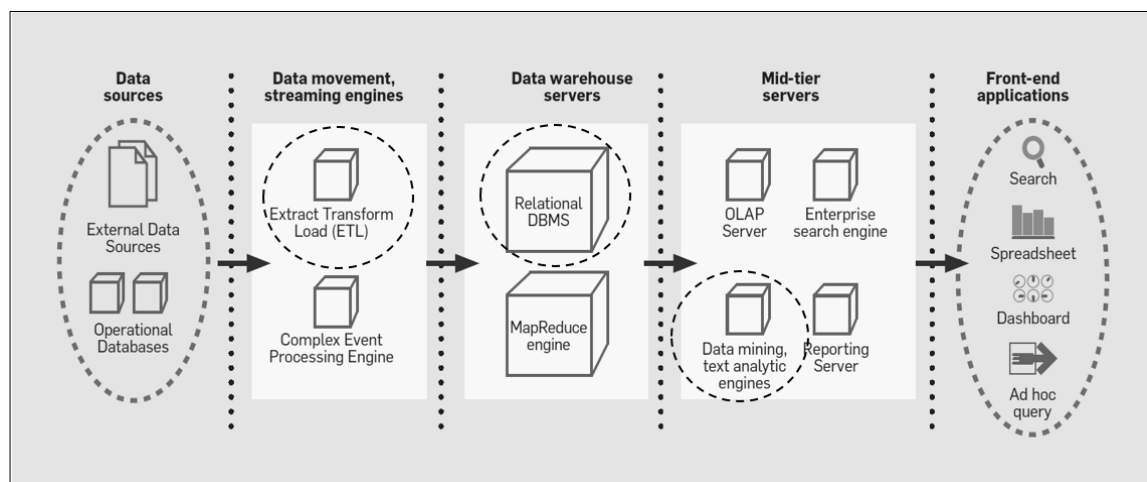


Figura 4: Arquitectura típica de un sistema de inteligencia de negocios

Fuente: Chaudhuri, S. et al (2011)

En general la arquitectura del subsistema deberá contener una etapa de recolección de datos sobre los cuales se llevarán a cabo las tareas de BI. En general estos datos provienen de diferentes fuentes, típicamente desde base de datos operacionales internas de la organización, y también desde base de datos externas a la organización. El recolectar los datos de diferentes fuentes, implican riesgos como por ejemplo datos de variada calidad, inconsistencia, código, formatos distintos, etc. como medida mitigadora, en la segunda parte de la arquitectura, es necesario preparar los datos a través del uso de herramientas de “ETL” (Extract-Transform-Load).

Posteriormente los datos sobre los cuales se desarrollarán las tareas de BI, a menudo son cargados en repositorios llamados Bodegas de Datos (Data Warehouse), y que son gestionados por Sistemas de Gestión de Base de Datos Relacionales (RDBMS). Para el modelo presentado, y a modo de prototipo se diseñará un DATA MART para los datos seleccionados.

Los motores de Minería de Datos (Data Mining) permiten un análisis en profundidad de los datos que van mucho más allá de lo que ofrece OLAP o de los servidores de informes, y proporcionará la capacidad de construir modelos de clasificación para ayudar a responder a preguntas tales como: ¿qué estudiantes están en riesgo de deserción? ¿Qué programas de apoyo son adecuados para el estudiante en riesgo de desertar? ¿Cómo desarrollar la persistencia en estudiantes con perfiles académicos frágiles? ¿Cuáles son las variables asociadas a los estudiantes sin la capacidad de permanecer en el sistema universitario?

### **8.3 Desarrollo del Modelo de Clasificación de perfiles de riesgos de los estudiantes**

#### **8.3.1 Introducción**

El Modelo de Clasificación será construido a partir de la unión de tres elementos a saber: el proceso KDD, el modelo de sistema referencial, y la arquitectura típica del subsistema BI, para formar finalmente el modelo mostrado en la figura 6.

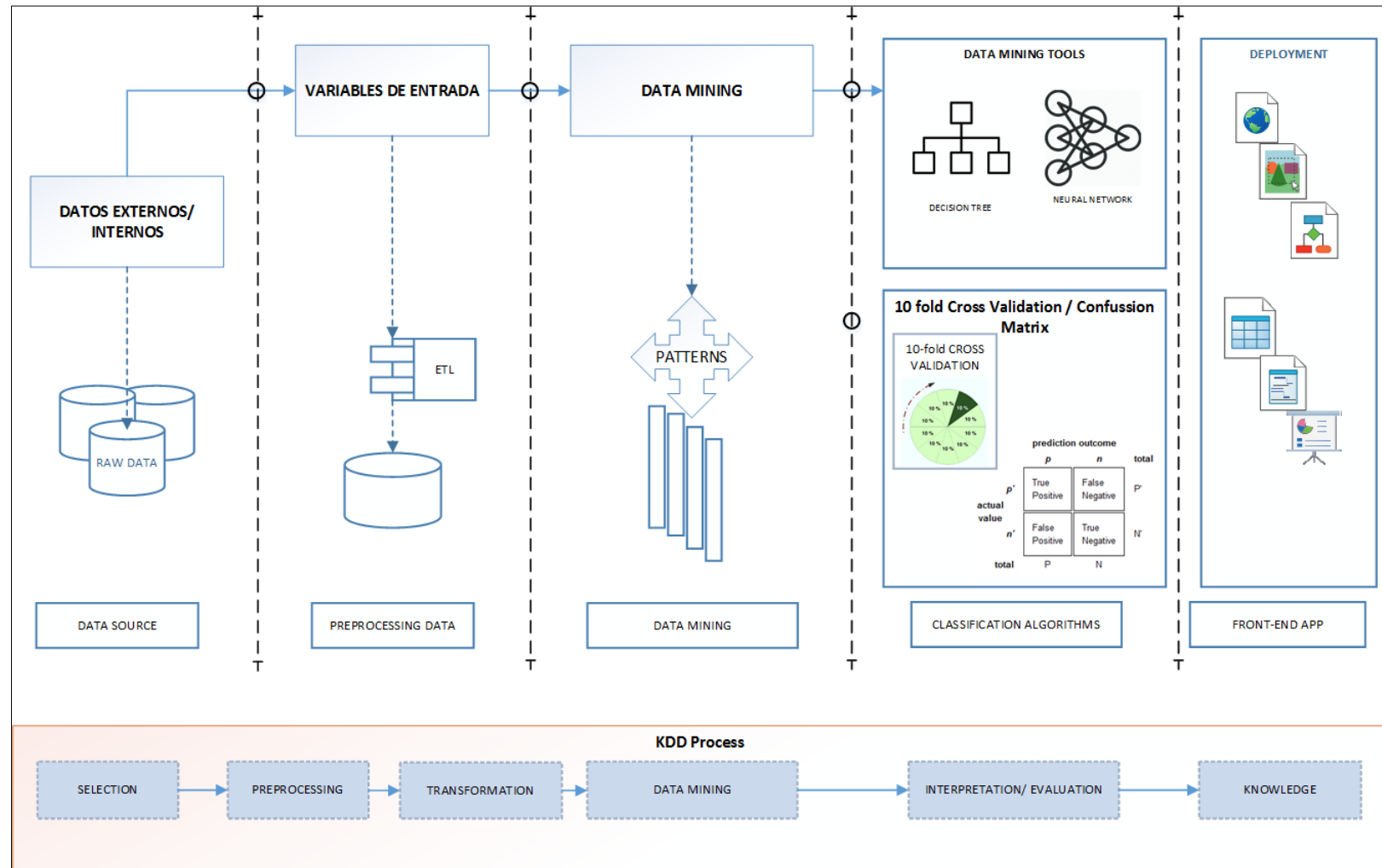


Figura 5: Componentes Modelo BI de Clasificación  
Fuente: Elaboración propia

### **8.3.2 Pasos en la construcción del Modelo**

La construcción del modelo se llevará a cabo en seis pasos que se describirán en el capítulo 4 de Desarrollo de la Propuesta. Los pasos serán los siguientes (Fayyad, U. et al. 1996; Ronald J. Brachman & Anand 1996; Chaudhuri et al. 2011; Han et al. 2012).

Paso 1: Construir el problema de negocios al que se dará solución con un Modelo de Adaptación de la Oferta de Programas de Apoyo Estudiantil a las características particulares de tres grupos, cada uno correspondiente a un perfil de riesgo de abandono: riesgo académico, riesgo social, riesgo socio-cultural. Ello permitiría focalizar la oferta de los programas de apoyo estudiantil a los perfiles de riesgo de y priorizar/adaptar la entrega de los servicios a las características de quienes conforman cada grupo.

Paso 2: Definir un conjunto de datos objetivo, donde se seleccionará un conjunto de datos, o bien se focalizará en un subconjunto de variables o muestras de datos en las que se deberá realizar el descubrimiento para dar solución al problema de negocios planteado en el paso 1.

Paso 3: Limpieza de datos y preprocesamiento, dado que las bases de datos reales generalmente son altamente susceptibles a datos ruidosos, desaparecidos e incoherentes, y esto debido principalmente a su tamaño típicamente de varios gigabytes o más, y su origen probablemente de múltiples fuentes heterogéneas. Por lo tanto, asegurar la calidad de los datos es crucial en esta etapa. La calidad de los datos se define en términos de exactitud, integridad, consistencia, puntualidad, credibilidad e interpretabilidad. (Han et al. 2012)

En este paso se usarán las principales técnicas de pre-procesamiento de datos: limpieza, integración, reducción y transformación de datos, tal como se muestra en la figura 3, que resume esta fase.

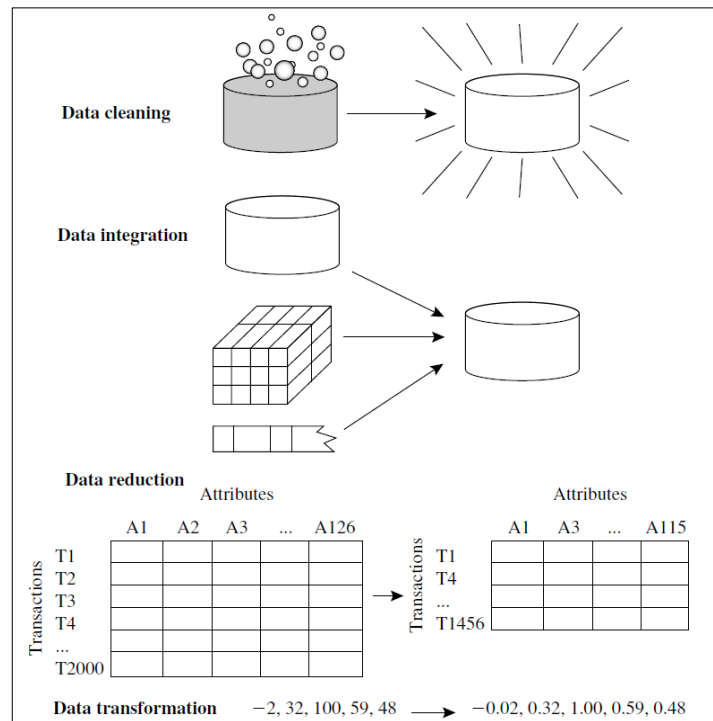


Figura 6: Formas de Pre-procesamiento de Datos

Fuente: (Han et al. 2012)

Paso 4: Minería de Datos (Data Mining). Corresponde al análisis exploratorio y la selección de modelos e hipótesis: la elección del algoritmo de minería de datos y el método o métodos de selección que se utilizarán para buscar patrones de datos. Este proceso incluye la decisión de qué modelos y parámetros pueden ser apropiados, y que coincide con un método particular de minería de datos con los criterios generales del proceso KDD. Además, el buscar patrones de interés en una forma de representación en particular, o un conjunto de tales representaciones, incluyendo reglas de clasificación o árboles, regresión y agrupación.

Paso 5: Interpretación y Evaluación, es interpretar patrones minados, posiblemente regresando a cualquiera de los pasos anteriores para una iteración adicional. Este paso también puede implicar la visualización de los patrones y modelos extraídos o la visualización de los datos dado los modelos extraídos.

Paso 6: Conocimiento Descubierto, está actuando sobre el conocimiento descubierto; usando el conocimiento directamente, incorporando el conocimiento en otro sistema para acciones futuras, o simplemente documentándolo y reportándolo a las partes interesadas. Este proceso también incluye buscar y resolver conflictos potenciales con conocimiento previamente creído (o extraído).

### **8.3.3 *Desarrollo Paso 1: Construcción del problema de negocios al que se dará solución con el Modelo.***

En primer lugar, es necesario la correcta comprensión y contexto del problema o solución de negocio a satisfacer. Este paso es muy importante puesto que permite dimensionar el problema, conocer las estrategias que se han empleado en su solución y los sistemas nacionales e internacionales de abordaje. Esto posibilita delimitar un problema real con implicancias económicas, sociales y culturales, que impactan a la empresa, el estado y las familias. Se idea una solución realizable con el conocimiento, tecnología e información disponible para delimitar objetivos y alcance de la propuesta.

En el capítulo 2 se detalla el conocimiento generado alrededor del problema, su valoración desde el ámbito de la gestión y la administración en la educación superior. Las altas tasas de abandono estudiantil son de gran impacto en el bienestar financiero de la Universidades, puesto que los retornos de los programas y proyectos de retención son muy bajos, y en algunos casos negativos. Además, las tasas de abandono o deserción, impactan directamente en los indicadores de calidad de la Universidades.

Para la construcción del problema o solución de negocio, se utilizó como base referencial un marco de trabajo integrado de análisis y desarrollo de un Sistema de Información Gerencial de apoyo a la toma de decisiones, como el que se muestra en la figura 6 y es necesario su entendimiento puesto que los Sistemas de Información crearán valor para la Universidad, en la medida que sea una solución organizacional y gerencial para los desafíos impuestos por el entorno. Este marco, desde el punto de vista del negocio, muestra al Sistema de Información interactuando con las dimensiones de administración, organización, y tecnología, para apoyar las funciones organizacionales y la toma de decisiones, que se conjugan con la solución y desafíos del negocio.

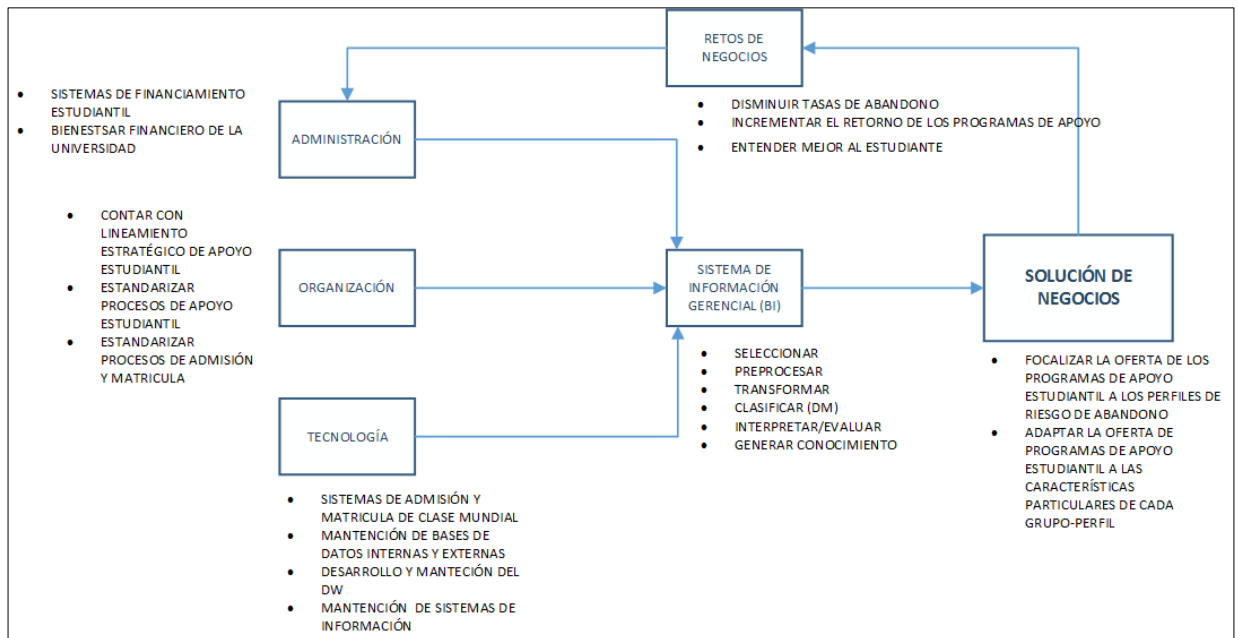


Figura 7: Marco referencial de diseño y análisis de un Sistema de Información

Fuente: Elaboración propia

El modelo a construir y su integración al marco referencial de análisis y desarrollo mostrado en la figura 6, deberá ser considerado como trabajo futuro para que llegue a ser parte de un Sistema de Información Gerencial mayor. Por ahora el Modelo deberá dar cuenta del cómo lograr, a través del proceso KDD, y principalmente de la tecnología de Minería de Datos, la generación de conocimiento para desarrollar la solución de negocios planteada.

Como solución de negocios se plantea: focalizar la oferta de los programas de apoyo estudiantil a los perfiles de riesgo de abandono, que permitirá a los encargados de la gestión de programas y proyectos de apoyos, llevar cabo tareas de priorización, dirección y gestión de recursos de forma más eficaz y eficiente. También adaptar la oferta de programas de apoyo estudiantil a las características particulares de tres grupos diferenciado por perfiles, lo que permitirá principalmente incrementar los retornos de estos programas de apoyo.

La literatura internacional da cuenta de que el abandono de la universidad es un problema multicausal, por ello pueden coexistir diversos modelos teóricos que permiten abordar la situación. Cada uno de los modelos se focaliza o enfatiza uno de los factores que se ha demostrado que influyen en el fenómeno: personales-familiares, académicos, institucionales, económicos, sociales-culturales. se analizó algunas estrategias de enfrentamiento que dan cuenta de estudiantes que típicamente presentan problemas de

inserción en el medio Universitario. A partir de ello se elaboró la definición de los perfiles de riesgo.

La primera definición fue de una construcción teórica que diferencie tres grupos de estudiantes. La segunda definición fue del grupo de variables que darían forma a cada perfil. En este primer momento se presenta la construcción conceptual que se relaciona a la literatura abordada.

### **8.3.4 Definición Conceptual de Perfiles de Riesgo de Deserción**

#### **8.3.4.1 Riesgo de Deserción por Integración Social:**

Define a los estudiantes que enfrentan problemas especiales para relacionarse y establecer vínculos sociales y académicos en la gran mayoría de las instituciones. La característica central es sentirse marginados de la vida institucional. Esto sucede con grupos minoritarios, adultos (Boshier 1973), personas con necesidades educativas especiales y sujetos de extracciones culturales muy diferenciadas de la población a la que la institución presta servicios. En este grupo la deserción es probablemente voluntaria y más frecuente en los primeros años de la carrera, debido a que se les dificulta establecer contacto con la comunidad social e intelectual de la institución y lograr amistades en ella, relaciones que tienden a realizarse en la etapa temprana de la carrera (Tinto 1989).

Para la persistencia de este grupo, son relevantes las intervenciones de integración social y académica que se originan en la interacción con otros estudiantes y miembros del personal docente, realizadas en el primer año.

#### **8.3.4.2 Riesgo de Deserción por Trayectoria Académica:**

Define a los estudiantes que enfrentan problemas especiales para integrarse a las exigencias de los programas formativos. La literatura nacional da cuenta de problemas para adaptarse al ritmo, exigencias y trabajo sistemático de lo académico: asistencia a clase, lecturas, estudio. Los problemas habituales refieren a hábitos de estudio, aprendizaje autónomo, capacidad reflexiva, relacionada a las exigencias y calidad de las escuelas de origen. En este grupo la deserción es más frecuente en periodos de transición y en este momento las instituciones pueden actuar con eficacia para prevenir el abandono temprano.

Medidas institucionales relativamente sencillas pueden producir efectos inmediatos y duraderos en la retención estudiantil. Emplear alumnos de los últimos años como consejeros, proporcionar sesiones de asesoramiento y orientación tempranas, promover la formación de

grupos de estudio y establecer tutores académicos para conjuntos de nuevos estudiantes, puede ayudarlos a adaptarse a la vida universitaria (Tinto 1989).

#### *8.3.4.3 Riesgo de Deserción por Características Socio-Culturales*

Define a los estudiantes que iniciaron un proceso de adquisición y adaptación a nuevos códigos culturales y simbólicos (de los Ríos & Canales 2007). Desconocen la vida universitaria y sus requerimientos, lo que produce un desajuste de expectativas que se entremezcla con dificultades académicas, emocionales, sociales y económicas. Esto les deja en un estado de vulnerabilidad, cuestionan las habilidades y competencias académicas que poseen fluctuando entre la valoración extremas a éxitos y fracasos (Canales & De los Ríos 2009).

Sus dificultades económicas se relacionan a la situación laboral frágil por bajos salarios e inestabilidad laboral de los adultos con los que conviven. Habitualmente trabajan para solventar parte de los gastos. Sus condiciones de vida (hacinamiento, delincuencia, horas de viaje para estudiar) dificultan su progresión formativa (Canales & De los Ríos 2009).

Los soportes que encuentran en la Universidad son: ayudas financieras (créditos y becas), servicios médicos, psicológicos y dentales, recursos docentes y de infraestructura (lugares de estudio, bibliotecas y recursos computacionales).

Estos estudiantes requieren soporte afectivo y contención pues éstos disminuyen el estrés asociado a los “fracasos” académicos y evitan valoraciones extremas de sus éxitos y fracasos.

La motivación se produce cuando tienen propósitos claros: anticipan beneficios asociados a la obtención del título profesional, aspiran al bienestar y calidad de vida (Canales & De los Ríos 2009).

#### ***8.3.5 Desarrollo Paso 2: Definición de las fuentes de datos internas y externas. Comprensión y conocimiento de los datos a usar en el modelo de clasificación.***

Las fuentes de datos internas corresponden a las bases de datos operacionales de la Universidad, y que almacenan los datos de sus estudiantes de dimensión tipo académica, personal, y económica principalmente. Las bases de datos de dimensión personal y económica generalmente son alimentadas por las unidades de admisión y matrícula de la Universidad. En tanto las bases de datos de dimensión académica son alimentadas por las unidades académicas de las diferentes facultades y/o escuelas.

Las fuentes de datos externas, corresponden a los datos obtenidos a través de los portales de las organizaciones que los mantienen. Entre ellas están:

- Servicio de Información de Educación Superior (SIES) disponible en portales de servicios como Mi Futuro y el Consejo Nacional de Educación (CNEd).
- Sistema de Medición de la Calidad de la Educación (SIMCE) disponible a través del sitio de la Agencia de Calidad de la Educación.
- Bases de datos PSU del DEMRE, solicitable a través de su portal de investigadores.

Las fuentes internas y externas se representan en la siguiente figura:

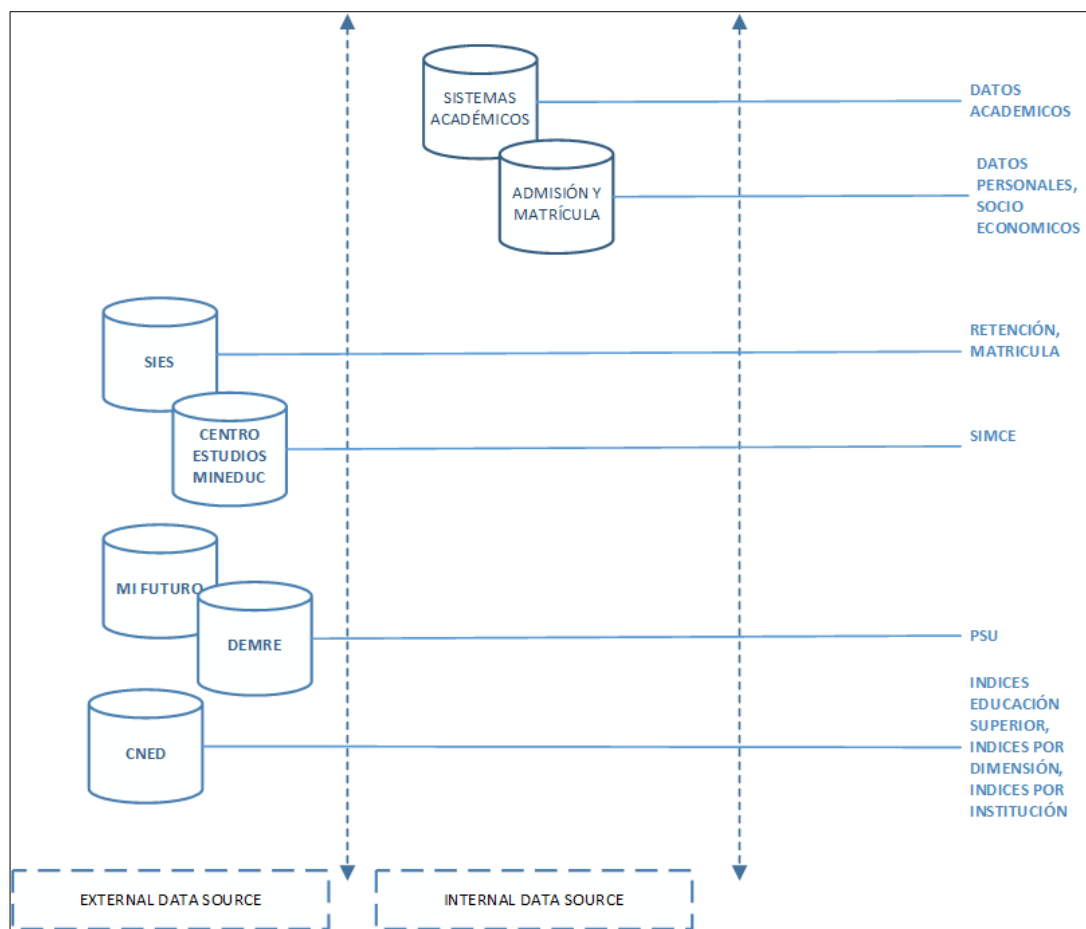


Figura 8: Fuentes de Datos Externas/Internas  
Fuente: Elaboración propia

#### 8.3.5.1 *Objetos de datos y tipos de atributos (Data Objects and Attribute Types)*

Es necesario en este paso definir los Data Objects del Modelo, en donde un Data Object representa una entidad, que, para el caso de una base de datos en una Universidad, los objetos podrían ser estudiantes, profesores, y cursos. Los Data Objects son descritos por sus Atributos. Si los data objects son almacenados en una base de datos, entonces ellos se convierten en tuplas. Esto es, las filas de la base de datos corresponden a los Data Objects, y las columnas son sus atributos.

Por otra parte, un atributo es el campo de datos que representa una característica, rasgo, o propiedad del data object. Los atributos que describen un objeto de estudiante pueden incluir, por ejemplo, dirección, edad, NSE, tipo de establecimiento de origen, puntaje PSU, etc. Los valores observados para un atributo dado se conocen como observaciones. Un conjunto de atributos utilizados para describir un objeto dado se denomina vector de atributo (o vector de entidad).

Un atributo asume valores que al transformarse en información son llamados datos. Estos datos varían en valor y asumen una distribución del atributo (o variable). Si el objeto o relación se establece en una función de un atributo se denomina univariable, una distribución bivariada implica dos atributos, y habitualmente se habla de una multivariada cuando hay más de tres.

El tipo de un atributo está determinado por el conjunto de posibles valores, (categorial o numérico, nominal u ordinal, binario, etc.) que el atributo puede tener. En el anexo 3 se muestra la clasificación según atributos de los data objects.

#### 8.3.5.2 *Atributos Continuos versus Discretos*

Los algoritmos de clasificación desarrollados a partir del campo del aprendizaje automático a menudo hablan de atributos como discretos o continuos. Cada tipo puede ser procesado de manera distinta.

Un atributo discreto tiene un conjunto finito o infinito de valores contables, que pueden o no representarse como números enteros. Los atributos edad, raza, examen psu, e ingreso familiar tienen un número finito de valores, por lo que son discretos. Tenga en cuenta que los atributos discretos pueden tener valores numéricos, como 0 y 1 para atributos binarios o, los valores 0 a 110 para el atributo edad.

Un atributo es infinito contable si el conjunto de valores posibles es infinito pero los valores pueden ser puestos en una correspondencia uno a uno con números naturales. Por ejemplo, el ID de estudiante de atributo es infinito contable. El número de estudiantes puede crecer hasta el infinito, pero en realidad, el conjunto real de valores es contable (donde los valores se pueden poner en correspondencia uno a uno con el conjunto de enteros). Los códigos postales de la dirección son otro ejemplo.

Si un atributo no es discreto, es continuo. Los términos atributo numérico y atributo continuo se usan a menudo indistintamente en la literatura. (Esto puede ser confuso porque, en el sentido clásico, los valores continuos son números reales, mientras que los valores numéricos pueden ser enteros o números reales). En la práctica, los valores reales se representan utilizando un número finito de dígitos. Los atributos continuos se representan normalmente como variables de punto flotante.

### *8.3.5.3 Definición de los tipos de atributos de los Objetos Perfiles de Riesgo de Deserción*

Tras tener una construcción teórica que daba cuenta de un objeto que definía a los estudiantes en riesgo de deserción (perfiles) se analizó los factores que influyen en la deserción según la literatura internacional. Luego se observó el valor que las variables que componen estos factores asumieron en los estudios nacionales (ver anexo 1), para definir empíricamente el tipo de atributos de cada perfil. Estos atributos se presentan a continuación.

#### *8.3.5.3.1 Atributos del Perfil Riesgo por Integración Social:*

Está compuesto por los atributos que pongan en la condición de distinto-extraño al grupo mayoritario con frágiles recursos académicos.

Su operacionalización como data Object se presenta en el anexo 4

#### *8.3.5.3.2 Riesgo por Trayectoria Académica*

Está compuesto por los atributos que ponen en condición de desventaja para adaptarse al ritmo y exigencia académica.

Su operacionalización como data Object se presenta en el anexo 5.

### 8.3.5.3.3 Riesgo por Vulnerabilidad Socio-cultural

Está compuesto por los atributos que sitúen al estudiante en condición de vulnerabilidad por no haber tenido condiciones que permitan desarrollar la capacidad de aprender y organizar redes sociales de apoyo. Su carencia se vincula a la insuficiencia de símbolos asociados a la vida universitaria.

Su operacionalización como data Object se presenta en el anexo 6.

## 8.3.6 *Desarrollo Paso 3: Pre Procesamiento y Transformación de los Datos (Data Preprocessing and Data Transformation)*

### 8.3.6.1 *Limpieza y Selección de Datos (Data Cleaning)*

En esta etapa se tiene por objetivos: seleccionar los datos relevantes y eliminar o al menos reducir ruidos, tales como: ingresos erróneos, información incompleta, outliers, inconsistencias. Generalmente los algoritmos de Minería de Datos se aplican sobre tablas planas, conocidas como tablas de contingencia o cubos multidimensionales.

Las rutinas de limpieza de datos trabajan para "limpiar" los datos, rellendo los valores faltantes, suavizando datos ruidosos, identificando o eliminando valores atípicos, y resolviendo inconsistencias. Si los usuarios creen que los datos están sucios, es poco probable que confíen en los resultados de cualquier minería de datos que se haya aplicado. Además, los datos sucios pueden causar confusión para el procedimiento de minería, lo que resulta en una salida poco confiable. Aunque la mayoría de las rutinas de minería tienen algunos procedimientos para tratar datos incompletos o ruidosos, no siempre son robustos. En su lugar, pueden concentrarse en evitar el sobreajuste de los datos a la función que se está modelando. Por lo tanto, un paso de preprocesamiento útil es ejecutar sus datos a través de algunas rutinas de limpieza de datos.

Para el caso de los registros incompletos las posibilidades son las siguientes:

- ignorar el registro
- poner un dato manualmente
- poner un ID global
- completar con valor significativo, por ejemplo, valor medio del atributo
- adivinar el valor más probable para el dato, por ejemplo, valor del atributo para un registro similar

- ignorar la tupla: método poco efectivo a menos que falten muchos atributos en la misma fila
- llenar los valores manualmente: no es practicable cuando el set de datos presenta muchos valores faltantes.
- usar la media del atributo
- usar la media por clases
- usar el valor más probable

Para el caso de Datos ruidosos: en ruidos se puede suavizar las variaciones de los datos, por ejemplo, discretizando en bins, y el uso de técnicas de clustering para identificar outliers. en el caso de Inconsistencias: chequear datos manualmente, realizar test cruzados, chequear rangos.

#### 8.3.6.2 Integración de Datos (*Data Integration*):

Al tener varias fuentes de datos para el Modelo, implicaría integrar múltiples bases de datos, cubos de datos o archivos (es decir, integración de datos). Sin embargo, algunos atributos que representan un concepto dado pueden tener nombres diferentes en diferentes bases de datos, causando inconsistencias y redundancias. Por ejemplo, el atributo de identificación del estudiante puede ser referido como *student\_id* en una fuente de datos, y como *std\_id* en otra.

Tener una gran cantidad de datos redundantes puede ralentizar o confundir el proceso de descubrimiento de conocimiento. Claramente, además de la limpieza de los datos, se deben tomar medidas para evitar las redundancias durante la integración de los datos. Normalmente, la limpieza de datos y la integración de datos se realizan como un paso de preprocesamiento cuando se preparan datos para un almacén de datos. La limpieza de datos adicionales se puede realizar para detectar y eliminar redundancias que pueden haber resultado de la integración de datos. La integración se encarga de: reunir los datos, remover inconsistencias, reducir redundancias, de diferentes representaciones.

#### 8.3.6.3 Reducción de Datos (*Data Reduction*):

La reducción de datos obtiene una representación reducida del conjunto de datos que es mucho menor en volumen, pero produce los mismos resultados analíticos (o casi los mismos). Las estrategias de reducción de datos incluyen la reducción de la dimensionalidad y la reducción de la numeridad. La reducción de datos se encarga de encontrar un sub-espacio tal que la exactitud de las herramientas de data mining es igual o mejor que al usar el espacio original. De esta forma el beneficio neto es de mayor eficiencia con, al menos, la misma exactitud.

en un espacio multidimensional, de muchos atributos, no todos son relevantes para cierta tarea. Además, existen redundancias no aparentes, cuya identificación permite simplificar la aplicación de las herramientas de data mining. Entre las técnicas para reducir dimensionalidad se encuentran:

- resumir en forma agregada algunas dimensiones
- eliminar dimensiones irrelevantes o redundantes
- proyectar datos a sub-espacios
- tomar muestras
- encontrar modelos paramétricos o no paramétricos que permitan representar los datos en forma resumida

#### 8.3.6.4 *Transformación y Discretización de Datos (Data Transformation and Data Discretization)*

para lograr un mejor rendimiento de las herramientas de Minería de Datos, la mayor parte de las veces es necesario transformar los datos, puesto que algunos algoritmos:

- necesitan normalizar datos a mismos rangos numéricos
- sólo operan con datos continuos, mientras que otros sólo con datos discretos.
- no toleran ruidos, especialmente outliers
- necesitan trabajar con datos previamente agregados para reducir complejidad computacional.

los datos en esta etapa son transformados y consolidados de tal forma que quedan listos para los procesos de minería de datos.

la transformación puede ser llevado a cabo por los siguientes procesos:

- *Smoothing*: roll up, binning, regresión, clustering
- *Normalización*: se modifica la escala de los atributos de tal forma que todos los valores queden dentro de un rango específico, (ejemplo: 0 y 1, -1 y 1)
- *Construcción de características*: nuevos atributos son construidos desde el mismo set de datos para mejorar el proceso de data mining.

#### 8.3.7 *Desarrollo Paso 4: Minería de Datos (Data Mining)*

En este estudio, se usan dos métodos populares de predicción / clasificación (y comparados entre sí): redes neuronales artificiales y árboles de decisión. Estos métodos de predicción se seleccionan debido a su capacidad superior de modelar los problemas de

predicción del tipo de clasificación y su popularidad en la literatura de data mining publicada recientemente.

#### 8.3.7.1 Árboles de Decisión (*Decision Tree*):

Los árboles de decisión son poderosos algoritmos de clasificación que son cada vez más populares debido a sus características de explicación intuitiva. Los árboles de decisión más populares son ID3, C4.5, C5, CART (Classification and Regression Trees), y CHAID (CHi-squared Automatic Interaction Detector).

Para el caso de este estudio se usará el algoritmo C5, dado que es un clasificador que clasifica los datos en menos tiempo que otros clasificadores. Para generar el árbol de decisión, el uso de memoria es mínimo y también mejora la precisión, además proporciona selección de características, validación cruzada y facilidades reducidas de poda de errores. Por lo tanto, el alcance adicional de este algoritmo se logra mediante la implementación de nuevas características como la validación cruzada y la complejidad del modelo. Al implementar todas las características del algoritmo usando paquetes weka, se mejora la precisión en la clasificación (Pandya 2015).

El nodo inicial del árbol de decisión se llama nodo raíz, los nodos intermedios del árbol se llaman nodos hoja y el último nodo del árbol se denomina nodo final. Las variables en cada nodo del árbol de decisión se ponen a prueba con el conjunto de datos de formación utilizando un algoritmo de separación. Comenzando con la pregunta de qué variable debe probarse en la raíz del árbol, el algoritmo del árbol de decisión construye el árbol desde la base (de hecho, los árboles de decisión generalmente se representan como un árbol invertido donde la raíz está en la parte superior, Por lo tanto, también podría ser llamado de arriba a abajo). El número de ramas a crear para cada nodo depende del algoritmo específico utilizado y del número de valores de la variable seleccionada.

Por ejemplo, para modelar el abandono o deserción de un estudiante, el proceso de modelado se podrá iniciar estableciendo la relación entre los abandonos y los factores de riesgos como, por ejemplo: el género, la asistencia, el grado previo al semestre, la educación de los padres, los ingresos de los padres, becas, si es primer hijo y si el estudiante está trabajando o no.

Luego para este estudio se sugiere desarrollar un método híbrido: donde el proceso de modelado es una combinación de técnicas de árbol de decisión y regresión logística. En primer lugar, el algoritmo del árbol de decisiones se utiliza para identificar los factores que

influyen en los abandonos. Una vez identificados los factores, se utiliza la técnica de regresión logística para cuantificar los abandonos y el efecto de cada factor de riesgo.

La siguiente tabla podría ser una muestra para el diseño del árbol de decisión:

Tabla 1: Ejemplo de variables de muestra

| Variable           | Tipo de Variable | Descripción           |
|--------------------|------------------|-----------------------|
| Genero             | Nominal          | Masculino, Femenino   |
| Prev NEM           | Nominal          | 1.....7               |
| ParentEd           | Nominal          | Educado/no educado    |
| ParentIncome       | Nominal          | Bajo, medio, alto     |
| Beca               | Nominal          | Obtenida/ no obtenida |
| PrimerHijo en la U | Nominal          | Verdadero / falso     |
| workEst            | Nominal          | Si / No               |

Luego el árbol de decisión podrá ser el siguiente, mostrado en la figura 9, en donde el ingreso de los padres es un factor de riesgo importante para la deserción. El bajo nivel de ingresos tiene una mayor influencia que los niveles medios y altos. El hecho de que el estudiante sea el primer hijo de la familia en ingresar a la educación superior también influye en la función de abandono.

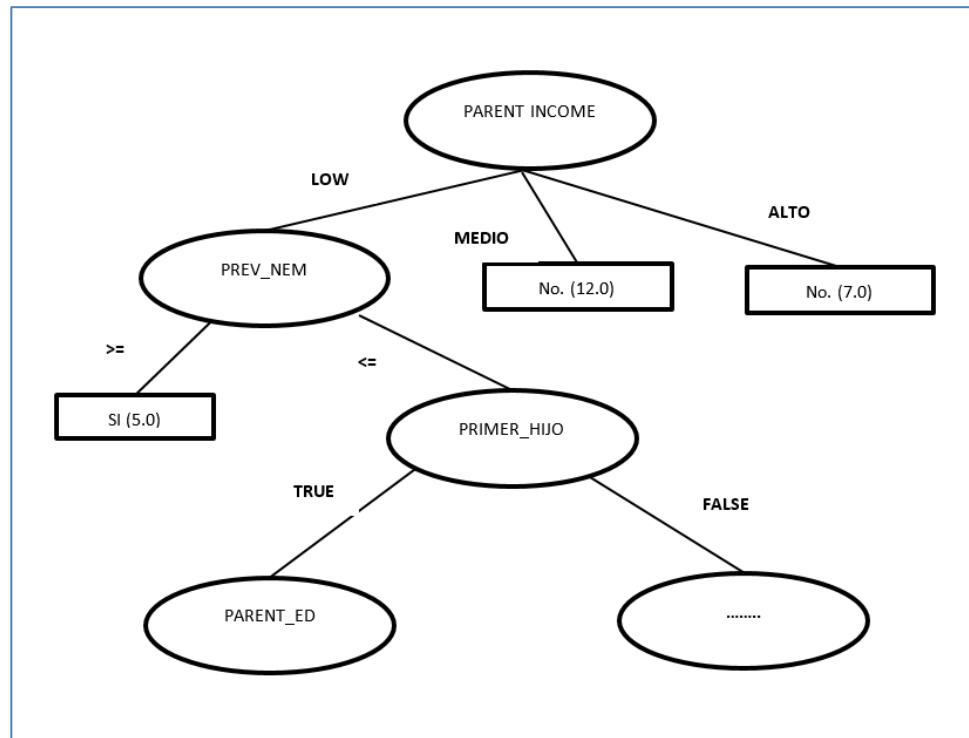


Figura 9: Árbol de decisión para perfil de riesgo académico  
(Fuente: elaboración propia)

Para encontrar una manera óptima de clasificar un conjunto de aprendizaje, lo que necesitamos realizar es minimizar las preguntas que se hacen (es decir, minimizar la profundidad del árbol). Por lo tanto, necesitamos alguna función que pueda medir qué preguntas proporcionan la división más equilibrada. La métrica de ganancia de información es esa función.

Dada una tabla de datos que contiene los atributos y la clase de los atributos, podemos medir la homogeneidad (o heterogeneidad) de la tabla basada en las clases. Decimos que una tabla es pura o homogénea si contiene solamente una sola clase. Si una tabla de datos contiene varias clases, entonces decimos que la tabla es impura o heterogénea. Hay varios índices para medir el grado de impureza cuantitativamente. Los índices más conocidos para medir el grado de impureza son la entropía, el índice de Gini y el error de clasificación.

La entropía de una tabla pura (consta de una sola clase) es cero porque la probabilidad es 1 y  $\log(1) = 0$ . La entropía alcanza el valor máximo cuando todas las clases de la tabla tienen probabilidad igual.

$$Entropía = \sum_j -P_j \log_2 P_j$$

El índice de Gini de una tabla pura consta de una sola clase es cero porque la probabilidad es 1 y  $1-1^2 = 0$ . Al igual que la entropía, el índice de Gini también alcanza el valor máximo cuando todas las clases de la tabla tienen probabilidad igual.

$$\text{Índice de Gini} = 1 - \sum_j P_j^2$$

Similar al índice de entropía e índice de Gini, el índice de error de clasificación de una tabla pura (consta de una sola clase) es cero porque la probabilidad es 1 y  $1-\max(1) = 0$ . El valor del índice de error de clasificación siempre está entre 0 y 1.

$$\text{Error de Clasificación} = 1 - \max\{P_j\}$$

De hecho, el índice máximo de Gini para un número dado de clases es siempre igual al máximo del índice de error de clasificación porque para un número de clases  $n$ , ponemos la probabilidad es igual a:

$$p = \frac{1}{n}$$

Y el máximo índice de Gini ocurre a:

$$1 - n \frac{1}{n^2} = 1 - \frac{1}{n}$$

Mientras que el índice máximo de error de clasificación también ocurre en:

$$1 - \max\left\{\frac{1}{n}\right\} = 1 - \frac{1}{n}$$

Para determinar el mejor atributo para un nodo en particular en el árbol, usamos la medida denominada Ganancia de información. La ganancia de información, Ganancia ( $S, A$ ) de un atributo  $A$ , con respecto a una colección de ejemplos  $S$ , se define como

$$\text{Ganancia}(S, A) = \text{Entropía}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \text{Entropía}(S_v)$$

En donde los Valores ( $A$ ) son el conjunto de todos los valores posibles para el atributo  $A$ , y  $S_v$  es el subconjunto de  $S$  para el cual el atributo  $A$  tiene el valor  $v$  (es decir,  $S_v = \{s \in S \mid A(s) = v\}$ ). El primer término de la ecuación para la Ganancia es sólo la entropía de la colección original  $S$ , y el segundo término es el valor esperado de la entropía después de que  $S$  se divide usando el atributo  $A$ . La entropía esperada descrita por este segundo término es simplemente la suma de la Entropías de cada uno, ponderado por la fracción:

$$\frac{|S_v|}{|S|}$$

que pertenecen a la Ganancia ( $S, A$ ) es, por tanto, la reducción esperada de la entropía causada por conocer el valor del atributo  $A$ .

$$Split\ Information(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

Luego la

$$Razón\ de\ Ganancia(S, A) = \frac{Ganancia(S, A)}{Split\ Information(S, A)}$$

El proceso de selección de un nuevo atributo y particionamiento de los ejemplos de entrenamiento se repite ahora para cada nodo descendente no terminal. Atributos que se han incorporado más arriba en el árbol se excluyen, por lo que cualquier atributo puede aparecer como máximo una vez a lo largo de cualquier camino a través del árbol. Este proceso continúa para cada nuevo nodo de hoja hasta que se cumpla cualquiera de dos condiciones:

1. Cada atributo ya se ha incluido a lo largo de este camino a través del árbol, o
2. Los ejemplos de entrenamiento asociados con este nodo hoja tienen el mismo valor de atributo objetivo (es decir, su entropía es cero).

A continuación, se muestra un ejemplo a un algoritmo de un árbol de decisión

- Create a root node for the tree
- If all examples are positive, Return the single-node tree Root, with label = +.
- If all examples are negative, Return the single-node tree Root, with label = -.
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.
- Otherwise Begin
  - A = The Attribute that best classifies examples.
  - Decision Tree attribute for Root = A.
  - For each possible value,  $v_i$ , of A,
    - Add a new tree branch below Root, corresponding to the test  $A = v_i$ .
    - Let  $\text{Examples}(v_i)$  be the subset of examples that have the value  $v_i$  for A
    - If  $\text{Examples}(v_i)$  is empty
      - Then below this new branch add a leaf node with label = most common target value in the examples
    - Else below this new branch add the subtree ID3 ( $\text{Examples}(v_i)$ , Target\_Attribute, Attributes – {A})
- End
- Return Root

#### 8.3.7.2 Redes Neuronales Artificiales (Artificial Neural Networks, ANN)

Las Redes Neuronales Artificiales son técnicas de análisis inspiradas biológicamente, capaces de modelar funciones no lineales extremadamente complejas.

En este estudio se usará una popular arquitectura de red neuronal llamada Perceptrón multi capa (multi-layer perceptron, MLP), con propagación hacia atrás, y algoritmo de aprendizaje supervisado. Esta arquitectura es una de las más estudiada, y una de las más fuertes en funciones de predicción y clasificación.

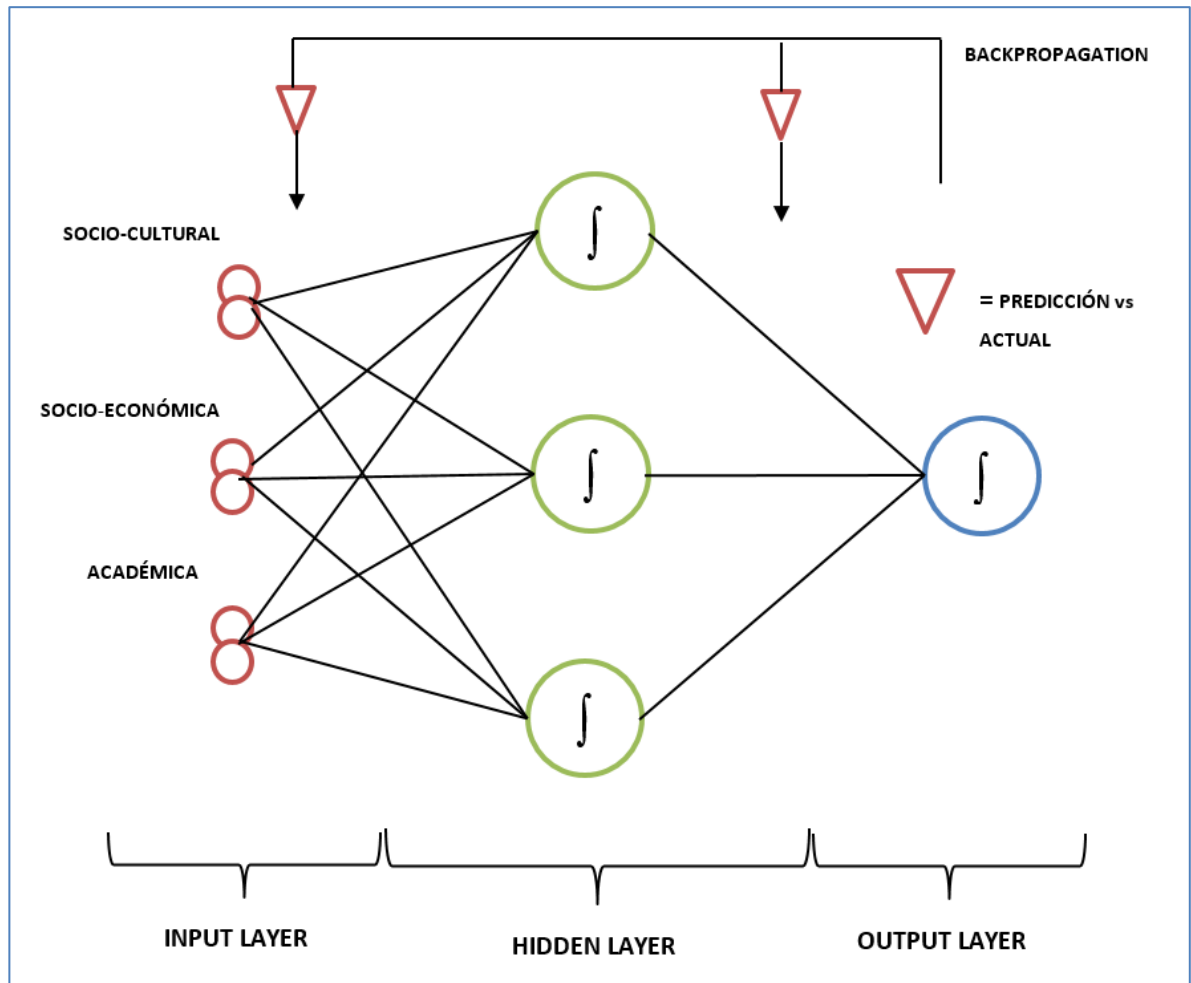


Figura 10: Arquitectura de Red Neuronal usada en este estudio  
Fuente: elaboración propia

Las neuronas (es decir, los elementos de procesamiento) en estas capas están completamente interconectadas desde la entrada a la salida. El número de neuronas en la capa de entrada depende del tipo y el número de variables de entrada en el conjunto de datos. El número de neuronas en la capa de salida depende del número de etiquetas de clase asociadas con el problema de clasificación que se está estudiando (tres en este caso). Los números de las capas ocultas y el número de neuronas en cada capa oculta se determinan mediante un proceso de experimentación iterativo (en este estudio se usaron 10 neuronas y 8 neuronas respectivamente para las capas ocultas 1 y 2).

### 8.3.8 Desarrollo paso 5: Evaluación e interpretación:

Con el fin de minimizar el sesgo asociado con el muestreo aleatorio de las muestras de datos de entrenamiento y retención en la comparación de la precisión predictiva de dos o más métodos, una práctica común es utilizar la validación cruzada k veces (k-fold cross-validation). En la validación cruzada k-fold, el conjunto de datos completo es aleatoriamente (a menudo estratificado en la variable de salida) dividido en k subconjuntos mutuamente excluyentes de aproximadamente el mismo tamaño. El modelo de clasificación es entrenado y probado k veces; Cada vez que es entrenado en todos menos en un fold y probado en el único fold restante. Los resultados de la prueba de todos los fold se compilan en una matriz de confusión. La estimación de la precisión global de la validación cruzada se calcula simplemente como la media de las k medidas de precisión individual como en la siguiente ecuación:

$$CV = \frac{1}{k} \sum_{i=1}^k A_i$$

Donde CV significa exactitud de la validación cruzada, k es el número de veces utilizados, y A es la medida de exactitud de cada vez. En este estudio, se establece el valor de k a 10, que pasa a ser una práctica muy común en muchos estudios de minería de datos. Para comparar los modelos de predicción, se deberá utilizar un criterio de rendimiento común: exactitud de predicción que es la proporción de clasificar correctamente los casos en todos los casos utilizados para fines de prueba.

También se deberá llevar a cabo un análisis de sensibilidad. El análisis de sensibilidad es un proceso iterativo mediante el cual se pueden evaluar las relaciones de causa y efecto entre la variable independiente y la variable dependiente. En el proceso de realizar el análisis de sensibilidad, después de que el modelo es entrenado a un nivel satisfactorio, el "aprendizaje" (es decir, el proceso de desarrollo del modelo) está deshabilitado para que los parámetros del modelo no se vean afectados. La idea fundamental detrás del análisis de sensibilidad es medir el nivel de degradación que puede ocurrir en el desempeño del predictor si no se incluye una variable independiente en el modelo. Por lo tanto, la medida de sensibilidad de una variable independiente específica, es la relación del error del modelo de predicción sin la variable independiente con el error del modelo que incluye esta variable predictora. Cuanto más sensible es el modelo a una variable particular, mayor será la disminución del rendimiento en ausencia de esa variable y, por tanto, mayor será la relación de importancia. Esta medida de sensibilidad se aplica a todas las técnicas de aprendizaje

automático, incluyendo los dos tipos de modelos utilizados en este estudio. La fórmula específica para esta medida de sensibilidad se da en la siguiente ecuación:

$$Si = \frac{Vi}{V(Ft)} = \frac{V(E(Ft|Xi))}{V(Ft)}$$

Donde V (Ft) es la varianza de salida incondicional. En el numerador, el operador de esperanza E llama a una integral sobre Xi; Es decir, sobre todas las variables de entrada, pero Xi, luego el operador de varianza V implica una integral adicional sobre Xi. La importancia de la variable se calcula entonces como la sensibilidad normalizada.

La generación o descubrimiento de conocimiento a través del modelo presentado en este estudio, podría ser utilizado para la implementación de programas de apoyo, como por ejemplo: “Iniciativas para una graduación exitosa”, en la cual su meta podría ser: “aumentar la graduación a tiempo a una tasa de un 5% por año”, y en donde algunos indicadores para el control de gestión, podrían ser:

- Identificar a los estudiantes de primer año que estén en riesgo de abandonar antes de comenzar sus estudios.
  - ¿Quién está en riesgo y por qué?
  - ¿Por qué? Así quienes están a cargo de la gestión académica puedan intervenir temprano para ayudar, proporcionar orientación y apoyo.
- Aumentar la disponibilidad de asesores profesionales y consejeros capacitados con un conjunto de herramientas tecnológicas para ayudar a los estudiantes a mantenerse conectados con estos profesionales.
- Identificar a los estudiantes que están seleccionando carreras de riesgo según sus perfiles de riesgo.

Por ejemplo, en la figura 11 se muestra un árbol de decisión completo, que corresponde al resultado del trabajo desarrollado por el autor William R. Veitch del distrito escolar de Colorado Springs en EE.UU. El estudio de Veitch incluye todas las variables como potenciales predictoras, y en donde la variable que más impacta en la deserción, son aquellas asociadas con el perfil de riesgo académico.

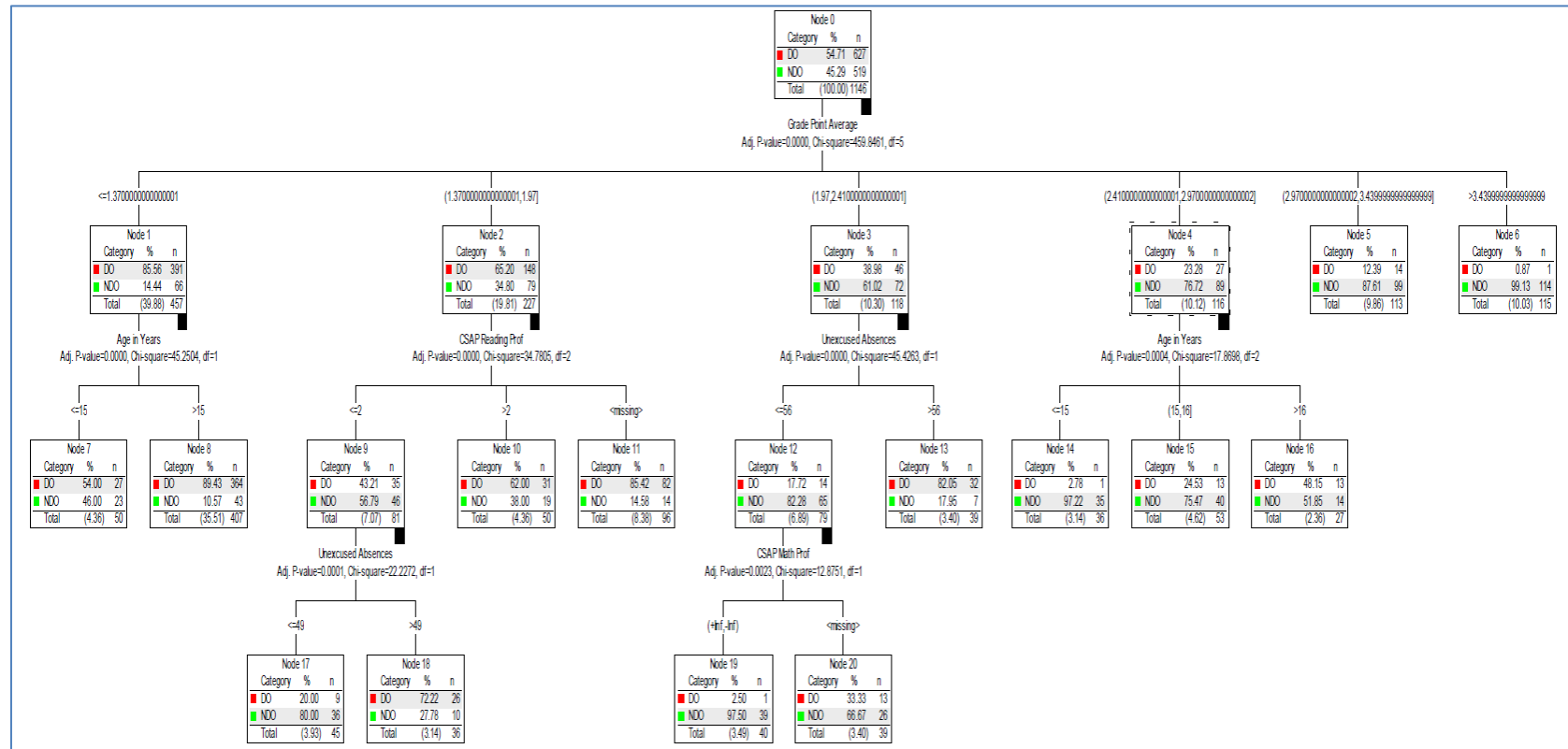


Figura 11: Ejemplo de un Árbol de decisión completo para perfil de riesgo académico  
Fuente: (Veitch et al. 2004)

#### 8.4 Alineamiento Estratégico de la cartera de proyectos con los perfiles de riesgo

Luego de obtener la clasificación de los estudiantes por perfiles de riesgo de deserción o abandono, es necesario alinear los programas y proyectos relativos al manejo de la deserción, con los respectivos grupos de interés. Para ello la organización deberá alinear el portafolio y la cartera de proyectos con los objetivos y estrategia corporativa, como una forma efectiva de asegurar que los proyectos están adecuadamente enfocados al logro de los objetivos y KPI planteados.

La figura 12 muestra el alineamiento que debe existir entre la estrategia, y los respectivos proyectos derivados de los lineamientos e iniciativas estratégicas.

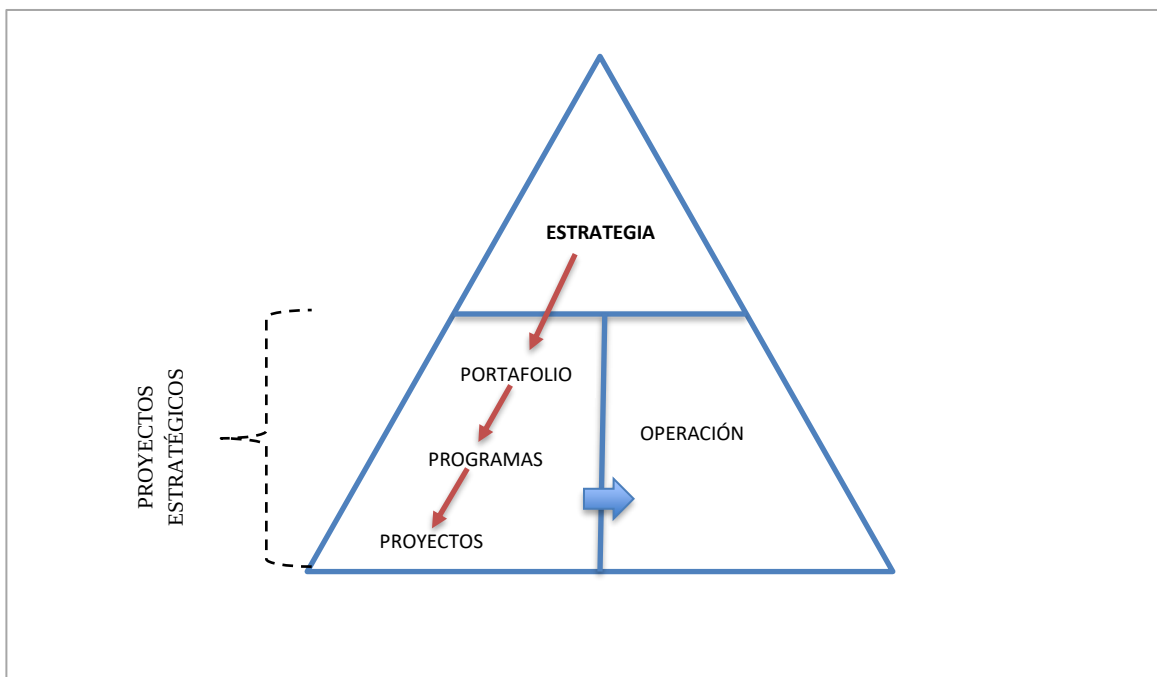


Figura 12: Alineamiento estratégico del portafolio  
(Fuente: Elaboración propia)

De manera de unir la estrategia corporativa, los portafolios, los proyectos y los perfiles de deserción, es necesario la definición de un marco de referencia de gobierno para una adecuada gestión de proyectos estratégicos asociados al tratamiento de la deserción al interior de las universidades. Este marco de referencia se muestra en la figura 13.

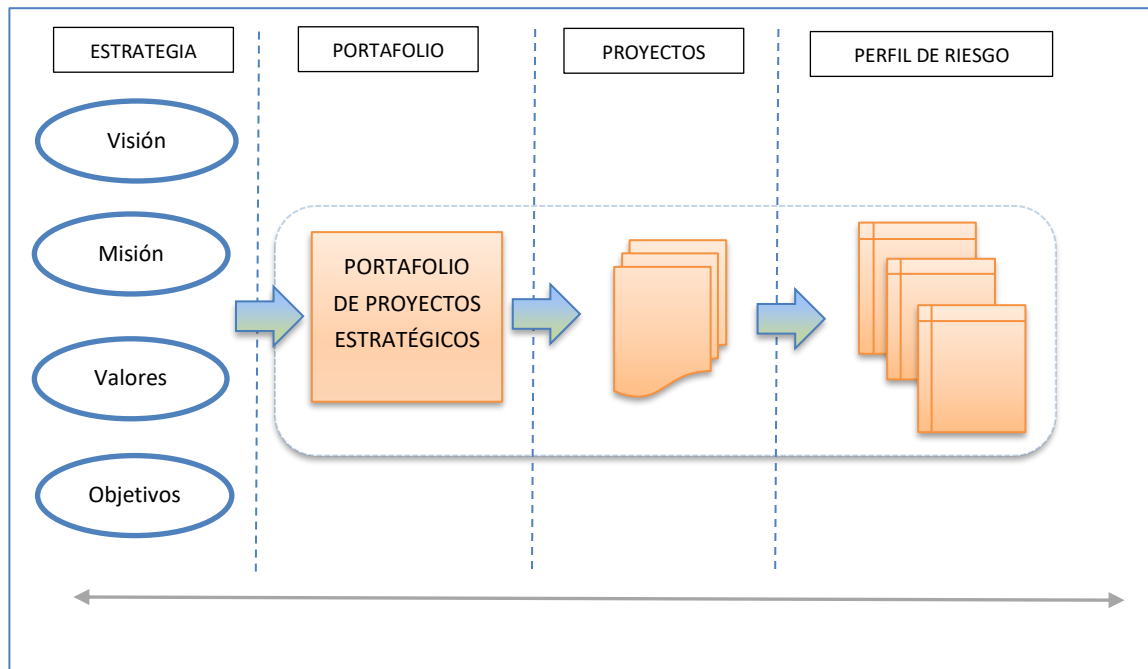


Figura 13: Marco de gobierno para proyectos estratégicos  
(Fuente: Elaboración propia)

### 8.5 Programas y Proyectos para evitar el abandono estudiantil

Las universidades en general tienen ciertas actividades para evitar la deserción y/o aumentar la retención. Estas actividades generalmente son a través de talleres de tipo psicopedagógicos como: estrategias de aprendizaje, desarrollo cognitivo y manejo del estrés. De tipo de apoyo económico como créditos y becas; o de tipo docente como la asignación de tutores al ingreso a la universidad, programas de nivelación de competencias, programas de nivelación en matemática para carreras de ingeniería. Estando la mayoría de estas actividades, no agrupadas en proyectos de tipo estratégicos, sino más bien son aisladas entre sí.

El modelo busca clasificar a los estudiantes de manera temprana por perfiles de deserción, a objeto las universidades puedan ajustar sus programas y proyectos de manera anticipada por cada grupo de riesgo. El modelo no realiza de forma autónoma la asociación entre perfiles de deserción y proyectos. Esta acción deberá realizarse por ahora de forma manual por quienes estén a cargo de la gestión docente en las universidades. Por lo tanto, es necesario que la Institución de Educación Superior diseñe de forma correcta su portafolio de proyectos estratégicos, así como sus indicadores y KPI para la gestión de la deserción o el abandono de los estudiantes de primer año en una Universidad del tipo masiva.

## **8.6 Indicadores o KPI estadísticos asociados a la Deserción**

En Chile, la institución responsable de la recolección y procesamiento de los datos de matrícula es el Sistema de Información de Educación Superior (SIES), organismo que depende de la División de Educación Superior del Ministerio de Educación.

A partir de las bases de matrícula que maneja este organismo, es posible calcular las tasas de retención y deserción a nivel de instituciones y sistema.

En este contexto, la retención se entiende como la contraparte de la deserción. Las estadísticas de retención informan sobre cuantos estudiantes continúan sus estudios y logran avanzar hacia la obtención de su título. Operacionalmente, la tasa de retención al primer año se entiende como el cociente entre el número de estudiantes que ingresan como alumnos de primer año a una carrera o programa en un año determinado, y el número de esos mismos estudiantes que se mantienen como estudiantes antiguos en la misma institución al año siguiente.

Por lo anteriormente expuesto, los siguientes indicadores o KPI debieran ser los observables por el modelo de gestión, dado que son los indicadores que controla el SIES.

- Retención de 1er año por tipo de institución y cohorte
- Retención de 1er año por género y cohorte
- Retención de 1er año por rango de edad y cohorte
- Retención de 1er año por rango de PSU y cohorte
- Retención de 1er año por establecimiento de origen y cohorte
- Retención de 1er año por quintil de ingreso y cohorte
- Retención de 1er año por presencia de beneficios y cohorte

Ver anexo 7 para mayores detalles.

## 9 CONCLUSIONES

La deserción estudiantil es un fenómeno que provoca y genera altos costos a las instituciones de educación superior, al estado y a sus familias. Por consiguiente, teniendo en cuenta las diferentes definiciones de deserción existentes, es deseable para las Universidades distinguir qué tipo o tipos de deserción son los más costosos, e impactan en su “salud financiera”. Al lograr este conocimiento permitirá enfocar mejor los estudios y el diseño de programas y proyectos eficientes para manejar las altas tasas de abandono estudiantil.

Aunque el modelo presentado en este trabajo no establece relaciones causales, este entrega una clasificación temprana por tipo de perfil de deserción, antes que los estudiantes inicien el semestre académico, y con ello lograr ajustar y alinear los programas y proyectos asociados a la gestión de los indicadores estadísticos de la deserción en las universidades chilenas de forma anticipada, a objeto lograr mayores rentabilidades de los proyectos, evidenciadas a través de la mejora de los indicadores y KPI indicados en el punto 8.6 de este documento.

La minería de datos es una herramienta muy útil para una amplia variedad de problemas del mundo real, dominios e industrias donde se están recopilando y almacenando grandes cantidades de datos. Aunque las instituciones educativas no se han aprovechado tanto como en otros dominios (por ejemplo, banca, marketing, atención médica y medicina / biología), se están moviendo ahora a un ritmo creciente para utilizar esta metodología para una variedad de propósitos. Algunas de las áreas de aplicación de minería de datos emergentes notables en el campo de la educación incluyen la evaluación de las necesidades del estudiante, la gestión de la retención, y la mejora en la asignación de asignaturas, entre otras. A medida que aumenta la conciencia sobre las capacidades de la minería de datos, los administradores, los profesionales y los investigadores de las instituciones educativas identificarán la forma de utilizarlo en el análisis y la solución de problemas relacionados con su dominio y aparentemente insolubles.

El éxito en la minería de datos depende de seguir una metodología sólida, prestando la debida atención a cada paso en el proceso, y ser crítica, sobre todo, desde el inicio hasta el final del proceso de experimentación. Los pasos más importantes son los primeros en la metodología KDD, es decir, entender el dominio del problema de negocio y entender los datos. Para algunos expertos, hasta el 85% del tiempo total del proyecto se gasta en estas primeras fases (Turban et al., 2010). El presente estudio fue otro ejemplo para confirmar esta afirmación. Siguiendo una metodología sólida, como KDD ayudó mucho en la organización de presente trabajo y permitió mantenerse enfocado a lo largo del estudio.

En diversos trabajos revisados de la bibliografía, los resultados de validación cruzada indican que las técnicas de minería de datos son capaces de predecir significativamente mejor que su contraparte estadística. Diversos estudios demuestran que la regresión logística, es decir, el algoritmo de árbol de decisión C5 predice con una precisión cercana al 95%, mientras que la regresión logística es capaz de predecir con una precisión cercana a 83 %. Esto nos advierte que, al elegir un método sobre otro para un problema de predicción, además de la precisión de la predicción, también se deben considerar factores como la eficiencia (tiempo necesario para construir un modelo), interpretabilidad (facilidad de comprensión del modelo desarrollado), despliegue (facilidad de despliegue) El modelo de uso real) y la justificación teórica.

El enfoque (y tal vez la limitación) de este estudio es el hecho de que pretende predecir a través de la clasificación, el abandono de los estudiantes utilizando datos institucionales. A pesar de que aprovecha los resultados de los estudios teóricos anteriores, este estudio no tiene por objeto desarrollar una nueva teoría, sino que pretende mostrar la viabilidad de los métodos de minería de datos como un medio para proporcionar una forma alternativa de comprender y predecir la deserción de los estudiantes en las instituciones de educación superior. Desde el punto de vista práctico, un sistema de información de apoyo a la gestión, que abarca estos modelos de predicción se puede utilizar como una ayuda a la toma de decisión de los departamentos de gestión de la admisión de estudiantes en las Universidades masivas, y que son sensibles a la retención de estudiantes. Claro está que deben ser sobre plataformas informáticas sólidas y probado prestigio a nivel mundial, como por ejemplo: Microstrategy, Pentahoo, Rapidminer, etc.

Las potenciales trabajos futuros de este estudio incluyen: (i) extender los métodos y conjuntos de modelado predictivo con otras técnicas más recientes, (ii) mejorar las fuentes de información incluyendo los datos de estudios institucionales basados en encuestas intencionalmente elaborados y administrados cuidadosamente para fines de retención, y (iii) el despliegue del sistema informático como una ayuda a la toma de decisión para que los administradores evalúen su idoneidad y usabilidad en el mundo real.

## 10 BIBLIOGRAFÍA

- Acuña, C., 2012. *Acceso y Deserción en la Educación Superior, Caso Aplicado a Chile*. Universidad de Chile.
- Aguirre, N., 2012. *Factores que Predicen el Rendimiento Académico en la Escuela de Ingeniería de la Universidad de Chile*. Universidad de Chile.
- Alegre, M.Á. & Benito, R., 2010. Los Factores del Abandono Educativo Temprano. España en el Marco Europeo. *Revista de Educación*, Núm. ext., pp.65–92.
- Allan, J.F., McKenna, J. & Dominey, S., 2014. Degrees of resilience: profiling psychological resilience and prospective academic achievement in university inductees. *British Journal of Guidance & Counselling*, 42(1), pp.9–25. Available at: <http://www.tandfonline.com/doi/abs/10.1080/03069885.2013.793784> [Accessed September 27, 2016].
- Alzoabi, Z., Diko, F. & Hanna, S., 2011. Suggested Model for Business Intelligence in Higher Education. In A. A. R. El Sheikh & Mouhib Alnoukar, eds. *Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications: Cross-Disciplinary Applications*. Hershey, Pensilvania: IGI Global, pp. 223–239.
- Arrau, F., 2003. *Deserción en la Educación Superior en Chile*, Santiago.
- Baker, R.W., McNeil, O. V. & Siryk, B., 1985. Expectation and reality in freshman adjustment to college. *Journal of Counseling Psychology*, 32(1), pp.94–103. Available at: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-0167.32.1.94> [Accessed September 26, 2016].
- Baker, R.W. & Schultz, K.L., 1992. Experiential Counterparts of Test-Indicated Disillusionment During Freshman Adjustment to College. *Nacada Journal*, 12(2), pp.13–22.
- Baker, R.W. & Siryk, B., 1984. Measuring Adjustment to College. *Journal of Counseling Psychology*, 31(2), pp.179–189.
- Baker, R.W. & Siryk, B., 1989. *Student Adaptation to College Questionnaire (SACQ)*, Los Angeles: Western Psychological Services.
- Barahona, P. & Aliaga, V., 2013. Variables predictoras del rendimiento académico de los alumnos de primer año de las carreras de Humanidades de la Universidad de Atacama, Chile. *Rev Int Investig Cienc Soc*, 9(2), pp.207–220.
- Barnett, M., 2004. A Qualitative Analysis of Family Support and Interaction among Black College Students at an Ivy League University. *The Journal of Negro Education*, 73(1), p.53. Available at: <http://www.jstor.org/stable/10.2307/3211259?origin=crossref>.
- Barr, J.E. & Rasor, R., 1999. *Freshmen Persistence as Measured by Reaching Academic*

- Achievement Benchmarks.*, Sacramento, CA. Available at:  
<http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED428798&site=ehost-live>.
- Barrios, M., 2011a. Deserción Universitaria en Chile. Incidencia del Financiamiento y otros Factores Asociados. *Revista CIS*, (14), pp.59–72. Available at:  
<http://www.techo.org/wp-content/uploads/2013/02/barrios.pdf>.
- Barrios, M., 2011b. *Deserción y Financiamiento en las Universidades Chilenas*. Pontificia Universidad Católica. Available at:  
<https://repositorio.uc.cl/bitstream/handle/11534/1480/566190.pdf?sequence=1>.
- Barrios, M., Meneses, F. & Paredes, R., 2011. *Financial Aid and University Attrition in Chile*, Santiago.
- Bellei, C., 2013. El “Fin de Lucro” como Política Educacional. In J. E. García-Huidobro & A. Falabella, eds. *Los fines de la Educación*. Santiago: Ediciones Universidad Católica de Chile, pp. 85–114.
- Bernardi, F. & Requena, M., 2010. Desigualdad y Puntos de Inflexión Educativos: el Caso de la Educación Post-obligatoria en España. *Revista de Educacion*, Núm. ext., pp.93–118.
- Boshier, R., 1973. Educational Participation and Dropout: a Theoretical Model. *Adult Education Quarterly*, 23(4), pp.255–282. Available at:  
<http://aeq.sagepub.com/cgi/doi/10.1177/074171367302300401>.
- Canales, A. & De los Ríos, D., 2009. *Entendiendo la Permanencia de Estudiantes Vulnerables en el Sistema Universitario*, Santiago.
- Carrasco, E., Zúñiga, C. & Espinoza, J., 2014. Elección de carrera en estudiantes de nivel socioeconómico bajo de universidades chilenas altamente selectivas. *Calidad en la educación*, 40(40), pp.95–128. Available at:  
[http://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0718-45652014000100004&lng=en&nrm=iso&tlng=en](http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-45652014000100004&lng=en&nrm=iso&tlng=en) [Accessed September 26, 2016].
- Catterall, J.S., 1998. Risk and Resilience in Student Transitions to High School. *American Journal of Education*, 106(2), pp.302–333.
- Centro de Estudios, 2012. *Deserción en la Educación Superior en Chile*, Santiago.
- Centro de Estudios Mineduc, 2012a. *Análisis Módulo Educación Encuesta CASEN 2011*, Santiago. Available at: [http://static1.mineduc.cl/doc\\_mailing/20121026\\_analisis\\_casen\\_final.pdf](http://static1.mineduc.cl/doc_mailing/20121026_analisis_casen_final.pdf).
- Centro de Estudios Mineduc, 2012b. *Serie Evidencias: Realidad educativa en Chile: ¿Qué aprendemos de la Encuesta CASEN 2011?*, Santiago.
- Centro de Microdatos, 2008. *Estudio sobre Causas de la Deserción Universitaria*, Santiago. Available at:  
[http://www.opech.cl/educsuperior/politica\\_acceso/informe\\_final\\_causas\\_desercion\\_un](http://www.opech.cl/educsuperior/politica_acceso/informe_final_causas_desercion_un)

iversitaria.pdf.

- Centro Interuniversitario de Desarrollo CINDA, 2006. *Repitencia y Deserción Universitaria en América Latina 1ª*, Santiago: Alfabetas Artes Gráficas.
- Cepe, Universidad Alberto Hurtado & CIAE, 2013. *Universidades Chilenas: Institucionalidad y Regulación*, Santiago.
- Chaudhuri, S., Dayal, U. & Narasayya, V., 2011. An overview of business intelligence technology. *Communications of the ACM*, 54(8), p.88.
- CIDE, 2001. *III Encuesta a Actores del Sistema Educativo*, Santiago.
- CIDE, 2003. *IV Encuesta a Actores del Sistema Educativo*, Santiago.
- Corominas, E., 2001. La Transición de los Estudios Universitarios. Abandono o Cambio en el Primer Año de Universidad. *Revista de Investigación Educativa*, 19(1), pp.127–151. Available at: <http://dugi-doc.udg.edu/bitstream/handle/10256/9842/TrasicionEstudios.pdf?sequence=1>.
- Credé, M. & Niehorster, S., 2012. Adjustment to College as Measured by the Student Adaptation to College Questionnaire: A Quantitative Review of its Structure and Relationships with Correlates and Consequences. *Educational Psychology Review*, 24(1), pp.133–165. Available at: <http://link.springer.com/10.1007/s10648-011-9184-5> [Accessed September 26, 2016].
- Custodio, L., 2010. Caracterización de los Desertores de Udelar (Año 2006): desde la Inversión y el Consumo hacia la Exclusión Académica y la Deserción Voluntaria. In T. Fernández, ed. *La desafiliación en la Educación Media y Superior de Uruguay: conceptos, estudios, políticas*. Montevideo: Comisión Sectorial de Investigación Científica, Universidad de la República, pp. 153–168.
- Departamento de Investigación e Información Pública - Consejo Nacional de Educación, 2016. Estadísticas y bases de datos INDICES [base de datos en línea]. *Indicadores Educación Superior*. Available at: [http://www.cned.cl/public/Secciones/SeccionIndicesPostulantes/CNED\\_IndicesTabla\\_u\\_MatriculaSistema.html?IdRegistro=005](http://www.cned.cl/public/Secciones/SeccionIndicesPostulantes/CNED_IndicesTabla_u_MatriculaSistema.html?IdRegistro=005) [Accessed September 24, 2016].
- Díaz, C., 2009. Factores de Deserción Estudiantil en Ingeniería: Una Aplicación de Modelos de Duración. *Informacion Tecnologica*, 20(5), pp.129–145.
- Díaz, C., 2008. Modelo Conceptual para la Deserción Estudiantil Universitaria Chilena. *Estudios Pedagógicos*, XXXIV(2), pp.65–86.
- Donoso, S. & Schiefelbein, E., 2007. Análisis de los Modelos Explicativos de Retención de Estudiantes en la Universidad: una Visión desde la Desigualdad Social. *Estudios Pedagógicos*, 33(1), pp.7–27.
- Elias, M., 2008. Los Abandonos Universitarios: Retos ante el Espacio Europeo de Educación Superior. *Estudios sobre Educación ESE*, (15), pp.101–121.

- Erdoğan, Ş.Z. & Timor, M., 2005. A Data Mining Application in a Student Database. *Journal of Aeronautics & Space Technologies*, 2(2), pp.53–57. Available at: <http://search.ebscohost.com/login.aspx?direct=true&AuthType=cookie,ip,uid&db=afh&AN=29406812&site=ehost-live>.
- Ethington, C., 1990. A psychological model of student persistence. *Research in Higher Education*, 31(3), pp.279–293.
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth Padharic, 1996. From data mining to knowledge discovery in databases. *Advances in Knowledge Discovery and Data Mining*, 17(3), pp.1–36.
- Fernández, T., 2010. *La Desafiliación en la Educación Media y Superior de Uruguay: Conceptos, Estudios y Políticas* 1ª. T. Fernández, ed., Montevideo: Central de Impresiones S.A.
- Fishbein, M. & Ajzen, I., 1975. *Belief, attitude, intention, and behavior: an introduction to theory and research* 1ª., Reading, Massachusetts: Addison-Wesley. Available at: <http://people.umass.edu/ajzen/f&a1975.html>.
- Floyd, C., 1996. Achieving Despite the Odds: a Study of Resilience Among a Group of Africa American High School Seniors. *The Journal of Negro Education*, 65(2), pp.181–189.
- Gajardo, L., 2006. Desigualdad en el Acceso a la Educación Superior en Chile. *Revista Central de Sociología*, (1), pp.11–29.
- Gladieux, L. & Swail, W., 1998. Postsecondary Education: Student Success, Not Just Access. In *The Forgotten Half Revisited: American Youth And Young Families, 1998-2000*. Washington, DC: American Youth Policy Forum.
- Goldfinch, J. & Hughes, M., 2007. Skills, learning styles and success of first-year undergraduates. *Active Learning in Higher Education*, 8(3), pp.259–273.
- Gómez, D., 2013. *Hacia la Construcción Colectiva de un Marco Conceptual para Analizar, Predecir, Evaluar y Atender el Abandono Estudiantil en la Educación Superior: Síntesis*, Medellín.
- González, L. & Uribe, D., 2002. Estimaciones sobre la “Repitencia” y Deserción en la Educación Superior Chilena. Consideraciones sobre sus Implicaciones. *Calidad en la Educación*, 17(1), pp.77–90.
- González, L., Uribe, D. & González, S., 2005. *Estudio sobre la Repitencia y Deserción en la Educación Superior Chilena*, Santiago.
- Han, J., Kamber, M. & Pei, J., 2012. *Data Mining: Concepts and Techniques* Third Edit. J. Han, M. Kamber, & J. Pei, eds., Waltham, Boston: Morgan Kaufmann.
- Hart, M. et al., 2007. Introducing students to business intelligence: Acceptance and perceptions of OLAP software. *Issues in Information Science and Information Technology*, 4(1), pp.105–123.

- Hays, R.B. & Oxley, D., 1986. Social network development and functioning during a life transition. *Journal of Personality and Social Psychology*, 50(2), pp.305–313. Available at: <http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.50.2.305>.
- Hernández, C., 2013. *Hacia la Construcción Colectiva de un Marco Conceptual para Analizar, Predecir, Evaluar y Atender el Abandono Estudiantil en la Educación Superior: Marco Conceptual*, Medellín.
- Hernández, R., Fernández, C. & Baptista, P., 2006. *Metodología de la Investigación 4ª*, México, D.F.: McGraw-Hill Interamericana.
- Jarrett, R., 1997. Resilience among Low-Income African American Youth: An Ethnographic Perspective. *Ethos*, 25(2), pp.218–229. Available at: <http://doi.wiley.com/10.1525/eth.1997.25.2.218>.
- Johnson, M.L. et al., 2015. Models and messengers of resilience: a theoretical model of college students' resilience, regulatory strategy use, and academic achievement. *Educational Psychology*, 35(7), pp.869–885. Available at: <http://www.tandfonline.com/doi/full/10.1080/01443410.2014.893560> [Accessed September 27, 2016].
- Kovačić, Z., 2010. Early Prediction of Student Success: Mining Students Enrolment Data. In *Informing Science and Information Technology Education Conference*. Santa Rosa, California: Informing Science Institute, pp. 647–665.
- Larroucau, T., 2013. *Estudio de los Factores Determinantes de la Deserción en el Sistema Universitario Chileno*. Universidad de Chile. Available at: <http://repositorio.uchile.cl/handle/2250/114843>.
- Laudon, K. & Laudon, J., 2012. *Sistemas de Información Gerencial 12da edici.*, México, D.F.: Pearson Educación.
- Libertad y Desarrollo, 2015. Casen 2013: Continúa el alza en Cobertura en Educación Superior. 30 de enero. Available at: <http://lyd.org/centro-de-prensa/noticias/2015/01/casen-2013-continua-alza-en-cobertura-de-educacion-superior/>.
- de los Ríos, D., 2000. *Indicadores de Calidad y Eficiencia en la Educación Universitaria: algunas Propuestas para el Sistema de Acreditación Chileno.*, Santiago.
- de los Ríos, D. & Canales, A., 2007. Factores Explicativos de la Deserción Universitaria. *Calidad en la educación*, (26), pp.173–201.
- Masten, A.S., 1994. Resilience in Individual Development: Successful Adaptation despite Risk and Adversity. In M. Wang & E. Gordon, eds. *Educational resilience in inner city America: Challenges and prospects*. Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers, pp. 3–25.
- Mena, L., Fernández, M. & Riviére, J., 2010. Desenganchados de la Educación: Procesos, Experiencias, Motivaciones y Estrategias de Abandono y del Fracaso Escolar. *Revista*

de Educación, Núm. ext., pp.119–145.

- Meneses, F. & Blanco, C., 2010. *Financial Aid and Higher Education Enrollment in Chile: A Government Policy Analysis*, Munich. Available at: <https://mpra.ub.uni-muenchen.de/23321/>.
- Ministerio de Desarrollo Social, 2013. *Casen 2013: Educación. Síntesis de Resultados.*, Santiago.
- Mischel, W., Zeiss, R. & Zeiss, A., 1974. Internal-External Control and Persistence: Validation and Implications of the Stanford Preschool Internal-External Scale. *Journal of Personality and Social Psychology*, 29(2), pp.265–278.
- Navarrete, S., Candia, R. & Puchi, R., 2013. Factores Asociados a la Deserción / Retención de los Estudiantes Mapuche de la Universidad de la Frontera e Incidencia de los Programas de Apoyo Académico. *Calidad en la educación*, 39(2), pp.44–80. Available at: <http://dx.doi.org/10.4067/S0718-45652013000200003>.
- OECD, 2008. *Education at a Glance 2008: OECD Indicators*, Paris.
- OECD, 2009. *Education at a Glance 2009: OECD Indicators*, Paris. Available at: [http://www.oecd-ilibrary.org/education/education-at-a-glance-2009\\_eag-2009-en](http://www.oecd-ilibrary.org/education/education-at-a-glance-2009_eag-2009-en).
- Oliva, M.A., 2008. Política Educativa y Profundización de la Desigualdad en Chile. *Estudios Pedagógicos*, 34(2), pp.207–226. Available at: [http://www.scielo.cl/scielo.php?script=sci\\_arttext&pid=S0718-07052008000200013&lng=en&nrm=iso&tlng=en](http://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0718-07052008000200013&lng=en&nrm=iso&tlng=en) [Accessed September 17, 2016].
- Pandya, R., 2015. C5 . 0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning. , 117(16), pp.18–21.
- Pascarella, E.T. & Chapman, D.W., 1983. A Multiinstitutional, Path Analytic Validation of Tinto's Model of College Withdrawal. *American Educational Research Journal*, 20(1), pp.87–102. Available at: <http://aer.sagepub.com/cgi/doi/10.3102/00028312020001087>.
- Pascarella, E.T. & Terenzini, P.T., 2005. *How college affects students: A third decade of research*. 1ª., San Francisco: Jossey-Bass.
- Poy, R., 2010. Efectos del Credencialismo y de las Expectativas Sociales sobre el Abandono Escolar. *Revista de educación*, Núm. ext., pp.147–169. Available at: <http://dialnet.unirioja.es/servlet/articulo?codigo=3342424&info=resumen&idioma=ENG>.
- Rolando, R., 2007. *Retención de Primer Año en Educación Superior*, Santiago.
- Ronald J. Brachman & Anand, T., 1996. The process of Knowledge Discovery in Databases: A first sketch. *AAAI Press / The MIT Press*, pp.37–57. Available at: <http://www.aaai.org/Papers/Workshops/1994/WS-94-03/WS94-03-001.pdf>.
- Saldaña, M. & Barriga, O., 2010. Adaptación del modelo de deserción universitaria de

- Tinto a la Universidad Católica de la Santísima Concepción, Chile. *Revista de Ciencias Sociales*, XVI(4), pp.616–628.
- Sánchez, G., Navarro, W. & García, A., 2009. Factores de Deserción Estudiantil en la Universidad Surcolombiana. *Revista Paideia Surcolombiana*, (14), pp.97–103.
- Santelices, M. et al., 2011. *Determinantes de Deserción en la Educación Superior Chilena, con énfasis en Efecto de Becas y Créditos 1ª*. Centro de Estudios MINEDUC, ed., Santiago: GRÁFIKA.
- Schoenbach, V.J., 1999. Diseños de estudio analíticos. In *Comprendiendo los Fundamentos de la Epidemiología*. Chapel Hill, North Carolina: School of Public Health, pp. 209–252. Available at: <http://www.epidemiolog.net/es/endesarrollo/DisenosDeEstudioAnaliticos.pdf>.
- Şen, B., Uçar, E. & Delen, D., 2012. Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, 39(10), pp.9468–9476. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0957417412003752>.
- El Sheikh, A.A.R. & Alnoukar, Mouhib, 2011. *Business Intelligence and Agile Methodologies for Knowledge-Based Organizations: Cross-Disciplinary Applications: Cross-Disciplinary Applications 1ª*, Hershey, Pensilvania: Business Cience Reference.
- Superby, J.F., Vandamme, J.-P. & Meskens, N., 2006. Determination of factors influencing the achievement of the first-year university students using data mining methods. In *8th International Conference on Intelligent Tutoring Systems, Educational Data Mining Workshop*. Jhongli, Taiwan: Louvain School of Management.
- Terenzini, P.T. & Pascarella, E.T., 1991. Twenty years of research on college students: Lessons for future research. *Research in Higher Education*, 32(1), pp.83–92. Available at: <http://link.springer.com/10.1007/BF00992835>.
- Tinto, V., 1989. Definir la Deserción: una Cuestión de Perspectiva. *Revista de la Educación Superior*, XVIII((3)71), pp.33–51.
- Tinto, V., 1987. *Leaving College: Rethinking the Causes and Cures of Student Attrition 1ª*, Chicago: University of Chicago Press.
- Tinto, V., 1993. Reflexiones sobre el Abandono de los Estudios Superiores. *Perfiles Educativos*, (62), pp.56–63. Available at: <http://132.248.192.201/seccion/perfiles/1993/n62a1993/mx.peredu.1993.n62.p56-63.pdf>.
- Uribe, D., 2004. Oferta educativa y oferta de graduados de educación superior. In J. J. Brunner & P. Meller, eds. *Oferta y demanda de profesionales y técnicos en Chile*. Santiago: Ril Editores, pp. 131–170.
- Veitch, W.R., Ph, D. & Veitch, W.R., 2004. Identifying Characteristics of High School

Dropouts : Data Mining With A Decision Tree Model Colorado Springs ( CO ) School District 11 Presented at the Annual Meeting of the American Educational Research Association San Diego , CA Identifying Characteristics of High School Dropouts : Data Mining With A Decision Tree Model. , pp.1–11.

Walker, C., Gleaves, A. & Grey, J., 2006. Can students within higher education learn to be resilient and, educationally speaking, does it matter? *Educational Studies*, 32(3), pp.251–264. Available at: <http://www.tandfonline.com/doi/full/10.1080/03055690600631184> [Accessed September 27, 2016].

Yu, C.H. et al., 2007. *A Data-Mining Approach to Differentiate Predictors of Retention.*, Austin, TX. Available at: <http://eric.ed.gov/?id=ED496657>.

Zúñiga, D. et al., 2009. Modelos de Predicción del Rendimiento Académico de los Estudiantes de Medicina en el Ciclo Básico y Preclínico: Un Estudio Longitudinal. *Revista Médica de Chile*, 137(10), pp.1291–1300. Available at: <http://dx.doi.org/10.4067/S0034-98872009001000003>.

## 11 ANEXOS.

ANEXO 1: Cuadro Resumen Valor Asumido por las Variables Asociadas a la Deserción Estudiantil en Estudios Chilenos

| Estudio<br>Autores                                                                                                                                                                                                        | Grupo<br>especifico       | Indicador                                    | Retención             |           | Deserción           |           | Rendimiento |           |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------|----------------------------------------------|-----------------------|-----------|---------------------|-----------|-------------|-----------|
|                                                                                                                                                                                                                           |                           |                                              | Variables protectoras |           | Variables de riesgo |           |             |           |
|                                                                                                                                                                                                                           |                           |                                              | R directa             | R inversa | R directa           | R inversa | R directa   | R inversa |
| Factores que Predicen el Rendimiento Académico en la Escuela de Ingeniería de la Universidad de Chile.<br>Aguirre, N. (2012)                                                                                              | Ingeniería civil Santiago | PSU de Ciencias – modulo física-, matemática | Explica               |           |                     |           |             |           |
| Modelos de Predicción del Rendimiento Académico de los Estudiantes de Medicina en el Ciclo Básico y Preclínico: Un Estudio Longitudinal.<br>Zúñiga, D., Mena, B., Oliva, R., Pedrals, N., Padilla, O. y Bitran, M. (2009) | Medicina Santiago         | NEM                                          | ✓                     |           |                     |           |             |           |
|                                                                                                                                                                                                                           |                           | PSU de matemática y biología                 | ✓                     |           |                     |           |             |           |
|                                                                                                                                                                                                                           |                           | Estilo activo                                |                       | ✓         |                     |           |             |           |
|                                                                                                                                                                                                                           |                           | Procesamiento abstracto                      | ✓                     |           |                     |           |             |           |
|                                                                                                                                                                                                                           | Ingeniería                | Colegio municipal                            |                       |           | ✓                   |           |             |           |

|                                                                                                                                                                                                             |                      |                                        |   |  |   |  |   |   |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------|----------------------------------------|---|--|---|--|---|---|
| Factores de Deserción Estudiantil en Ingeniería: Una Aplicación de Modelos de Duración.<br>Díaz, C. (2009)                                                                                                  | Santiago             | PPA                                    | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | Crédito universitario                  | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | PSU                                    | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | Media científico-humanista             | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | Ingresos familiares altos              | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | Beca                                   |   |  | ✓ |  |   |   |
|                                                                                                                                                                                                             |                      | Edad de egreso (mayor)                 |   |  | ✓ |  |   |   |
|                                                                                                                                                                                                             |                      | Mayor lugar preferencia en postulación |   |  | ✓ |  |   |   |
| Factores Asociados a la Deserción / Retención de los Estudiantes Mapuche de la Universidad de la Frontera e Incidencia de los Programas de Apoyo Académico.<br>Navarrete, S., Candia, R. y Puchi, R. (2013) | Mapuche Araucanía    | Ruralidad                              |   |  | ✓ |  |   |   |
|                                                                                                                                                                                                             |                      | Menor lugar preferencia                | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | Establecimiento científico humanista   | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | Promedio PSU                           | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | NSE                                    | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | Al segundo año (PGA_PPA)               | ✓ |  |   |  |   |   |
|                                                                                                                                                                                                             |                      | Avance curricular                      | ✓ |  |   |  |   |   |
| Variables predictoras del rendimiento académico de los alumnos de primer año                                                                                                                                | Humanida des Atacama | Trabaja                                |   |  |   |  |   | ✓ |
|                                                                                                                                                                                                             |                      | Conformidad con la carrera             |   |  |   |  | ✓ |   |

|                                                                                                                                                         |                       |                                                 |  |  |   |         |   |  |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|-------------------------------------------------|--|--|---|---------|---|--|
| de las carreras de Humanidades de la Universidad de Atacama, Chile. Barahona, P. y Aliaga, V. (2013)                                                    |                       | PSU                                             |  |  |   |         | ✓ |  |
|                                                                                                                                                         |                       | NEM                                             |  |  |   |         | ✓ |  |
| Adaptación del modelo de deserción universitaria de Tinto a la Universidad Católica de la Santísima Concepción, Chile. Saldaña, M. y Barriga, O. (2010) | Ingeniería Concepción | Ingresos familiares y                           |  |  | ✓ |         |   |  |
|                                                                                                                                                         |                       | PSU de Lenguaje y Matemática                    |  |  | ✓ |         |   |  |
|                                                                                                                                                         |                       | Promedio semestral                              |  |  |   | ✓       |   |  |
|                                                                                                                                                         |                       | Créditos aprobados                              |  |  |   | ✓       |   |  |
|                                                                                                                                                         |                       | Cobertura arancel (beca)                        |  |  |   | ✓       |   |  |
|                                                                                                                                                         |                       | Ingresos familiares                             |  |  |   | Explica |   |  |
| Factores Explicativos de la Deserción Universitaria. de los Ríos, D. y Canales, A. (2007)                                                               | Cualitativo nacional  | Rendimiento estudiantes                         |  |  |   | Explica |   |  |
|                                                                                                                                                         |                       | Baja integración a las carreras o instituciones |  |  | ✓ |         |   |  |
|                                                                                                                                                         |                       | Baja calidad institucional                      |  |  | ✓ |         |   |  |
|                                                                                                                                                         |                       | Falta de claridad vocacional                    |  |  | ✓ |         |   |  |
|                                                                                                                                                         |                       | Postular a carreras no prioritarias             |  |  | ✓ |         |   |  |
|                                                                                                                                                         |                       | Problemas hábitos de estudio-disciplina-        |  |  | ✓ |         |   |  |
|                                                                                                                                                         |                       | Escuelas municipales y                          |  |  | ✓ |         |   |  |

|                                                                                 |                    |                                                       |  |  |            |   |  |  |
|---------------------------------------------------------------------------------|--------------------|-------------------------------------------------------|--|--|------------|---|--|--|
|                                                                                 |                    | subvencionadas (ritmo)                                |  |  |            |   |  |  |
|                                                                                 |                    | Padres y redes cercanas sin experiencia universitaria |  |  | ✓          |   |  |  |
|                                                                                 |                    | Expectativas profesionales no cumplidas               |  |  | ✓          |   |  |  |
|                                                                                 |                    | Cambio de condiciones económicas familiares           |  |  | ✓          |   |  |  |
|                                                                                 |                    | Cambio función familiar (hijo, muerte sostenedor)     |  |  | ✓          |   |  |  |
| Estudio sobre Causas de la Deserción Universitaria. Centro de Microdatos (2008) | Población Nacional | Problemas vocacionales (primera pref-información)     |  |  | ✓          |   |  |  |
|                                                                                 |                    | Situación económica de las familias                   |  |  | ✓          |   |  |  |
|                                                                                 |                    | Rendimiento académico                                 |  |  | ✓          |   |  |  |
|                                                                                 |                    | Hombres                                               |  |  | ✓          |   |  |  |
| Deserción y Financiamiento en las Universidades Chilenas.                       | Población Nacional | Becas                                                 |  |  | Sin efecto |   |  |  |
|                                                                                 |                    | Crédito (+aval del estado)                            |  |  |            | ✓ |  |  |

|                                                                                                                 |                    |                                        |   |  |   |   |  |  |
|-----------------------------------------------------------------------------------------------------------------|--------------------|----------------------------------------|---|--|---|---|--|--|
| Barrios, M. (2011)                                                                                              |                    | Padres sin estudios universitarios     |   |  | ✓ |   |  |  |
|                                                                                                                 |                    | Vivir solos                            |   |  | ✓ |   |  |  |
|                                                                                                                 |                    | aumento de la edad en hombres          |   |  | ✓ |   |  |  |
|                                                                                                                 |                    | mujeres casadas                        |   |  | ✓ |   |  |  |
| Deserción Universitaria en Chile. Incidencia del Financiamiento y otros Factores Asociados. Barrios, M. (2011a) | Población Nacional | Mayor ingreso familiar                 |   |  | ✓ |   |  |  |
|                                                                                                                 |                    | Ayuda económica                        |   |  |   | ✓ |  |  |
|                                                                                                                 |                    | Mujeres                                |   |  |   | ✓ |  |  |
|                                                                                                                 |                    | Estudiantes solteros                   |   |  |   | ✓ |  |  |
|                                                                                                                 |                    | NEM alto                               |   |  |   | ✓ |  |  |
|                                                                                                                 |                    | PSU alta                               |   |  |   | ✓ |  |  |
|                                                                                                                 |                    | Mayor edad del estudiante              |   |  | ✓ |   |  |  |
|                                                                                                                 |                    | Escuelas municipales                   |   |  | ✓ |   |  |  |
|                                                                                                                 |                    | Universidad acreditada                 |   |  | ✓ |   |  |  |
|                                                                                                                 |                    | Mujer                                  | ✓ |  |   |   |  |  |
|                                                                                                                 |                    | Mayor ingreso familiar                 | ✓ |  |   |   |  |  |
|                                                                                                                 |                    | Instituciones acreditadas              | ✓ |  |   |   |  |  |
|                                                                                                                 |                    | Ingresar al programa en primera opción | ✓ |  |   |   |  |  |
|                                                                                                                 |                    | Mejor puntaje NEM                      | ✓ |  |   |   |  |  |
|                                                                                                                 |                    | Mejor PSU                              | ✓ |  |   |   |  |  |

|                                                                                               |                       |                                                               |   |   |  |   |        |  |
|-----------------------------------------------------------------------------------------------|-----------------------|---------------------------------------------------------------|---|---|--|---|--------|--|
|                                                                                               |                       | Ayuda estudiantil<br>1° CAE<br>2° Fondo solidario<br>3° Becas | ✓ |   |  |   |        |  |
|                                                                                               |                       | Edad de ingreso del<br>estudiante                             |   | ✓ |  |   |        |  |
|                                                                                               |                       | Ser soltero                                                   |   | ✓ |  |   |        |  |
|                                                                                               |                       | Padres que no<br>completaron la<br>educación escolar          |   | ✓ |  |   |        |  |
|                                                                                               |                       | Vivir de manera<br>independiente                              |   | ✓ |  |   |        |  |
|                                                                                               |                       | Duración del<br>programa                                      |   | ✓ |  |   |        |  |
|                                                                                               |                       | Egresar de colegios<br>municipales                            |   | ✓ |  |   |        |  |
| Acceso y Deserción en<br>la Educación Superior,<br>Caso Aplicado a Chile.<br>Acuña, C. (2012) | Población<br>Nacional | Quintil SIMCE                                                 |   |   |  |   | Acceso |  |
|                                                                                               |                       | Quintil PSU                                                   |   |   |  | ✓ |        |  |
|                                                                                               |                       | Quintil PSU<br>establecimiento                                |   |   |  | ✓ | ✓      |  |
|                                                                                               |                       | Quintil SIMCE<br>establecimiento                              |   |   |  | ✓ | ✓      |  |
|                                                                                               |                       | Universidad<br>acreditada                                     |   |   |  | ✓ |        |  |
|                                                                                               |                       | Proporción<br>financiamiento<br>beca o crédito<br>(+crédito)  |   |   |  | ✓ |        |  |
|                                                                                               |                       | un hermano en<br>educación superior                           |   |   |  | ✓ |        |  |

|                                                                                                                            |                    |                                                          |  |  |      |      |  |  |
|----------------------------------------------------------------------------------------------------------------------------|--------------------|----------------------------------------------------------|--|--|------|------|--|--|
| <p>Estudio de los Factores Determinantes de la Deserción en el Sistema Universitario Chileno.<br/>Larroucau, T. (2013)</p> |                    | 2 o más hermanos en prebásica y básica                   |  |  | ✓    |      |  |  |
|                                                                                                                            |                    | padres con educación inferior a educación media          |  |  | ✓    |      |  |  |
|                                                                                                                            | Población Nacional | Puntaje NEM                                              |  |  |      | Bajo |  |  |
|                                                                                                                            |                    | Mejor ranking de notas EM                                |  |  |      | Bajo |  |  |
|                                                                                                                            |                    | PSU matemática y lenguaje.                               |  |  |      | Bajo |  |  |
|                                                                                                                            |                    | Mayor nivel ingreso del núcleo                           |  |  |      | ✓    |  |  |
|                                                                                                                            |                    | Financiamiento con crédito, becas o combinación de ambos |  |  |      | ✓    |  |  |
|                                                                                                                            |                    | Nivel educacional de la madre                            |  |  |      | Bajo |  |  |
|                                                                                                                            |                    | La edad del estudiante                                   |  |  | Bajo |      |  |  |
|                                                                                                                            |                    | Estudiante hombre                                        |  |  | Bajo |      |  |  |
|                                                                                                                            |                    | PSU matemática del colegio                               |  |  |      | ✓    |  |  |
|                                                                                                                            |                    | Colegio con menor grado vulnerabilidad económica         |  |  |      | ✓    |  |  |
|                                                                                                                            |                    | Universidades selectivas del CRUCH                       |  |  |      | ✓    |  |  |

|  |  |                                                |  |  |                       |   |  |  |
|--|--|------------------------------------------------|--|--|-----------------------|---|--|--|
|  |  | Empleabilidad de carreras                      |  |  |                       | ✓ |  |  |
|  |  | Carreras “Ciencias”                            |  |  | + deser               |   |  |  |
|  |  | Carreras “Educación y Humanidades”.            |  |  | - deser               |   |  |  |
|  |  | PSU de matemática en carreras de Ciencias      |  |  | Quintiles intermedios |   |  |  |
|  |  | PSU de matemática en “Educación y Humanidades” |  |  | Mayor quintil         |   |  |  |
|  |  | El ingreso familiar a carreras de “Ciencias”   |  |  |                       | ✓ |  |  |
|  |  |                                                |  |  |                       |   |  |  |

## ANEXO 2: Pasos en el Proceso proceso KDD (Fayyad, U. et al. 1996)

Entre los pasos básicos para el desarrollo del proceso están los siguientes:

- Paso 1: Desarrollar una comprensión del dominio de la aplicación y del conocimiento previo relevante a obtener, e identificar la meta u objetivo del proceso KDD desde el punto de vista del cliente, es decir, entendimiento de las necesidades del negocio.
- Paso 2: Crear un conjunto de datos de destino: seleccionar un conjunto de datos o enfocar un subconjunto de variables o muestras de datos en el que se va a realizar el descubrimiento.
- Paso 3: Limpieza de datos y el pre-procesamiento. Las operaciones básicas incluyen eliminar el ruido, recopilar la información necesaria para modelar o contabilizar el ruido, decidir sobre estrategias para manejar los campos de datos faltantes y contabilizar la información de la secuencia temporal y los cambios conocidos.
- Paso 4: Reducción y proyección de datos. Encontrar características útiles para representar los datos dependiendo del objetivo de la tarea. Con los métodos de reducción de la dimensionalidad o de transformación, se puede reducir el número efectivo de variables bajo consideración, o se pueden encontrar representaciones invariantes para los datos
- Paso 5: Los objetivos del proceso KDD descritos en el paso 1, se corresponde con un método particular de minería de datos. Por ejemplo, resumen, clasificación, regresión, agrupación, etc.
- Paso 6: Es el análisis exploratorio, el modelo y la selección de hipótesis. La elección del algoritmo de minería de datos y el método o métodos de selección que se utilizarán para buscar patrones de datos. Este proceso incluye decidir qué modelos y parámetros pueden ser apropiados (por ejemplo, los modelos de datos categorizados son diferentes de los modelos de vectores sobre los reales) y hacer coincidir un método particular de minería de datos con los criterios generales del proceso KDD, por ejemplo, el usuario final podría estar más interesado en entender el modelo que sus capacidades predictivas.
- Paso 7: Minería de datos. Buscar patrones de interés en una forma representacional particular o un conjunto de tales representaciones, incluyendo reglas de clasificación o árboles, regresión y agrupación. El usuario puede ayudar significativamente al método de minería de datos realizando correctamente los pasos anteriores.
- Paso 8: Es interpretar patrones, posiblemente regresando a cualquiera de los pasos 1 a 7 para iteración adicional. Este paso también puede implicar la visualización de los

patrones y modelos extraídos o la visualización de los datos dados los modelos extraídos.

- Paso 9: Está actuando sobre el conocimiento descubierto, usando el conocimiento directamente, incorporando el conocimiento en otro sistema para acciones futuras, o simplemente documentándolo y reportándolo a las partes interesadas. Este proceso también incluye buscar y resolver conflictos potenciales con conocimiento previamente creído (o extraído).

### ANEXO 3: Caracterización de los objetos de datos mediante atributos

*Atributo Nominal:* Nominal significa "relativo a los nombres". Los valores de un atributo nominal son símbolos o nombres de cosas. Cada valor representa algún tipo de categoría, código o estado, por lo que los atributos nominales también tienen el carácter de categóricos. Los valores no tienen ningún orden significativo. Ejemplos: el estado marital y la región de procedencia pueden ser dos atributos que describen a un objeto del estudiante.

*Atributo Binario:* Un atributo binario es un atributo nominal con sólo dos categorías o estados: 0 ó 1, donde 0 típicamente significa que el atributo está ausente, y 1 significa que está presente. Los atributos binarios se denominan booleanos si los dos estados corresponden a verdadero y falso

Un atributo binario es simétrico si ambos estados son igualmente valorables y llevan el mismo peso; es decir, no hay preferencia sobre qué resultado debe ser codificado como 0 ó 1. Uno de estos ejemplos podría ser el atributo género que tiene los estados masculino y femenino. Un atributo binario es asimétrico si los resultados de los estados no son igualmente importantes.

*Atributo Ordinal:* Un atributo ordinal es un atributo con valores posibles que tienen un orden o rango significativo entre ellos, pero no se conoce la magnitud entre valores sucesivos.

*Atributo numérico:* Un atributo numérico es cuantitativo; es decir, es una cantidad medible, representada en valores enteros o reales. Los atributos numéricos pueden ser de escala de intervalo o de escala de razón.

*Atributo de Escala de Intervalos:* Los atributos escalonados en intervalos se miden en una escala de unidades de igual tamaño. Los valores de los atributos escalonados tienen orden y pueden ser positivos, 0 o negativos. Por lo tanto, además de proporcionar una clasificación de los valores, tales atributos nos permiten comparar y cuantificar la diferencia entre los valores. Dado que este tipo de atributos son numéricos, podemos calcular su valor medio, además de la moda como medidas de tendencia central.

*Atributo de Escala de Razón:* Un atributo de escala de razón es un atributo numérico con un punto cero inherente en su origen. Es decir, si una medida es la escala de razón, podemos hablar de un valor como un múltiplo (o relación) de otro valor. Además, los valores

están ordenados, y también podemos calcular la diferencia entre los valores, así como la media, la mediana y la moda.

## Anexo 4: Definición del Primer Perfil de Riesgo

Tabla 2: Tabla de Operacionalización de Variables para el Perfil Riesgo por Integración Social

| <b>Dimensión</b> | <b>Variables</b>         | <b>Descripción</b>                                                          | <b>Tipo de Dato</b>                                                                                 | <b>Dominio</b>  | <b>Indicador</b>           |
|------------------|--------------------------|-----------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------|-----------------|----------------------------|
| Personal         | Minoría Étnica           | Grupo étnico de origen de estudiantes                                       | Nominal Binario                                                                                     | [0 - 1]         | 1: etnia                   |
| Personal         | Cultura Minoritaria      | Grupo cultural de procedencia de estudiantes                                | Nominal Binario                                                                                     | [0 - 1]         | 1: cultura minoritaria     |
| Personal         | Procedencia              | Procedencia urbana o rural                                                  | Nominal Binario                                                                                     | [0 - 1]         | 1: rural                   |
| Personal         | Discapacidad             | Condición mental motriz o perceptiva de desigualdad                         | Nominal Binario                                                                                     | [0 - 1]         | 1: discapacidad            |
| Personal         | Inmigración              | Procedencia de otra nación.                                                 | Nominal Binario<br>1: país africano, latinoamericano, asiático.<br>0: país europeo, norteamericano. | [0 - 1]         | 1: inmigrante discriminado |
| Académica        | Puntaje PSU              | Obtención de puntaje PSU (X) mayor que la media ( $\mu$ ) de la Universidad | Numérico                                                                                            | 480 - 680       | $X \geq \mu$               |
| Académica        | Promedio NEM             | NEM mayor que la media de la universidad                                    | Numérico                                                                                            | 4,2 - 5,8       | $X \geq \mu$               |
| Socioeconómica   | Ingresos Familiares      | distinto a la media                                                         | Numérico                                                                                            | 80.000 - 2.6M\$ | $X \neq \mu$               |
| Institucional    | Acreditación universidad | años de acreditación                                                        | Numérico                                                                                            | 2-7             | $X < 4$                    |

---

|               |                         |                                             |          |     |         |
|---------------|-------------------------|---------------------------------------------|----------|-----|---------|
|               |                         | universidad > 4<br>años                     |          |     |         |
| Institucional | Acreditación<br>carrera | años de<br>acreditación<br>carrera > 4 años | Numérico | 2-7 | $X < 4$ |

---

### Anexo 5: Definición del Segundo Perfil de Riesgo

Tabla 3: Tabla de Operacionalización de Variables para el Perfil Riesgo por Trayectoria Académica

| <b>Dimensión</b> | <b>Variable</b>                                                      | <b>Descripción</b>                                                                                   | <b>Tipo de Dato</b> | <b>Dominio</b> | <b>Indicador</b>                      |
|------------------|----------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|---------------------|----------------|---------------------------------------|
| Académica        | Puntaje PSU del estudiante bajo el promedio de la carrera            | Obtención de puntaje PSU bajo la media de la carrera                                                 | Ordinal Numérico    | 450 - 680      | X < a definir en cada grupo normativo |
| Académica        | Puntaje PSU del estudiante bajo la media de la distribución nacional | obtención de puntaje PSU menor a la mediana                                                          | Ordinal Numérico    | 150-850        | X <500                                |
| Académica        | Bajo NEM del estudiante                                              | puntaje NEM bajo la mediana                                                                          | Ordinal Numérico    | 213 - 825      | X<519                                 |
| Académica        | Bajo puntaje SIMCE del estudiante en segundo medio                   | puntaje SIMCE bajo 245<br>(En estricto se establece por año de rendimiento y sector de conocimiento) | Ordinal Numérico    | 200 -300       | X<245                                 |
| Académica        | Ranking de notas bajo                                                | Puntaje bajo la media de su contexto                                                                 | Ordinal Numérico    | 250 - 850      | X<500                                 |
| Académica        | Bajo puntaje PSU del establecimiento de origen                       | Obtención de puntaje PSU promedio del establecimient                                                 | Ordinal Numérico    | 150-850        | X <500                                |

---

|           |                                                                   |                                                                                                                                  |                     |                                                         |                    |
|-----------|-------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|---------------------|---------------------------------------------------------|--------------------|
|           |                                                                   | o menor a la mediana                                                                                                             |                     |                                                         |                    |
| Académica | Bajo puntaje SIMCE de segundo medio del establecimiento de origen | puntaje SIMCE promedio del establecimiento o bajo 245 (En estricto se establece por año de rendimiento y sector de conocimiento) | Ordinal<br>Numérico | 200 -300                                                | X<245              |
| Académica | Dependencia del establecimiento                                   | Dependencia municipal o particular sub.                                                                                          | Nominal<br>Numérico | 1--- municipal<br>2--- subvencionada<br>3--- particular | 1 y 2              |
| Personal  | Padres que no completaron la educación escolar                    | No haber terminado la educación básica.                                                                                          | Nominal<br>Binario  | [0 - 1]                                                 | 1-- no completaron |

---

### Anexo 6: Definición del Tercer Perfil de Riesgo

Tabla 4: Tabla de Operacionalización de Variables para el Perfil Riesgo por Vulnerabilidad Socio-cultural

| <b>Dimensión</b>   | <b>Variable</b>                                                      | <b>Descripción</b>                                   | <b>Tipo de Dato</b> | <b>Dominio</b>                                                           | <b>Indicador</b>                      |
|--------------------|----------------------------------------------------------------------|------------------------------------------------------|---------------------|--------------------------------------------------------------------------|---------------------------------------|
| Personal Social    | Nivel educativo de los padres                                        | Nivel educativo alcanzado por los padres.            | Nominal Binario     | [0 - 1]                                                                  | 1: ed. inferior a educación media     |
| Personal Social    | Nivel educativo de los padres                                        | Padres que no ingresaron a una universidad           | Binario             | [0 - 1]                                                                  | 1: padres no ingresaron a una U       |
| Personal Educativa | Tipo de colegio de egreso                                            | Dependencia de los colegios de egreso                | Nominal Binario     | [0 - 1]                                                                  | 1: egresado de liceo municipal        |
| Personal Social    | Estado civil por género                                              | Estado civil en relación al genero                   | Nominal Numérico    | 1:hombre soltero<br>2:hombre casado<br>3:mujer soltera<br>4:mujer casada | 4: mujer casada                       |
| Personal Social    | Vivir solos                                                          | Con quien viven los estudiantes.                     | Nominal Binario     | [0 - 1]                                                                  | 1: vivir solos                        |
| Académica          | Puntaje PSU del estudiante bajo el promedio de la carrera            | obtención de puntaje PSU bajo la media de la carrera | Ordinal Numérico    | 450 - 680                                                                | X < a definir en cada grupo normativo |
| Académica          | Puntaje PSU del estudiante bajo la media de la distribución nacional | obtención de puntaje PSU menor a la mediana          | Ordinal Numérico    | 150-850                                                                  | X <500                                |
| Académica          | Bajo NEM del estudiante                                              | puntaje NEM bajo la mediana                          | Ordinal Numérico    | 213 – 825                                                                | X<519                                 |

|           |                                                                   |                                                                                                                                   |                     |                                                         |                      |
|-----------|-------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|---------------------|---------------------------------------------------------|----------------------|
| Académica | Bajo puntaje SIMCE del estudiante en segundo medio                | puntaje SIMCE bajo 245<br>En estricto se establece por año de rendimiento y sector de conocimiento                                | Ordinal<br>Numérico | 200 -300                                                | X<245                |
| Académica | Ranking de notas bajo                                             | Puntaje bajo la media de su contexto                                                                                              | Ordinal<br>Numérico | 250 – 850                                               | X<500                |
| Académica | Bajo puntaje PSU del establecimiento de origen                    | Obtención de puntaje PSU promedio del establecimiento menor a la mediana                                                          | Ordinal<br>Numérico | 150-850                                                 | X <500               |
| Académica | Bajo puntaje SIMCE de segundo medio del establecimiento de origen | puntaje SIMCE promedio del establecimiento bajo 245<br>(En estricto se establece por año de rendimiento y sector de conocimiento) | Ordinal<br>Numérico | 200 -300                                                | X<245                |
| Académica | Dependencia del establecimiento                                   | Dependencia municipal o particular sub.                                                                                           | Nominal<br>Numérico | 1--- municipal<br>2--- subvencionada<br>3--- particular | 1 y 2                |
| Económica | Cambio función familiar (hijo, muerte sostenedora)                | estudiantes que asumen nuevo rol por cambio en las condiciones familiares                                                         | Nominal<br>Binario  | [0 - 1]                                                 | 1: cambio de función |
| Económica | ingresos familiares en                                            | (suma de ingreso                                                                                                                  | Ordinal<br>Numérico | [0 - 125.558]                                           | X < 125.559          |

---

|                                  |                                                                                                  |
|----------------------------------|--------------------------------------------------------------------------------------------------|
| los dos<br>primeros<br>quintiles | familiares y<br>división por el<br>número de<br>miembros) en<br>los dos<br>primeros<br>quintiles |
|----------------------------------|--------------------------------------------------------------------------------------------------|

---

## Anexo 7: ejemplos de indicadores manejados por SIES

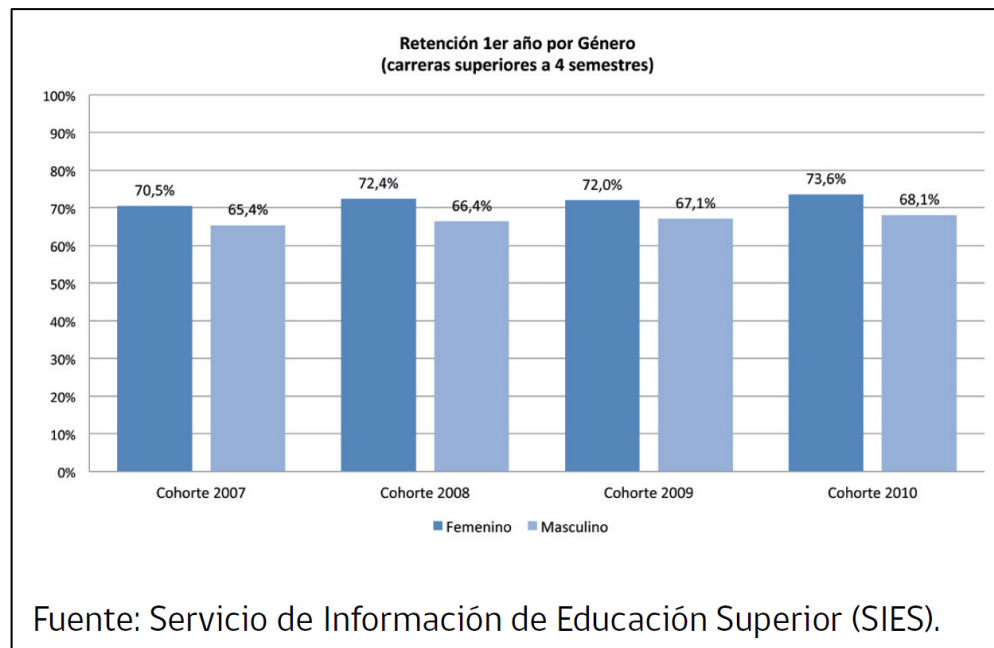
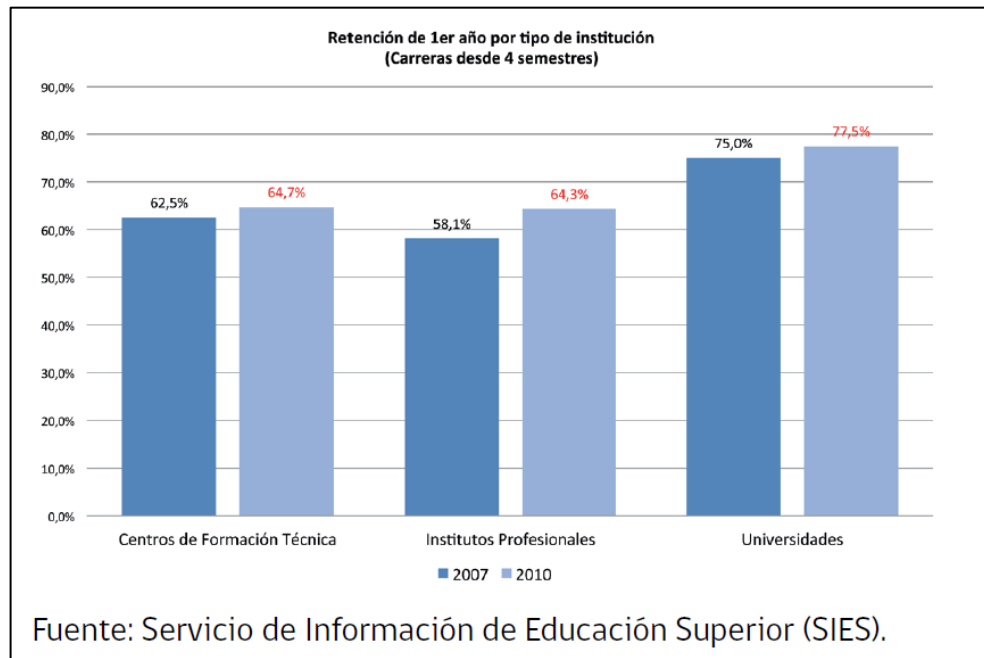


**Tabla N° 1: Retención de 1er año por rango de edad y cohorte**

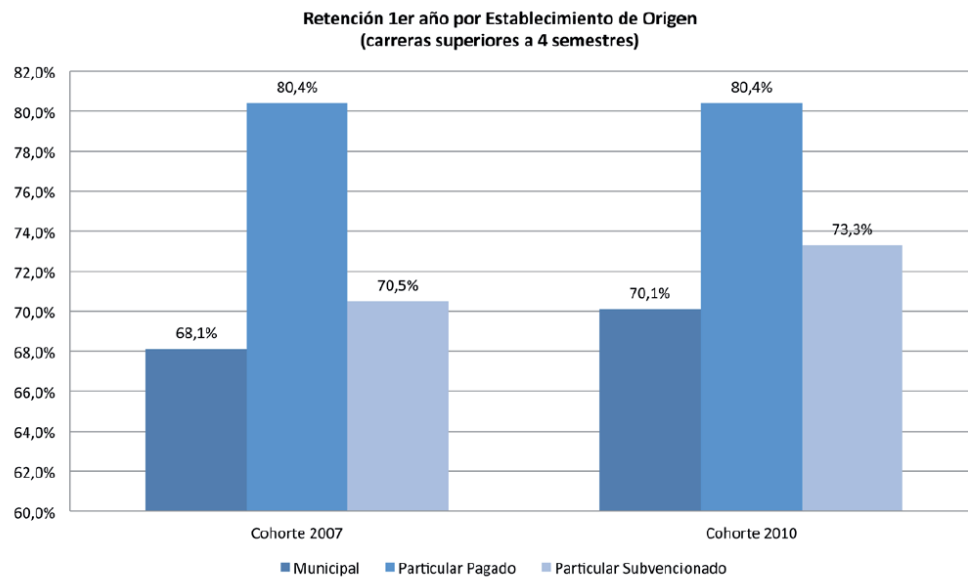
| Rango de Edad | Cohorte 2007 | Cohorte 2008 | Cohorte 2009 | Cohorte 2010 |
|---------------|--------------|--------------|--------------|--------------|
| 15 a 19 años  | 74,8%        | 76,3%        | 76,7%        | 77,7%        |
| 20 a 24 años  | 62,9%        | 65,3%        | 65,7%        | 67,5%        |
| 25 a 29 años  | 57,0%        | 59,2%        | 59,6%        | 62,6%        |
| 30 a 34 años  | 59,2%        | 59,6%        | 59,7%        | 62,3%        |
| 35 a 39 años  | 63,3%        | 59,7%        | 59,8%        | 62,5%        |
| 40 y más años | 61,8%        | 56,3%        | 54,0%        | 61,1%        |
| <b>Total</b>  | <b>67,9%</b> | <b>69,4%</b> | <b>69,6%</b> | <b>70,9%</b> |

Fuente: Servicio de Información de Educación Superior (SIES).

**Tabla N° 2: Retención de 1er año por rango PSU, cohorte 2008**

| Rango PSU para Universidades | Cohorte 2007 | Cohorte 2008 | Cohorte 2009 | Cohorte 2010 |
|------------------------------|--------------|--------------|--------------|--------------|
| Sin PSU                      | 62,7%        | 62,2%        | 60,7%        | 68,4%        |
| $X < 475$                    | 63,7%        | 66,1%        | 66,4%        | 68,1%        |
| $475 \leq X < 500$           | 76,4%        | 77,0%        | 77,3%        | 78,5%        |
| $500 \leq X < 550$           | 79,1%        | 79,2%        | 79,2%        | 80,4%        |
| $550 \leq X < 600$           | 82,9%        | 82,3%        | 81,5%        | 83,1%        |
| $600 \leq X < 650$           | 84,4%        | 83,5%        | 83,7%        | 84,3%        |
| $650 \leq X < 700$           | 86,7%        | 84,7%        | 85,6%        | 85,6%        |
| $700 \leq X < 750$           | 90,3%        | 88,5%        | 89,4%        | 88,8%        |
| $750 \leq X < 800$           | 95,4%        | 93,1%        | 93,7%        | 92,9%        |
| $800 \leq X$                 | 97,5%        | 96,3%        | 96,3%        | 95,3%        |
| <b>Total</b>                 | <b>75,0%</b> | <b>74,6%</b> | <b>74,5%</b> | <b>77,5%</b> |

Fuente: Servicio de Información de Educación Superior (SIES).

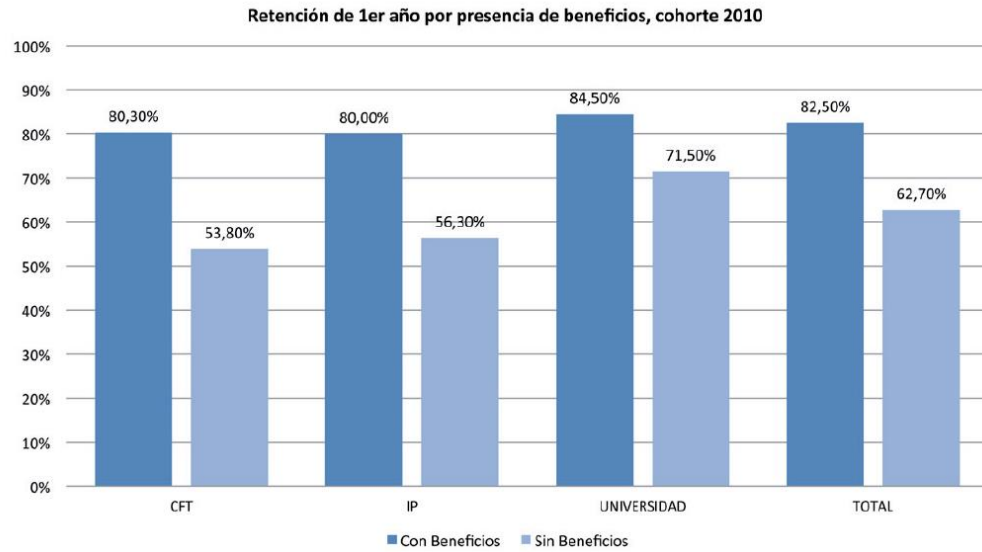


Fuente: Servicio de Información de Educación Superior (SIES).

**Tabla N° 3: Retención de 1er año por quintil de ingresos, cohorte 2008.**

| Quintil (según FUAS) | Cohorte 2008 | Cohorte 2009 | Cohorte 2010 |
|----------------------|--------------|--------------|--------------|
| Q1                   | 76,1%        | 76,2%        | 73,8%        |
| Q2                   | 78,2%        | 79,5%        | 77,1%        |
| Q3                   | 78,2%        | 78,5%        | 77,8%        |
| Q4                   | 79,0%        | 79,1%        | 78,7%        |
| Q5                   | 77,2%        | 77,2%        | 77,7%        |
| Sin información      | 62,0%        | 59,2%        | 61,9%        |
| <b>Total</b>         | <b>69,4%</b> | <b>69,6%</b> | <b>70,9%</b> |

Fuente: Servicio de Información de Educación Superior (SIES).



Fuente: Servicio de Información de Educación Superior (SIES).