



**“Identificación de grupos de alumnos según variables académicas y socioeconómicas en las escuelas de pregrado de la Facultad de Economía y Negocios de la Universidad de Chile”**

**Seminario para optar al título de Ingeniero Comercial, Licenciatura en Ciencias de la Administración**

**Alumna:**

**Francisca Del Solar Rosselot**

**Profesor Guía:**

**César Ortega Gutiérrez**

**Santiago, Enero 2016**

**Universidad de Chile**

**Facultad de Economía y Negocios**

**Escuela de Economía y Administración**

## Tabla de contenido

1. Introducción.....	5
2. Revisión de literatura.....	7
2.1. Contexto Data Mining (DM) .....	7
2.1.1. Definición Data Mining .....	7
2.1.2. Metodologías que aplican DM .....	9
2.2. Minería de datos en la Educación (EDM) .....	29
2.2.1. Historia y evolución .....	29
2.2.2. Principales Enfoques EDM.....	32
2.2.3. Principales Aplicaciones EDM.....	33
2.2.4. Objetivos .....	38
2.2.5. Diferencia entre EDM y otras disciplinas de DM .....	39
2.2.6. Desafíos y Tendencias.....	40
3. Estudio en Estudiantes FEN de la Universidad de Chile.....	42
3.1. Objetivos del Estudio.....	42
3.2. Metodología de Investigación .....	43
3.3. Pregunta de Investigación .....	56
4. Resultados del Estudio.....	56
5. Conclusiones y Propuestas.....	80
Bibliografía.....	82
Anexos.....	84
Anexo 1 .....	84
Anexo 2 .....	85
Anexo 3 .....	85
Anexo 4 .....	86
Anexo 5 .....	86
Anexo 6. ....	87

## Índice de Tablas

Tabla 1. Clasificación de Aplicaciones de EDM.....	38
Tabla 2. Porcentajes de Reprobación por Semestres de Inscripción de Cursos .....	45
Tabla 3. Cantidad de Ud/Créditos Reprobados. ....	46
Tabla 4. Porcentajes de Reprobación histórica por cursos.....	46
Tabla 5. Cantidad de créditos según semestres para la Malla Nueva .....	48
Tabla 6. Cantidad de Ud según semestres para la Malla Antigua .....	49
Tabla 7: Atributos Seleccionados.....	50
Tabla 8. Tramo según cantidad de créditos y de Ud.....	51
Tabla 9. Variables diferenciadoras más relevantes, EEA-2SEM .....	58
Tabla 10. Variables diferenciadoras más relevantes, ESIA-2SEM .....	60
Tabla 11. Variables diferenciadoras más relevantes, EEA-3SEM .....	63
Tabla 12. Variables diferenciadoras más relevantes, ESIA-3SEM .....	64
Tabla 13. Variables diferenciadoras más relevantes, EEA-4SEM .....	66
Tabla 14. Variables diferenciadoras más relevantes, ESIA-4SEM .....	68
Tabla 15. Variables diferenciadoras más relevantes, EEA-5SEM .....	70
Tabla 16. Variables diferenciadoras más relevantes, ESIA-5SEM .....	72
Tabla 17. Variables diferenciadoras más relevantes, EEA-6SEM .....	74
Tabla 18. Variables diferenciadoras más relevantes, ESIA-6SEM .....	75

## Índice de Ilustraciones

<b>Ilustración 1.</b> Visión General de los Pasos que Componen el Proceso de KDD, extraído del artículo de Fayyad, Piatestsky-Shapiro y Smyth, 1996 (pág. 41) .....	13
<b>Ilustración 2.</b> Esquema de los 4 niveles de CRISP-DM ([CRISP-DM, 2000]).....	15
<b>Ilustración 3.</b> Diagrama de proceso que muestra la relación entre diferentes fases de CRISP-DM. Extraído del libro CRISP-DM 1.0, 2000 (pág. 10). ....	19
<b>Ilustración 4.</b> Descripción de Áreas de DM actuales y Técnicas para minar distintos tipos de datos, extraída de Venkatadri, M., y Lokanatha, C., Reddy de su artículo de 2011 “A Review on Data Mining from Past to the Future” (Pág. 20) .....	26
<b>Ilustración 5.</b> Tendencias en Data Mining, extraída de Venkatadri, M., y Lokanatha, C., Reddy, en su artículo de 2011 “A Review on Data Mining from Past to the Future”. (Pág. 21 y 22).....	27
<b>Ilustración 6.</b> Número de artículos publicados hasta el año 2009, agrupados de acuerdo al año. Extraído de Romero, C., y Ventura, S., en su artículo de 2010 “Educational Data Mining: A Review of the State of the Art”. (Pág. 603) .....	31
<b>Ilustración 7.</b> El Ciclo de Aplicación de la Minería de Datos en los Sistemas Educativos, extraído de Romero, C., y Ventura, S., en su artículo de 2007 “Educational data mining: A survey from 1995 to 2005”. (Pág.136).....	33
<b>Ilustración 8.</b> Evolución del porcentaje de reprobaciones según grupo de dependencia a través de los cinco semestres - EEA .....	76
<b>Ilustración 9.</b> Evolución del Factor de Avance según grupo de dependencia .....	77
<b>Ilustración 10.</b> Factor de Rendimiento entre las Escuelas a través de los semestres .....	78
<b>Ilustración 11.</b> Porcentaje de reprobación entre las Escuelas a través de los semestres .....	79

## 1. Introducción

La toma de decisiones en cualquier aspecto de la vida del ser humano, está ligada a las preferencias y a la racionalidad de éste (Hastie, 2001). En el caso de la toma de decisiones de los alumnos en la educación superior, estas tienden a estar influenciadas por la poca madurez y poca experiencia que los estudiantes poseen. Por ejemplo, un alumno no relaciona el tiempo, esfuerzo y el nivel intelectual necesario para enfrentar de manera exitosa el nivel de la carga académica que inscribirá en un semestre determinado. Estas decisiones pueden desencadenar consecuencias no deseadas a lo largo de su etapa universitaria, tales como constantes reprobaciones a lo largo de la carrera, llegando incluso a la deserción o eliminación académica de esta.

Por lo anterior, el principal objetivo de este estudio es lograr identificar cuáles son los distintos grupos de estudiantes diferenciados por semestre de avance al momento de inscribir sus cursos. Sumado a lo anterior, se busca encontrar características comunes y relevantes que identifiquen a estos grupos. Todo lo anterior a partir de la utilización de las herramientas de minería de datos y la metodología KDD.

Para lograr a los objetivos de esta investigación, se utilizaron herramientas de administración de datos como Access y SPSS Modeler el cual es una herramienta de Minería de datos, para encontrar las variables más relevantes y luego generar los distintos grupos de estudiantes. Con la primera herramienta se exploró y administró la muestra de datos con la que

se contaba. Luego, se utilizó el programa SPSS Modeler para lograr generar conocimiento mediante los algoritmos y métodos que este software posee, y de esta forma encontrar las características que describen a los grupos de alumnos de la Facultad de Economía y Negocios.

La estructura del informe consta inicialmente de una introducción de lo que es la Minería de Datos, cuáles son sus metodologías más importantes y en qué áreas se ha desempeñado. Luego da a conocer la Minería de Datos en la educación, finalizando este apartado se expone un estudio relacionado con recomendación a los estudiantes al momento de inscribir los cursos. En la sección siguiente se presenta la metodología escogida (KDD) y se aplica al estudio en sí, para luego proseguir con la exposición de los resultados obtenidos, describiendo cada uno de los clústeres encontrados por semestre de avance. Para finalizar este trabajo se entregarán las conclusiones.

Se espera que este estudio entregue los cimientos para futuras investigaciones y que los resultados obtenidos sean utilizados para profundizar esta área de investigación. Por otro lado, se busca seguir creciendo en el uso de la herramienta de minería de datos en un campo muy importante como lo es la educación superior. Lo anterior con el objetivo de encontrar resultados que generen impactos y mejoras tanto en los sistemas educacionales como en los resultados académicos de los alumnos.

## 2. Revisión de literatura

### 2.1. Contexto Data Mining (DM)

En esta sección se hará un recorrido de la Minería de datos, MD o DM en inglés, desde su definición hasta los desafíos futuros que tiene en la actualidad. Se destacará la gran evolución que ha tenido esta corriente desde sus inicios y con ello la inmensa difusión que se ha logrado alcanzar, llegando a áreas como la banca, clasificación de imágenes médicas, juegos, Ingeniería, entre otras.

La generación continua de datos es una realidad a la que nos enfrentamos día a día, donde el crecimiento exponencial sobrepasa las capacidades humanas de retenerlos y más aún de utilizarlos. Bajo este contexto, surge la necesidad de crear nuevos métodos y herramientas para que todos los actores de esta sociedad del conocimiento logren beneficiarse de que se encuentren los datos disponibles, generando y extrayendo información útil con el posterior objetivo de transformarlo en conocimiento. Uno de sus principales retos es trabajar con grandes volúmenes de datos, los que pueden contener sus propios problemas, como ruido, datos faltantes, volatilidad, entre otros

#### 2.1.1. Definición Data Mining

El concepto de minería de datos no es nuevo. A partir de los años setenta se inició con la utilización de este término, iniciándose bajo la idea de encontrar correlaciones sin una hipótesis previa en bases de datos con distintos tipos de ruido. Ya en los años ochenta, autores como Rakesh Agrawal, Gio

Wiederhold, Robert Blum y Gregory Piatetsky-Shapiro, entre otros, empezaron a consolidar los términos de Data Mining y Knowledge Discovery in Databases (KDD).

Vemos que el concepto ya tiene varias décadas, pero ha tomado una mayor fuerza estos últimos años dado el gran avance que ha experimentado la tecnología y su exponencial crecimiento.

En línea con lo anterior, se realizará un pequeño recorrido de las distintas definiciones que se le han asignado al concepto de minería de dato, nombrando las más relevantes.

La minería de datos es el descubrimiento eficiente de los patrones previamente desconocidos en grandes bases de datos (Gartner Group 1994) (Business Week 1994). [1]

Los autores Fayyad, Gregory Piatetsky-Shapiro y Padhraic Smyth (1996) [2] describen la minería de datos como una etapa de descubrimiento en el proceso de KDD, la cual consiste en el *uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos pre procesados*.

Por otro lado, según Margaret Dunham (2002) [3] minería de datos se define como: *"La extracción no trivial de información implícita, previamente desconocida y potencialmente útil a partir de los datos"*.

Por su parte, el Instituto SAS entiende el concepto de DM como el proceso de *seleccionar, explorar, modificar, modelar y valorar* grandes cantidades de

datos, con el objetivo de descubrir patrones desconocidos. Este proceso es resumido con la sigla SEMMA [4].

Dada las definiciones, podemos decir que minería de datos se entiende como un campo multidisciplinario que incluye diversas áreas como el aprendizaje automático, el reconocimiento de patrones, la estadística y las bases de datos. En definitiva, DM consiste en analizar un conjunto de datos para luego aplicar algoritmos adecuados y con ello obtener nuevos patrones a partir de los datos iniciales, con la premisa que estos patrones deben ser útiles.

### 2.1.2. Metodologías que aplican DM

La minería de datos es una disciplina que ha ido creciendo con el paso de los años. Las organizaciones han entendido que los grandes volúmenes de datos que poseen en sus sistemas pueden ser analizados y explotados para la obtención de nuevo conocimiento. El principal objetivo de este campo es encontrar información oculta o implícita, que no es posible obtener mediante métodos estadísticos convencionales.

Los principales esfuerzos en el área de minería de datos se han centrado en la investigación de técnicas para la explotación de información y extracción de patrones. Sin embargo, se ha desarrollado en menor medida el hecho de cómo ejecutar este proceso hasta obtener un “nuevo conocimiento”, es decir, en las metodologías.

Muchos de los modelos conocidos en la actualidad como metodologías son en realidad *modelos de proceso*, definidos como un conjunto de actividades y tareas organizadas para llevar a cabo un trabajo. La principal diferencia que tiene este último con el concepto de metodología, es que el modelo establece *qué hacer* y la metodología especifica por su parte *cómo hacerlo*. [14]

En 1996, el modelo KDD se constituyó como el primer modelo aceptado en la comunidad científica que estableció las etapas principales de un proyecto de explotación de información.

Ya en el año 2000, junto al gran crecimiento en el área de minería de datos, surgieron tres nuevos modelos que plantean un enfoque sistemático para llevar a cabo el proceso: SEMMA, Catalyst y CRISP-DM.

Considerando por un lado la antigüedad y por el otro, su transversalidad en las disciplinas, considerándose la metodología más utilizada según un estudio publicado en el año 2007 por la comunidad KDnuggets (Data Mining Community's Top Resource), se profundizará a continuación en los modelos KDD y CRISP-DM, respectivamente.

#### 2.1.2.1. *KDD*

Una definición precisa sobre descubrimiento del conocimiento en bases de datos (KDD) es la de Fayyad, Piatetsky-Shapiro y Smyth en su artículo de 1996 [2] que describe al proceso como: "Proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y entendibles en los

datos". Profundizando en esta definición, "proceso" implica que el KDD comprende diversos pasos, los que incluyen la preparación de los datos, la búsqueda de patrones, la evaluación del conocimiento y el refinamiento, todos estos pueden repetirse en múltiples iteraciones. El concepto, "no trivial" hace referencia a que el proceso involucra investigación e inferencia, no es una simple y directa aplicación de cálculos sobre la data. Finalmente, por "válido" se entiende que los patrones descubiertos deben seguir siendo precisos, para integrar datos nuevos que representen un aporte sobre algo desconocido por el investigador.

Como se mencionó anteriormente, el proceso KDD posee diversas etapas, dentro de los cuales es importante mencionar, según el trabajo realizado por Brachman y Anand (1996), los siguientes nueve [5]:

**Etapa 1:** *Desarrollar y entender el dominio de aplicación*, indagar cuales son los conocimientos previos relevantes a saber e identificar el objetivo del proceso KDD.

**Etapa 2:** *Creación o selección de un conjunto de datos objetivos*, sobre los cuales se ejecutará el descubrimiento.

**Etapa 3:** *Pre-procesamiento de datos*, en el que se incluye limpieza de datos, recolección de información sobre el modelo y la definición de estrategias para manejar los datos faltantes.

**Etapa 4:** *Transformación de datos*, que incluye la reducción y proyección de los mismos, donde se trata de encontrar características útiles para

representar los datos según los objetivos de la tarea. Lo anterior se logra aplicando métodos de reducción o transformación de la dimensionalidad, con el objetivo final de reducir el número de variables.

**Etapa 5:** *Elección de la tarea de DM*, donde se pretende encontrar un método adecuado alineado a los objetivos que se quieren alcanzar, dentro de los cuales se encuentra el método de reducción, de clasificación, de regresión, de clustering, entre otras.

**Etapa 6:** *Elección de algoritmo de DM con el objetivo de hallar los patrones*, lo que implica decidir los modelos y parámetros apropiados, para luego alinearlos al criterio utilizado en el proceso KDD.

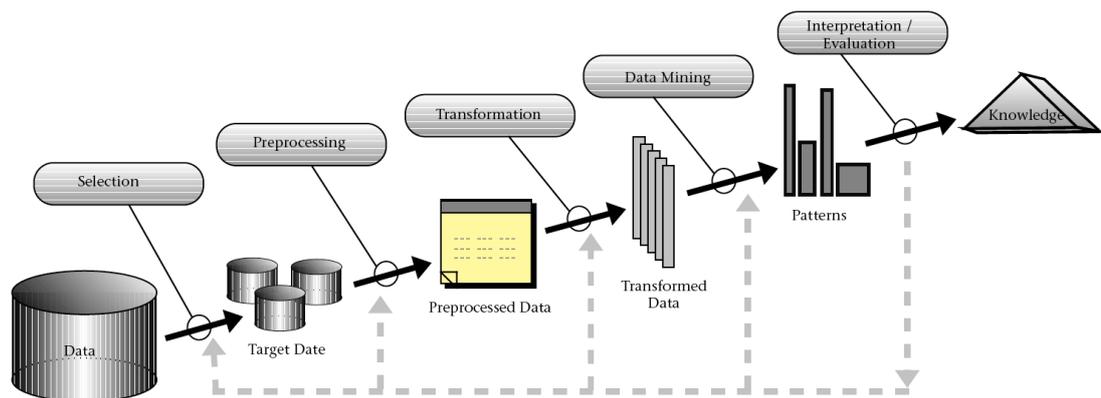
**Etapa 7:** *Data Mining*, en este subproceso se ejecuta el algoritmo, lo que conlleva a una búsqueda de patrones de interés representados de una manera particular, ya sea mediante reglas de asociación, arboles de decisión, clustering, entre otros.

**Etapa 8:** *Interpretación de patrones*, en esta etapa es posible visualizar los patrones extraídos, testeando su coherencia para finalmente validar los resultados del algoritmo con conocimiento experto. Cabe destacar la posibilidad que aquí se decida volver al primer paso o a una etapa previa.

**Etapa 9:** *Consolidación de conocimiento obtenido (difusión)*, en esta última etapa se obtiene el conocimiento en sí, incorporándolo a otro sistema para más acciones o simplemente para documentarlo.

A continuación se encuentra el diagrama de procesos del KDD, detallado anteriormente:

**Ilustración 1.** *Visión General de los Pasos que Componen el Proceso de KDD, extraído del artículo de Fayyad, Piatestsky-Shapiro y Smyth, 1996 (pág. 41)*



DM se considera por muchos autores un paso dentro del KDD (Knowledge Discovery in Databases), incluso, el más importante de estos. Según Margaret Dunham (2002) [3], los conceptos de KDD y DM están íntimamente relacionados, donde el primero se refiere al proceso de encontrar patrones en los datos y luego transformarlos en conocimiento útil que satisfaga los objetivos de investigación respectiva, y por otro lado el segundo hace alusión a usar algoritmos para extraer la información y los patrones derivados de ese proceso.

Concluyendo, vemos que ambos autores coinciden en que el KDD es un proceso que implica la utilización de bases de datos, y el Data Mining es una de las etapas de este proceso donde se utilizan algoritmos con el fin de encontrar patrones provenientes de los datos que respondan de una u otra

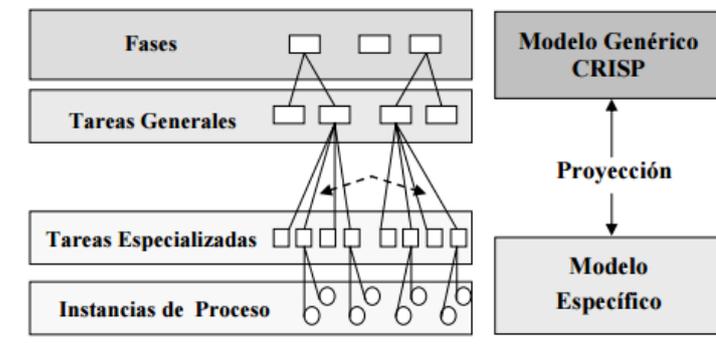
manera al problema planteado inicialmente por el investigador. Además es importante mencionar la existencia de una naturaleza iterativa e interactiva del descubrimiento del conocimiento, ya que el analista puede decidir si volver o seguir adelante al final de cada una de las etapas del proceso, es decir, es un proceso que envuelve numerosos pasos de la mano con diversas decisiones realizadas por el usuario o investigador.

#### *2.1.2.2. CRISP-DM*

CRISP-DM (*Cross Industry Standard Process for Data Mining*) fue creada por el grupo de empresas SPSS, NCR y Daimler Chrysler en el año 2000. Su creación surgió como manera de solucionar las falencias que tenía el Data Mining hasta ese año, siendo esta una guía de referencia para el desarrollo sistemático de proyectos de minería de datos. En la actualidad, es la metodología más utilizada en el desarrollo de proyecto de minería de datos.

CRISP-DM, está dividida en 4 niveles de abstracción organizados de forma jerárquica en tareas que van desde el nivel más general, hasta los casos más específicos y organiza el desarrollo de un proyecto de Data Mining, en una serie de seis fases.

**Ilustración 2.** Esquema de los 4 niveles de CRISP-DM ([CRISP-DM, 2000]).



La sucesión de fases no es necesariamente rigurosa. Como se aprecia en el esquema, cada una de las fases es estructurada en diversas tareas generales ubicadas en un segundo nivel. Estas se proyectan a tareas específicas, donde finalmente se describen las acciones que deben ser desarrolladas para situaciones específicas, pero en ningún momento se propone como realizarlas. En resumen, CRISP-DM establece un conjunto de tareas y actividades para cada fase que compone un determinado proyecto, pero no especifica cómo llevarlas a cabo.

Entrando de lleno a las fases que componen CRISP-DM, éstas son seis, donde la primera es probablemente una de las más importantes, pues el conocimiento y dominio del problema que se tenga en esta fase pueden depender el éxito o fracaso del proyecto. Más aún, no tomar las decisiones adecuadas durante la fase de comprensión de negocio, podría generar problemas tales como ineficiencias en decisiones a tratar en fases posteriores, incluso, acarrear el proyecto por direcciones no previstas.

Primeramente es importante mencionar que implica las flechas en el diagrama y la forma circular del modelo. Las primeras indican la

dependencia más relevante y frecuente entre las fases. El hecho de que este modelo este de manera circular, simboliza la naturaleza cíclica de la minería de datos. Ahora bien, son seis las fases que componen este modelo, según lo expresado en el libro CRISP-DM 1.0 (2000) [15]:

**Fase 1:** *Compresión del negocio.* Esta primera fase se centra en la compresión de los objetivos y requisitos del proyecto desde una perspectiva de negocio. El paso posterior es convertir este conocimiento en una definición del problema de minería de datos, y junto a ello la creación de un plan preliminar diseñado para alcanzar los objetivos.

En resumidas cuentas, las principales tareas a desarrollar en esta fase inicial son: determinar objetivos de negocio, valoración de la situación, realizar los objetivos de DM y realizar el plan de proyecto.

**Fase 2:** *Compresión de Datos.* Esta segunda fase comprende la recolección inicial de datos, con el objetivo de establecer un primer contacto con el problema, familiarizándose con ellos para identificar su calidad y establecer las relaciones más evidentes que permitan definir las primeras hipótesis de información oculta. Esta fase, junto a las próximas dos, son las que demandan el mayor esfuerzo y tiempo en un determinado proyecto DM.

En esta fase, las principales tareas son: recolectar los datos iniciales, descripción de los datos, exploración de los datos y verificación de la calidad de los mismos.

**Fase 3: Preparación de datos.** En esta fase, una vez efectuada la recolección inicial de datos, se procede a su preparación para adaptarlos a las técnicas de Data Mining que se utilicen posteriormente, tales como técnicas de visualización de datos, de búsqueda de relaciones entre variables u otras medidas para exploración de los datos.

La preparación de datos incluye las tareas generales de selección de datos a los que se va a aplicar una determinada técnica de modelado, limpieza de datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato. Estas tareas son ejecutadas en múltiples oportunidades y sin un orden establecido.

**Fase 4: Modelado.** En esta fase, se seleccionan y aplican diferentes técnicas de modelado y con ello se calibran los parámetros para obtener resultados óptimos. Una dificultad con la que nos podemos encontrar en esta fase, es el hecho de la existencia de técnicas que tiene requerimientos específicos para la forma de los datos, y por lo tanto es necesario volver a la fase de preparación de datos.

Las técnicas a utilizar en esta fase se eligen en función de los siguientes criterios:

- Ser apropiada al problema.
- Disponer de datos adecuados.
- Cumplir los requisitos del problema.
- Tiempo adecuado para obtener un modelo.
- Conocimiento de la técnica.

Las principales tareas que se llevan a cabo en esta fase son las siguientes: seleccionar la técnica de modelamiento, generar el plan de prueba, construir el modelo y por último evaluarlo.

**Fase 5: Evaluación.** En esta penúltima fase se evalúa el modelo. Es preciso revisar el proceso, teniendo en cuenta los resultados obtenidos, para repetir algún paso anterior en el que se haya posiblemente cometido un error. Es importante mencionar que en esta etapa se puede emplear múltiples herramientas para la interpretación de los resultados.

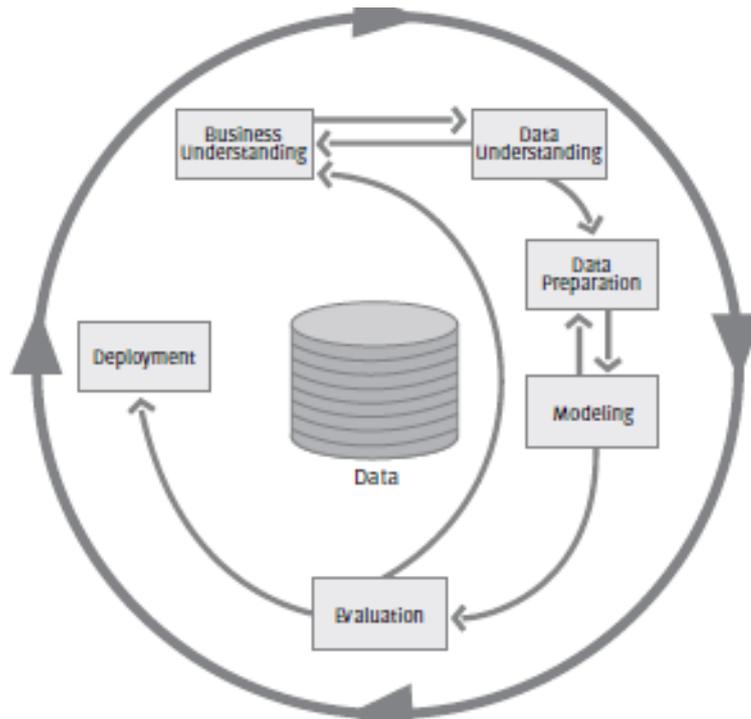
En definitiva, las tareas involucradas en esta fase del proceso son las siguientes: evaluar los resultados, revisión del proceso y la determinación de los pasos siguientes.

**Fase 6: Despliegue.** Esta última fase, una vez construido y validado el modelo, se transforma el conocimiento obtenido en acciones dentro del proceso de negocio. Esto depende de los requerimientos, pudiendo ser simples como la generación de un reporte o complejos como la implementación de un proceso de explotación de información que atraviese a toda la organización. En bastantes casos será el cliente final, y no el analista de datos, el que llevará a cabo esta última fase.

Las principales tareas llevadas a cabo en esta fase son las siguientes: plan de implementación, plan de monitoreo y mantenimiento, informe final y por último la revisión del proyecto.

A continuación se presenta el diagrama de proceso del modelo CRIPS-DM:

**Ilustración 3.** Diagrama de proceso que muestra la relación entre diferentes fases de CRISP-DM. Extraído del libro CRISP-DM 1.0, 2000 (pág. 10).



Resumiendo, podemos decir que CRISP-DM es un método probado para orientar los proyectos de minería de datos. Como metodología incluye las descripciones de fases normales de un proyecto, las tareas necesarias de cada fase y una explicación de las relaciones entre las tareas. Por otro lado, como un modelo de proceso, éste ofrece un resumen del ciclo de vital de minería de datos, explicado anteriormente.

Para finalizar, a pesar de la amplia aceptación y desarrollo que ha tenido el modelo CRIPS-DM en el marco de la minería de datos, éste todavía no

representa un proceso maduro que pueda definirse como una metodología establecida, es decir, que si bien establece un conjunto de tareas y actividades que deben ser llevadas a cabo en el proyecto, no establece con qué técnicas o modelos debe implementarse cada actividad (Marbán, 2003).

La principal diferencia entre estos modelos, es que CRISP-DM profundiza en mayor detalle las tareas y actividades a ejecutar en cada una de las etapas del proceso de la minería de datos. Por su parte, el KDD provee solo una guía general del trabajo a realizar en cada una de las fases.

Muchos investigadores del tema, se enfocan en la elección del algoritmo, es decir, ajustar modelos o determinar los patrones desde los datos disponibles. Estos modelos ajustados juegan el rol de “conocimiento inferido”. Dentro de este proceso de ajuste, existen dos tipos de formalismos matemáticos: estadísticas y lógica, donde según lo expresado por Fayyad, Piatetsky-Shapiro y Smyth (1996), el primero permite efectos no determinísticos en el modelo y el segundo es puramente determinístico. Por esta razón, el primer enfoque es el que usualmente se utiliza, dada la incertidumbre en los procesos que conlleva la generación de datos.

Un proyecto determinado de minería de datos posee diversas fases esenciales e imprescindibles para llevar a cabo esta etapa: Comprensión del negocio y del problema que se quiere resolver; Determinación, obtención y limpieza de los datos; Creación de los modelos matemáticos; Validación y comunicación de los resultados; Integración de los resultados.

### 2.1.3. Modelo data mining

Entrando más de lleno al DM, podemos ver que posee diversas tareas, estas se pueden dividir en dos categorías, las que predicen y las que describen. Las primeras tienen como objetivo estimar valores futuros o desconocidos a partir de alguna variable de interés, mientras que las segundas identifican patrones (correlaciones, tendencias, clústeres y trayectorias) que explican los datos.

Dentro de las tareas predictivas podemos ver la existencia de una variable “dependiente” que se pretende predecir y las “independientes” o “explicativas”, que son los atributos usados para la predicción. Por otro lado, las tareas descriptivas predicen patrones (correlaciones, tendencias, clústers y trayectorias) que resumen la relación existente entre los datos.

Dentro de estas dos categorías, podemos encontrar tareas más específicas de minerías de datos:

**Modelos predictivos:** Hace referencia a una construcción de un modelo para la variable dependiente como función de variables explicativas. Existen dos tipos de modelos, la clasificación (variables discretas) y la regresión (variables continuas).

- Clasificación: Consiste en la categorización de datos en grupos predefinidos o clases, que son determinados después de examinar la data. Los algoritmos de clasificación según lo expresado por Dunham (2002), requieren que las clases sean creadas en base a los valores de

determinados atributos, las que se describen por la observación de las características de los datos conocida su pertenencia a una clase específica.

- **Regresión:** Es utilizada para clasificar los datos en variables de predicción de valor real. Se asume que los datos objetivos son ajustables por medio de la utilización de algún tipo conocido de función y luego se determina la mejor función de este tipo que modela los datos disponibles.

**Reglas de asociación:** Son utilizadas para encontrar patrones que describen características fuertes de asociación entre los datos. Los patrones son representados en forma de regla. Además, es de suma importancia, verificar que las relaciones no sean casuales. El origen de esta tarea surge en respuesta a la necesidad de analizar las canastas de compra y la posible relación, existente entre los diversos productos.

**Análisis de conglomerado:** Conjunto de técnicas que permite clasificar a los objetos en grupos según homogeneidad de los mismos, por otro lado, cada grupo posee elementos con características que lo diferencian de los otros grupos. A cada uno de estos se les llama clúster o conglomerado. Se diferencia de la clasificación, ya que los grupos no han sido predefinidos, se definen por los mismos datos. Dado lo anterior, es necesario tener cierto conocimiento para interpretar los clústeres que fueron creados.

**Descubrimiento Secuencial:** Se utiliza para encontrar patrones secuenciales en los datos. Similares a las Reglas de Asociación, pero este se basa en el factor tiempo.

**Análisis de Series de Tiempo:** Se examina la variación del valor de un atributo en el tiempo. Se busca dejar todos los datos en la misma unidad de tiempo (diario, mensual, etc.).

**Summarization o Resumen:** Mapas de resumen de datos en sub-muestras con descripciones asociadas, donde solo se utiliza la información más relevante. También es conocida con el nombre de caracterización o generalización.

Se ve claramente que la minería de datos y el KDD provienen de múltiples disciplinas, las que evolucionan y se combinan para obtener lo mejor de ellas en la extracción del conocimiento útil de las bases de datos según el contexto en las que se utilicen. Destaca la administración de la base de datos, recuperación de la información, reconocimiento de patrones, visualización de datos, estadísticas, algoritmos, computación de alto desempeño, aprendizaje de máquinas, entre otros.

#### *2.1.4. Desafíos y tendencias*

Reafirmando que el Data Mining es la evolución de varias disciplinas, nacen las 2 principales tendencias de los inicios de este, según los autores Venkatadri y Lokanatha [6], la primera habla sobre la “Tendencia de Datos” en la cual se consideraba que los algoritmos de DM trabajaban mejor sobre una única base de datos de naturaleza numérica, para luego, gracias a la influencia estadística y técnicas de aprendizaje automático, se pudo trabajar con datos de diversos orígenes (no numérico) y de múltiples bases relacionadas. La segunda tendencia llamada, “Tendencia Computacional”,

indica que en un comienzo el DM basaba sus algoritmos generalmente solo en la estadística, hoy en día se basa en diversas técnicas computacionales como la “Inteligencia Artificial” o el “Reconocimiento de Patrones”. Así, podemos verificar que a través del tiempo ha existido una evolución que incorpora necesariamente a ambas tendencias, implementando diversas herramientas para trabajar de la manera más eficiente y eficaz con una base de datos variada, incorporando nuevas técnicas de DM, logrando de esta forma obtener un conocimiento útil para resolver problemas planteados desde un comienzo por el analista.

Dado los avances y logros que se han obtenido gracias a estos métodos, considerando las distintas áreas a las que han podido llegar, tales como salud, educación, retail, detección de fraude, entre otros varios, es que día a día nace el interés de diversos campos por incluirlo en sus estudios e incluso los que ya están buscan obtener información de mayor complejidad. Por lo anterior, surgen diversos desafíos para el DM, los que incluyen diversos aspectos, desde formatos de los datos a mayores complejidades de los negocios.

Los mismos autores, Fayyad, Piatetsky-Shapiro y Smyth (1996), y Margaret Dunham (2002), destacan los siguientes desafíos que se les presenta en la actualidad al Data Mining:

**Minería de Datos heterogéneos:**

- Hipertexto e Hipermedia DM: Colección de datos desde catálogos en línea, librerías digitales, y base de datos de información en línea que incluyen *hyperlinks*, marca de grupo de texto, y otras formas de datos. Existe una aplicación de DM encargada de descubrir patrones desde la web, llamada Web Mining. Las técnicas del DM que se ocupan en esta área son: clasificación y clustering.
- DM “Ubicuo”: Este campo tiene relación con la generación de datos que tienen su origen en datos móviles. Aquí se utilizan técnicas tradicionales de DM, específicamente la combinación de aprendizaje de máquinas y estadística.
- DM Multimedia: Los datos de este campo incluyen imágenes, videos, audio y animación. Las técnicas más aptas sobre este escenario son árboles de decisión basado en reglas, algoritmos de clasificación tales como redes neuronales artificiales, métodos de aprendizaje basado en instancias, máquinas de soporte vectorial, reglas de asociación y clustering.
- DM Espacial: Los datos de esta sección incluyen datos astronómicos y satelitales. Dentro de las técnicas utilizables se incluyen los sistemas OLAP (*on-line analytical processing*) y métodos de clustering, ambos espaciales.
- DM de Series de Tiempo: Este último punto aborda técnicas aplicadas a secuencias de datos medidos de manera sucesiva y en espacio de tiempos uniformes. Los algoritmos a utilizar suelen ser reglas de inducción o arboles de decisión.

**Utilización de los recursos computacionales:** Una de las principales características de DM, es la utilización de la computación avanzada y recursos de red, donde se encuentra la computación paralela, distribuida y de redes. Sobre los algoritmos utilizados, en esta área destacan los “algoritmo a priori”.

**Investigación y computación científica:** Dentro de la disciplina científica podemos encontrar la astronomía, bioinformática, química o imagenología médica, entre otras, las que destacan por la gran cantidad de datos que estas generan. Es por esta razón que el DM ha tenido que modificar y refinar sus distintas técnicas ad-hoc a esta disciplina.

**Tendencias de Negocios:** Esta área es la más desarrollada en el tema de DM, donde la mayoría de las aplicaciones se relacionan con técnicas de clasificación y predicción, con el objetivo de dar soporte a la toma de decisiones que enfrentan los trabajadores en los diversos cargos, al interior de la organización.

*Ilustración 4. Descripción de Áreas de DM actuales y Técnicas para minar distintos tipos de datos, extraída de Venkatadri, M., y Lokanatha, C., Reddy de su artículo de 2011 “A Review on Data Mining from Past to the Future” (Pág. 20)*

<b>Data mining type</b>	<b>Application Areas</b>	<b>Data Formats</b>	<b>Data mining Techniques/Algorithms</b>
Hypermedia data mining	Internet and Intranet Applications.	Hyper Text Data	Classification and Clustering Techniques
Ubiquitous data mining	Applications of Mobile phones, PDA, Digital Cam etc.	Ubiquitous Data	Traditional data mining techniques drawn from the Statistics and Machine Learning
Multimedia data mining	Audio/Video Applications	Multimedia Data	Rule based decision tree classification algorithms
Spatial Data mining	Network, Remote Sensing and GIS applications.	Spatial Data	Spatial Clustering Techniques, Spatial OLAP
Time series Data mining	Business and Financial applications.	Time series Data	Rule Induction algorithms.

Resumiendo, las tendencias más relevantes en la actualidad que destacan dentro del conjunto nombrado anteriormente son: Importancia alcanzada por los datos no estructurados; necesidad de integrar los algoritmos y resultados obtenidos en sistemas operaciones y sitios web; y el requerimiento de que los procesos funcionen en línea.

Por último, además de las tendencias actuales nombradas anteriormente, los autores Venkatadri y Lokanatha (2011), mencionan las futuras tendencias a las que se enfrentará el DM: Estandarización de Lenguajes en DM, Procesamiento de Datos, Objetos de Datos Complejos, Recursos Computacionales, Web Mining, Computación Científica y Tendencias de Negocios.

*Ilustración 5. Tendencias en Data Mining, extraída de Venkatadri, M., y Lokanatha, C., Reddy, en su artículo de 2011 "A Review on Data Mining from Past to the Future". (Pág. 21 y 22)*

<b>Data mining trends</b>	<b>Algorithms/ Techniques employed</b>	<b>Data formats</b>	<b>Computing Resources</b>	<b>Prime areas of applications</b>
<b>Past</b>	Statistical, Machine Learning Techniques	Numerical data and structured data stored in traditional databases	Evolution of 4G PL and various related techniques	Business
<b>Current</b>	Statistical, Machine Learning, Artificial Intelligence, Pattern Reorganization Techniques	Heterogeneous data formats includes structured, semi-structured and unstructured data	High speed networks, High end storage devices and Parallel, Distributed computing etc...	Business, Web, Medical diagnosis etc...
<b>Future</b>	Soft Computing techniques like Fuzzy logic, Neural Networks and Genetic Programming	Complex data objects includes high dimensional, high speed data streams, sequence, noise in the time series, graph, Multi-instance objects, Multi-represented objects and temporal data etc...	Multi-agent technologies and Cloud Computing	Business, Web, Medical diagnosis, Scientific and Research analysis fields (bio, remote sensing etc...), Social networking etc...

Como se ha mencionado anteriormente, Data Mining posee un gran número de áreas de aplicación, considerando relevante describir la que se utilizará posteriormente en este trabajo junto a otras dos clásicas.

- **Aplicación al Retail:** Una de las importantes y exitosas aplicaciones de DM es el conocido análisis de canastas de mercado. Por ejemplo, si un cliente compra determinado producto también llevará otro, cada vez que vaya a comprar el primer producto. También está presente la aplicación sobre la identificación del “Cliente Alfa”, la que entrega información pudiendo analizarla y modelar comportamientos predictivos.
- **Servicio de Salud:** El DM posee diversas herramientas utilizables en diversas ramas del área de la salud, lo que conlleva a obtener grandes beneficios para todas las partes involucradas.
- **Educación:** Esta es una de las áreas más recientes en la utilización de minería de datos. Se ha centrado en el desarrollo de métodos que utilicen principalmente datos de plataformas educativas y en el uso de esos métodos para comprender mejor a los estudiantes y el entorno en el que aprenden. Los métodos utilizados en esta área, tienden a diferir a los generalistas. Específicamente, se utiliza los métodos psicométricos integrados a los métodos de aprendizaje máquina y texto de minería de datos para lograr los objetivos planteados.

Es tal el grado de desarrollo y especialización del uso del DM, que algunas de sus ramas se valen por sí solas. El ejemplo más conocido, es el de la “Inteligencia de Negocios” o “Business Intelligence”, que hoy en día puede manejar grandes cantidades de información, logrando obtener ventaja competitiva en un negocio, debido a que permite y ayuda a identificar y desarrollar las oportunidades presentes en el entorno. Ahora bien, existen

otras disciplinas que sobresalen por su reciente aparición y junto a ello su exponencial crecimiento en la utilización de DM como herramienta de investigación, específicamente, destaca el área de educación, la cual ha crecido de manera significativa con su aparición hace menos de 20 años [7].

## 2.2. Minería de datos en la Educación (EDM)

Como se mencionó en la sección anterior, la minería de datos es un campo que nos permite descubrir información nueva y potencialmente útil proveniente de una gran cantidad de datos. Ésta se ha empleado en un variado abanico de campos, desde el campo del retail, específicamente la canasta de compra hasta la investigación contra el terrorismo.

Minería de datos en la educación se refiere al conjunto de técnicas, herramientas e investigación diseñados para la extracción automática del significado de grandes cantidades de datos generados o relacionados con las actividades de aprendizaje en los centros educativos [7].

### 2.2.1. Historia y evolución

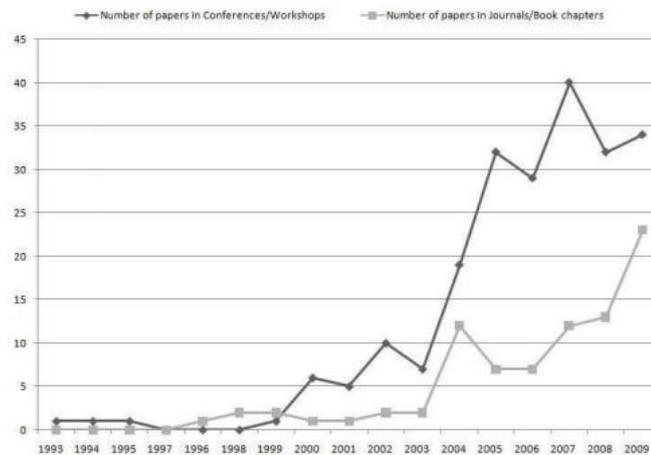
Dado el creciente interés en el desarrollo de técnicas para un análisis más detallado y preciso de las grandes cantidades de datos generados en los diversos centros educativos, nacen una serie de talleres desarrollados entre los años 2000 y 2007 en el marco de distintas conferencias internacionales. En el año 2008, un grupo de investigadores estableció un tipo de conferencia internacional anual sobre la investigación del EDM. Este mismo año, se introdujeron una serie de fuentes de datos educativos públicos, como la Pittsburgh Science of Learning's Centre, Datashop y el National Center for

Education Statistics, generando confianza y accesibilidad a los mismos y junto con ello un sustento para el desarrollo de esta rama. (Romero y Ventura, 2010)[9]

En el año 2010 se creó el “Manual de Minería de datos Educativos” con la participación de los autores C.Romero, S.Ventura, M. Pechenizky y R. Baker (Editorial CRC Press, Taylor & Francis Group). Un año después, nació como manera de consolidación, la Sociedad Internacional para la minería de datos educativos, con el fin de conectar a los investigadores del área e impulsar el crecimiento y desarrollo de la misma. Por último, en el año 2014 se lanzó otra revista llamada “Minería de Datos en la Educación: Aplicaciones y Tendencias” editado por A. Peña-Ayala (Springer, SCI Vol.524), con el objetivo de informar e intercambiar los resultados sobre las investigaciones de esta área.

El número de publicaciones sobre EDM ha experimentado un exponencial crecimiento en el último tiempo, lo cual se muestra en el siguiente gráfico donde se agrupan los artículos publicados del tema en cada año correspondiente, viéndose un claro salto del año 2005 al 2006.

**Ilustración 6.** Número de artículos publicados hasta el año 2009, agrupados de acuerdo al año. Extraído de Romero, C., y Ventura, S., en su artículo de 2010 "Educational Data Mining: A Review of the State of the Art". (Pág. 603)



A lo largo de la historia, EDM se ha desarrollado en dos grandes categorías o campos generales, las cuales son mencionadas por los autores Romero y Ventura (2007) [17]:

1. Estadística y visualización
2. Minería Web
  - Clustering, clasificación y detección de outliers.
  - Minería de reglas de asociación y minería de patrones secuenciales.
  - Minería de textos.

Por otro lado, la minería de datos puede ser aplicada a los datos provenientes de dos tipos de sistemas educativos: Educación tradicional y Educación a distancia. Las primeras están basadas en una relación cara a cara entre el estudiante y el profesor. En cambio los segundos, consiste en técnicas o métodos de acceso a los programas educativos para los

estudiantes que están separados por el tiempo y el espacio de sus profesores.

### 2.2.2. Principales Enfoques EDM

Dentro de los métodos más utilizados encontramos la predicción, el análisis de agrupaciones y la extracción de relaciones entre las variables, los que son universales para varias categorías. Ahora bien, las etapas del descubrimiento mediante modelos y la destilación de datos para el juicio humano se consideran predominantes dentro de esta rama [16].

Para clarificar los enfoques mencionados, se describirá los dos más relevantes según los autores Álvarez y Jiménez, para el área a la que se enfoca el siguiente trabajo:

**Descubrimiento mediante Modelos:** En este enfoque, se desarrolla un modelo mediante predicción, agrupamiento o ingeniería del conocimiento (usando métodos de razonamiento humano). Por ende, este modelo es utilizado como componente de otro análisis, como predicción o minería de datos. En el primer caso, las predicciones encontradas por el modelo creado son utilizadas como insumos o variables de entrada en las predicciones de una futura variable. En el caso de minería de relaciones, se estudian estas entre las predicciones del modelo creado y variables adicionales.

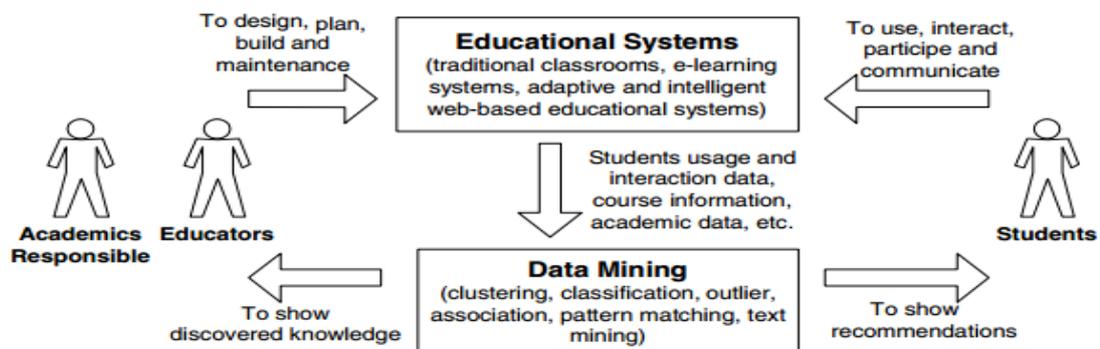
**Destilado de datos:** El objetivo principal es resumir y presentar la información de manera más atractiva e interactiva. Los datos son destilados para la interpretación humana por dos motivos claves: identificación y

clasificación. Cuando se realiza para lograr la identificación, los datos se muestran de manera que una persona puede identificar patrones conocidos, difíciles de expresar formalmente. Por otro lado, cuando se realiza para la clasificación, etiquetado manual, se pretende dar soporte a futuros modelos predictivos.

### 2.2.3. Principales Aplicaciones EDM

La aplicación de la minería de datos en el sistema educativo es un ciclo iterativo de formación de hipótesis, testeo y refinamiento. Esta puede estar orientada a diferentes participantes, cada uno visto desde un punto de vista en particular.

*Ilustración 7. El Ciclo de Aplicación de la Minería de Datos en los Sistemas Educativos, extraído de Romero, C., y Ventura, S., en su artículo de 2007 "Educational data mining: A survey from 1995 to 2005". (Pág.136)*



Este campo es integrado por dos principales participantes: los estudiantes y los instructores. Pero es posible encontrar a cinco participantes, donde cada uno posee un grado de relevancia particular, pero todos son imprescindibles para el buen desarrollo de esta rama. La conformación parte por los estudiantes, seguido por los educadores, los investigadores, las organizaciones y finalmente los administradores.

En cuanto a los estudiantes, el interés está en la comprensión de sus necesidades y utilizar métodos para mejorar experiencias y rendimientos educativos de los mismos. Los investigadores se centran en el desarrollo y la evaluación de las técnicas de minería de datos para la eficacia. Los administradores son los responsables de asignar los recursos para la implementación de los resultados en las diversas instituciones. Finalmente se encuentran los instructores, tutores, educadores o profesores, en sus distintos nombres. Estos intentan comprender el proceso de aprendizaje y con ello conocer los métodos que pueden utilizar para mejorar y eficientar sus formas de enseñanzas. Además según los autores Romero y Ventura (2010) [9], entre los objetivos que destacan para el uso de la minería de datos en este miembro están la obtención de realimentación objetiva sobre una instrucción, analizar el comportamiento de los estudiantes, identificar que estudiante requiere apoyo, predecir el desempeño del mismo, entre otras.

Por otro lado, la minería de datos en la educación también puede clasificarse según el tipo de dato que se utiliza o en otras palabras, en el medio ambiente en el cual se desarrolla. Según la tabla presentada por los autores Romero y Ventura (2010)<sup>1</sup>, existen 8 grandes grupos dentro de los cuales se clasifican las distintas referencias sobre EDM: Educación Tradicional, Educación basada en la Web (E-learning), Sistemas de Gestión de Aprendizaje, Sistemas Tutoriales Inteligentes, Sistemas Educativos Adaptativos, Pruebas/Cuestionarios, Textos/Contenidos y otros.

---

<sup>1</sup> Anexo 1: Tabla clasificatoria de estudios en EDM

Como último tipo de clasificación, en el mismo artículo “Educational Data Mining: A Review of the State of the Art” (2010), los autores Romero y Ventura elaboraron una lista de aplicaciones del EDM. Estas son once y se explicará el objetivo principal de cada una a continuación.

### **Aplicaciones EDM y sus principales objetivos:**

- A. Análisis y visualización de datos: El objetivo es realzar la información útil y toma de decisiones de apoyo. Las dos tareas más utilizadas son estadística y visualización de información.
- B. Proporcionar información para apoyar a los instructores (Feedback): El objetivo es proveer retroalimentación para apoyar la toma de decisiones de autores/profesores/administradores, específicamente acerca de cómo mejorar el aprendizaje de los estudiantes, organizar los recursos didácticos de manera más eficiente, entre otras, con el objetivo de que estos puedan adoptar medidas proactivas y / o correctivas apropiadas.
- C. Recomendar cursos o actividades a los estudiantes: El objetivo es ser capaz de hacer recomendaciones directamente a los estudiantes, generando nuevo conocimiento que pueda ser utilizado para hacer recomendaciones tales como la siguiente tarea, visita o problema a realizar. Por otro lado, este conocimiento puede ser utilizado para restringir el contenido, las interfaces y las secuencias de aprendizaje de cada alumno individualmente.

Los métodos utilizados en esta tarea son los de clasificación, los de asociación, los de clustering y últimamente el análisis de redes sociales.

- D. Predicción del rendimiento: El objetivo consiste en estimar el valor de una variable que describe el rendimiento futuro del estudiante a partir de cierta información disponible: Información histórica (evaluaciones previas) u otra información relacionada (social, actitudinal, ambiental, etc.).

Las principales técnicas con las que se ha abordado esta tarea son clasificación y predicción.

- E. Modelar el comportamiento del estudiante: El objetivo es desarrollar modelos cognitivos de los estudiantes, incluyendo un modelado de sus habilidades (destrezas) y el conocimiento declarativo.

Uno de los modelos más populares para representar el comportamiento de estudiantes son las redes bayesianas.

- F. Detección de comportamientos no deseados en estudiantes: El objetivo es descubrir / detectar aquellos estudiantes que tienen algún tipo de problema o comportamiento inusual, tales como: realización de acciones erróneas, baja motivación, el mal uso de los recursos, intento de hacer trampas en el sistema, abandono, fracaso escolar, etc.

Los métodos utilizados son clasificación, método de detección de anomalías, minería de datos de asociación y descubrimiento de subgrupos.

- G. Agrupación de estudiantes: El objetivo es crear grupos de estudiantes de acuerdo a sus características personalizadas, características personales, etc.
- H. Análisis de redes sociales: El Objetivo es el estudio de las relaciones entre los individuos, en lugar de atributos o propiedades individuales.
- I. Generación automática de mapas conceptuales: El objetivo es ayudar a los instructores/educadores en el proceso automático de desarrollar / construir mapas conceptuales. Estos se utilizan para representar gráficamente conceptos o ideas que tiene una relación jerárquica. Se trata de una forma estructurada de visualizar la información más relevante sobre un determinado tema.
- Se han utilizado dos tipos de técnicas para la generación de mapas de manera automática: minería de reglas de asociación y minería de textos.
- J. Planificación y programación: El objetivo es para mejorar el proceso educativo tradicional mediante la planificación de cursos futuros, ayudando con el itinerario de los cursos de cada estudiante, la planificación de la asignación de recursos, lo que ayuda en los procesos de admisión y de asesoramiento, desarrollo de plan de estudios, etc.
- K. Construcción de cursos: El objetivo es ayudar a los instructores y desarrolladores para llevar a cabo el proceso de construcción / desarrollo de cursos y aprender contenidos de forma automática.

Los autores clasifican estas once tareas en cinco objetivos según los resultados generales a los que se llegan en cada una. Lo anterior se resume en la siguiente tabla:

*Tabla 1. Clasificación de Aplicaciones de EDM*

<b>Objetivo</b>	<b>Tarea(s)</b>
Proporcionar información para los instructores	A y B
Proporcionar información para los estudiantes	C
Revelan características de los estudiantes	D, E, F y G
Estudian gráficos y relaciones entre los estudiantes y los conceptos	H y I
Ayudan a crear y/o planear los cursos	J y K

Extraído del artículo “Educational Data Mining: A Review of the State of the Art” (2010), de los autores Romero & Ventura. Elaboración propia.

#### 2.2.4. Objetivos

Los autores Baker y Yacef [8] identificaron los siguientes objetivos del EDM:

- *Predicción de los comportamientos de aprendizaje de los alumnos en el futuro.* El logro de este objetivo puede obtenerse mediante la creación de modelos de los estudiantes que incorporan las características de los mismos, incluyendo información detallada de su conocimiento, comportamientos y diversas motivaciones para aprender.
- *Descubrimiento o mejora de los modelos ya existentes en el campo.*
- *Estudio de los efectos del apoyo educativo,* logrado a través de los sistemas de aprendizaje.

- *Fomento del conocimiento científico sobre el aprendizaje y los alumnos,* mediante la construcción y la incorporación de los modelos de estudiantes, el campo de investigación en EDM y la tecnología y software utilizados.

#### 2.2.5. Diferencia entre EDM y otras disciplinas de DM

El proceso de DM es mayormente estandarizado para las distintas áreas ya que posee pasos definidos de forma transversal al campo al que se aplique.

Existen ciertos aspectos en los cuales se diferencia el EDM de los otros campos. Cristóbal Romero y Sebastián Aventura [9], destacan los siguientes:

- **Objetivos:** Dentro de los principales objetivos que tienen la minería de dato en sistemas educativos es mejorar los procesos de aprendizaje y orientar los aprendizajes de los estudiantes. Sumado a esto, se busca lograr una comprensión más profunda de los fenómenos educativos.
- **Datos:** Se requieren datos específicos del área de educación y por ende poseen información con significado intrínseco, relacionada con otros datos y múltiples niveles de jerarquía significativa.
- **Técnicas:** Existen técnicas que se pueden aplicar de manera directa a los datos, pero otras deben ser adaptadas. Por otro lado, existen técnicas específicas que pueden solo ser usadas en problemas relacionados con la educación.

### 2.2.6. Desafíos y Tendencias

El hecho de que exista un almacenamiento de datos educacionales, genera diversas ventajas para la investigación de esta rama, por ejemplo, el ahorro en tiempo que les genera a los analistas en las etapas de búsquedas y recopilación de datos, ya que como se mencionó en un inicio, estos se encuentran disponibles y accesibles.

Históricamente, utilizando los métodos tradicionales, el estudio de cómo la diferencia entre profesores o las diferencias de metodologías de clases tienen impacto sobre los aprendizajes particulares de los alumnos, ha sido bastante complejo, facilitándose en el último tiempo gracias a la minería de datos, según lo mencionado por A. Jiménez y H. Álvarez (2010) [16].

Entonces se puede decir que la minería de datos en la educación es una línea que nació para complementar los estudios realizados por los métodos tradicionales, como estudios psicológicos, sociológicos, entre otros. Ésta busca recoger la experiencia al interior de los procesos de aprendizaje, a través de la recopilación de los datos que se generen en estos, con la finalidad de recrear ese momento y realizar predicciones futuras.

Como toda línea investigativa, la minería de datos en la educación posee algunos puntos débiles en la actualidad, donde destacan en primer lugar la generalización. Este punto se refiere a que los resultados obtenidos de un estudio no se pueden extrapolar a otras instituciones. Por otro lado, el cuidado de privacidad de los datos que se utilizan en este tipo de investigación, ya que el fácil acceso a las herramientas puede poner en

riesgo a los estudiantes que proporcionaron algún tipo de información. Por último, al trabajar con datos e información que afectan de manera directa la vida de las personas, exige actuar con responsabilidad y prudencia, siendo necesario establecer un código de ética.

#### *2.2.6.1. Investigaciones Futuras*

Cada día esta área toma mayor fuerza y desarrollo en las investigaciones relacionadas al sistema educativo, por ende, año a año es necesario ir mejorando los diversos aspectos que la componen.

Destaca los siguientes puntos a indagar:

- Búsqueda y desarrollo de herramientas de minería de datos más fáciles de usar para los educadores.
- Necesidad de estandarización de métodos y datos relacionados con la educación.
- Integración con los sistemas E-learning.
- Desarrollar técnicas específicas de la minería de datos destinadas al uso investigativo de los sistemas educacionales.

En resumen, finalizando el capítulo de EDM, se ve claramente como esta rama ha evolucionado y creciendo con el pasar de los años, iniciando su recorrido en el año 1995 hasta la actualidad. Sumado a lo anterior, se ven distintos enfoques en las que esta línea debe enmarcar sus esfuerzos en la actualidad, para seguir entregando las herramientas necesarias y

adecuadas a cada futuro usuario que requiera utilizar esta ciencia en el desarrollo de su investigación.

A pesar del interés creciente de la utilización de las técnicas de minería de datos en el contexto educativo, en la actualidad existen pocos trabajos que han utilizado dichas herramientas de manera concreta a lo largo de su investigación. A partir de lo anterior, se expondrá a continuación uno de estos estudios, realizado en el año 2009, sobre una de las aplicaciones del EDM, llamado: “Recomendación en la Educación Superior mediante Técnicas de Minería de Datos” por los autores C. Vialardi, J. Bravo, L. Shafti y A. Ortigosa. [19]

### **3. Estudio en Estudiantes FEN de la Universidad de Chile**

#### **3.1. Objetivos del Estudio**

El presente estudio tiene por objetivo identificar y describir a los distintos grupos de alumnos de pregrado de las carreras de Ingeniería Comercial, Auditoría e Ingeniería en Información y Control de Gestión de la Facultad de Economía y Negocios de la Universidad de Chile (FEN), agrupados por semestre de avance académico al momento de inscribir cursos en un semestre determinado.

Además, se busca encontrar las características más relevantes que distinguen a estos grupos, para finalmente analizar las semejanzas y diferencias que se forman entre los mismos.

### 3.2. Metodología de Investigación

Este estudio se realizó utilizando la metodología del KDD mencionada en el primer capítulo. Esta técnica consta de 9 etapas, las cuales fueron desarrolladas y aplicadas al contexto del estudio presente (página 11 y 12).

**Etapa 1:** En esta etapa se identificó como principal objetivo analizar un conjunto de estudiantes de FEN, y con ello encontrar grupos de alumnos que posean atributos específicos diferenciadores y comunes entre los distintos periodos de inscripción de cursos. Dada la magnitud de datos existentes en el estudio, base de datos de la facultad, se decidió utilizar una herramienta de minería de datos, específicamente el KDD, considerando ser la metodología más adecuada dado del contexto y los objetivos que se quieren alcanzar.

**Etapa 2:** Con el objetivo de encontrar y distinguir los distintos grupos de alumnos por semestre de avance al momento de inscribir ramos, se utilizó la herramienta Microsoft Access para administrar la base de datos y el software SPSS Modeler como herramienta de minería de datos.

Los datos provienen de dos fuentes de información específicas; la primera es el Sistema de Administración Docente FEN y (2) Base del Departamento de Evaluación, Medición y Registro Educativo (DEMRE). La primera fuente de información contiene datos académicos de los estudiantes en su paso por la FEN. Mientras que la segunda fuente de información entrega los datos socio-demográficos que declaran los estudiantes al momento de rendir la prueba de selección universitaria (PSU), previo al ingreso a la Facultad.

Se consideró en la base de datos a analizar, a los alumnos que participaron en el proceso de inscripción de cursos de siete semestres, iniciando en el semestre de otoño 2012 y finalizando en otoño 2015. Para cada uno de estos procesos de inscripción de cursos los estudiantes se encuentran en distintos semestres de avance según la cantidad de Unidades Docentes (Ud) / créditos acumulados, cursados y aprobados, al momento de la inscripción de cursos. Los alumnos de la muestra pueden encontrarse dentro de diez semestres, según lo establecido en la malla curricular de la facultad.

Es importante mencionar, que el año 2012 la facultad realizó una reforma curricular, en la cual se realizaron diversas modificaciones a los planes de estudio de las tres carreras existentes. Dentro de los ajustes realizados, se encuentra la modificación de la unidad de medida de las horas de dedicación de cada uno de los cursos que componen las carreras, cambiando de Ud a de créditos.

Inicialmente se identificarían grupos para los diez semestres que tienen las tres carreras pertenecientes a la FEN. Sin embargo, luego de un análisis y una discusión entre el equipo de investigación, se acordó realizar el estudio sólo para los semestres dos, tres, cuatro, cinco y seis.

Lo anterior se fundamenta, en que los semestre dos al seis son los más complejos para los alumnos debido a que las mayores tasas de reprobación y en las mayores tasas de deserción se encuentran en dichos semestres mencionados anteriormente.

A partir de los datos proporcionados por la Secretaria de Estudios de FEN, se realizó un cálculo sobre el promedio de las reprobaciones por semestre en siete inscripciones de cursos, desde otoño del 2012 a otoño del año 2015.

*Tabla 2. Porcentajes de Reprobación por Semestres de Inscripción de Cursos*

	1er semestre	2do semestre	3er semestre	4to semestre	5to semestre	6to semestre	7mo semestre	8vo semestre	9vo semestre	10mo semestre
20151	16,40%	25,86%	14,32%	14,25%	11,13%	10,84%	7,30%	7,53%	3,27%	1,64%
20142	57,38%	15,20%	12,44%	10,61%	10,97%	8,50%	6,70%	6,88%	4,91%	2,03%
20141	16,18%	18,09%	11,48%	12,84%	10,81%	9,59%	8,09%	7,22%	4,61%	0,66%
20132	60,43%	17,74%	13,51%	12,91%	7,68%	9,15%	8,73%	5,97%	3,39%	0,77%
20131	11,73%	14,97%	11,77%	8,08%	9,76%	14,64%	8,15%	6,06%	3,33%	2,60%
20122	61,08%	24,68%	9,96%	18,43%	13,29%	11,94%	9,46%	7,38%	2,68%	2,30%
20121	11,69%	13,87%	14,03%	9,92%	15,10%	10,43%	9,62%	5,28%	3,06%	2,38%

Información desde la inscripción de ramos de otoño 2012 a otoño 2015. Datos extraídos de la Facultad Economía y Negocios de la Universidad de Chile. Elaboración propia.

Se puede apreciar que los mayores porcentajes de reprobación promedio por inscripción de cursos se encuentran en el segundo semestre. Además se refleja la drástica disminución de los porcentajes a partir del séptimo semestre. Es importante mencionar, que en todos los periodos de inscripción de cursos el semestre con mayor reprobación promedio es el segundo, excepto en los casos que el periodo de inscripción sea el semestre de primavera, donde pasa a tomar la delantera el primer semestre con la mayor reprobación promedio. Sin embargo, lo anterior se debe a que un gran porcentaje del ingreso de estudiantes a FEN se realiza el primer semestre. Mientras, que los segundos semestre existen muy pocos estudiantes de primer año, quienes a su vez reprueban un gran número de cátedras.

Por otra parte, se obtuvo el total de Ud/créditos reprobados por semestre de avance para cada inscripción de cursos. Con lo anterior, se puede observar que los primeros seis semestres llegan a ser más de un 75%<sup>2</sup> del total de créditos reprobados a lo largo de toda la carrera.

*Tabla 3. Cantidad de Ud/Créditos Reprobados.*

	1er semestre	2do semestre	3er semestre	4to semestre	5to semestre	6to semestre	7mo semestre	8vo semestre	9vo semestre	10mo semestre
20151	2540	850	1634	838	1066	646	632	559	332	104
20142	598	1642	734	1070	488	748	463	654	304	270
20141	2387	510	1198	614	1121	598	912	429	382	74
20132	668	1750	810	1128	605	1106	489	587	265	110
20131	1556	390	1027	758	1252	817	897	405	259	234
20122	608	2503	1251	2307	1297	1435	779	603	198	329
20121	2470	392	1470	1136	1823	910	831	404	245	225

Considerando Periodo de Inscripción de cursos y Semestre de avance. Datos obtenidos de la Facultad Economía y Negocios de la Universidad de Chile. Elaboración propia.

Con la intención de confirmar la decisión de los semestres seleccionados, se realizó un cálculo sobre los cursos más reprobados históricamente y de esta forma revisar si estos cursos tienen un patrón en común. Se seleccionaron los cursos que históricamente tenían más de un 15% de reprobación y que el número total de alumnos por cada curso superara los diez estudiantes.

*Tabla 4. Porcentajes de Reprobación histórica por cursos*

<sup>2</sup> Anexo 2: Detalle calculo porcentaje reprobación

<b>Nombre de Cátedra</b>	<b>% Reprobación</b>	<b>Semestre</b>
<b>CALCULO I</b>	31%	1er semestre
<b>ALGEBRA I</b>	28%	1er semestre
<b>ALGEBRA II</b>	27%	3er semestre
<b>INTRODUCCION A LA ECONOMIA</b>	22%	1er semestre
<b>INTRO. MICROECONOMIA / INTRO.TEO.PRECIOS</b>	22%	2do semestre
<b>MICROECONOMIA I / TEO. PRECIOS I</b>	22%	3er semestre
<b>CALCULO II</b>	21%	3er semestre
<b>CONTABILIDAD I</b>	22%	1er semestre
<b>INTRODUCCION A LA MACROECONOMIA</b>	20%	2do semestre
<b>ECONOMETRÍA I</b>	21%	5to semestre
<b>MICROECONOMIA II / TEORIA PRECIOS II</b>	20%	5to semestre
<b>ESTADISTICA II</b>	15%	3er semestre
<b>MACROECONOMIA I</b>	17%	4to semestre
<b>INVESTIGACION OPERATIVA</b>	17%	5to semestre
<b>CONTABILIDAD II / TOPICOS CONTABILIDAD</b>	19%	3er semestre
<b>CALCULO III</b>	22%	5to semestre
<b>ECONOMIA DEL TRABAJO</b>	17%	6to semestre

Datos extraídos de la Facultad Economía y Negocios de la Universidad de Chile.  
Elaboración propia.

Para realizar una mejor asignación de cada alumno al semestre de avance que se debiese encontrar según la cantidad de Ud/créditos acumulados en un determinado periodo de inscripción de cursos, se amplió el rango de Ud/créditos que se pueden obtener cursando en el primer semestre de las distintas carreras. Lo anterior se debe a la posible convalidación<sup>3</sup> de los tres primeros cursos de inglés que la facultad permite reconocer por medio de un examen de diagnóstico, utilizando puntaje obtenido en una prueba de diagnóstico. Para malla nueva se agregaron 6 créditos (2 créditos por cada curso de inglés) a la cota superior del primer semestre, quedando como límite superior 34 créditos. Mientras que para la malla antigua se agregaron 12 Ud a la cota superior del rango, ya que los cursos de inglés tienen 4 Ud, alcanzando 60 Ud el límite superior del primer semestre.

*Tabla 5. Cantidad de créditos según semestres para la Malla Nueva*

<b>Malla Nueva</b>		
<b>Tramos</b>	<b>Créditos</b>	<b>Semestres Inscripción Cursos</b>
(28 ; 32)		2do semestre 0
(58 ; 62)		3er semestre 1
(88 ; 92)		4to semestre 2
(118 ; 122)		5to semestre 3
(148 ; 152)		6to semestre 4
(178 ; 182)		7mo semestre 5
(208 ; 212)		8vo semestre 6
(238 ; 242)		9vo semestre 7
(268 ; 272)		10mo semestre 8

Datos extraídos de la Facultad Economía y Negocios de la Universidad de Chile.  
Elaboración propia.

<sup>3</sup> Anexo 3: Promedio de Ud/Créditos convalidados

*Tabla 6. Cantidad de Ud según semestres para la Malla Antigua*

<b>Malla Antigua</b>		
<b>Tramos Ud</b>	<b>Semestres</b>	<b>Inscripción Cursos</b>
(48 ; 52)	2do semestre	0
(98 ; 102)	3er semestre	1
(148 ; 152)	4to semestre	2
(198 ; 202)	5to semestre	3
(248 ; 252)	6to semestre	4
(298 ; 302)	7mo semestre	5
(348 ; 352)	8vo semestre	6
(398 ; 402)	9vo semestre	7
(448 ; 452)	10mo semestre	8

Datos extraídos de la Facultad Economía y Negocios de la Universidad de Chile.  
Elaboración propia.

Finalmente, la última decisión que se tomó antes de la etapa de pre-procesamiento de los datos fue seleccionar cuáles son los atributos o factores claves que describen a los distintos alumnos, para identificar los tipos de alumnos que se generan por semestre de avance al momento de inscribir los cursos.

Los atributos seleccionados para generar los grupos de alumnos con comportamientos similares, fueron elegidos considerando tres etapas. En la primera etapa, se realizó un análisis de los siete mejores trabajos realizados en el semestre de primavera 2015, en el curso “Business Intelligence and Analytics” de la misma facultad. Cada uno de los grupos de trabajo se interiorizó en el modelo del negocio y los datos. Finalmente se decidió a partir de un análisis previo, como por ejemplo el de correlación de las variables. En la segunda etapa, se consideraron los resultados encontrados

en un estudio anterior llamado “Carga Académica: Identificación de factores claves en una escuela de Economía y Negocios” [20]. Este trabajo buscaba identificar las variables claves que influyen el nivel de la carga académica que inscriben los estudiantes en FEN encontrando diferentes variables claves para una carga académica alta, media y baja. Por último, se utilizó para complementar las etapas anteriores el juicio experto del equipo de investigación a partir del conocimiento en profundidad del negocio y de los datos.

*Tabla 7: Atributos Seleccionados*

<b>Variab</b>	
<b>Secretaria de Estudios</b>	<b>DEMRE</b>
Cod_Alumno	Grupo_Dependencia
Periodo	Tiene_Trabajo_rem
%_Reprob Sem Anterior	Horas_que_dedica_trabajo
Sem_Avance	Grupo_familiar
Tramos	Ingreso_bruto_familiar
Factor_Rend Sem Anterior	¿Vive_sus_padres?
Tipo_Carrera	Ed_padre
	Ed_madre
	Sexo
	Régimen

Para clarificar en qué consisten los distintos atributos seleccionados, se detallaran inicialmente los pertenecientes a información proveniente de Secretaria de Estudios y luego los provenientes del DEMRE. Dentro del primer grupo, “Cod\_Alumno” representa un atributo que hace único al estudiante dentro de la facultad. El “Periodo” se refiere al semestre en el cual se inscriben los cursos. El atributo “%\_Reprob Sem Anterior” es el porcentaje de cursos que reprobó un determinado alumno el semestre anterior del periodo de inscripción de cursos correspondiente. “Sem\_Avance”

entrega el semestre en el que se encuentra un alumno al momento de realizar la inscripción establecida (periodo), este semestre se obtiene del total acumulado de Ud/créditos cursados y aprobados hasta dicho periodo. “Tramos” es una variable que se definió por conveniencia, generando siete tramos distintos. Los tramos aumentan de manera ascendente en diez Ud o seis créditos según la malla en la que se encuentre determinado alumno.

*Tabla 8. Tramo según cantidad de créditos y de Ud*

	<b>Créditos</b>	<b>Ud</b>
<b>Tramo 1</b>	menos de 8	menos de 13
<b>Tramo 2</b>	$8 < x \leq 14$	$13 < x \leq 23$
<b>Tramo 3</b>	$14 < x \leq 20$	$23 < x \leq 33$
<b>Tramo 4</b>	$20 < x \leq 26$	$33 < x \leq 43$
<b>Tramo 5</b>	$26 < x \leq 32$	$43 < x \leq 53$
<b>Tramo 6</b>	$32 < x \leq 38$	$53 < x \leq 63$
<b>Tramo 7</b>	$x > 38$	$x > 63$

El “Factor\_Rend Sem Anterior” es un porcentaje que representa la cantidad de cátedras cursadas y aprobadas sobre la cantidad total de cátedras cursadas, aprobadas y reprobadas.

Por último, el atributo “Tipo Carrera” indica la carrera que el alumno está estudiando, pudiendo ser Ingeniería Comercial (IC), Contador Auditor (CA) e Ingeniería en Información y Control de Gestión (IICG).

Por otro lado, los atributos seleccionados provenientes de la base de datos proporcionada por el DEMRE son diez. Inicialmente los atributos relacionados a la educación del alumno encontramos “Grupo\_Dependencia” que nos indica si el alumno proviene de un colegio Particular Pagado (1),

Particular Subvencionado (2) o de uno Municipal (3). Por su parte, el atributo “Régimen” indica si el alumno proviene de un colegio Masculino (1), Femenino (2) o Coeducacional (3).

Además, dentro de los atributos propios del alumno encontramos, “Sexo” que representa si el alumno es Hombre (1) o Mujer (2). “Tiene\_trabajo\_rem” indica si un alumno tiene trabajo o no (1), y si este es permanente (3) o temporal (2). De manera complementaria a este último atributo, “Horas\_que\_dedica\_trabajo” entrega un número de dos dígitos indicando cuantas horas el alumno le dedica al trabajo.

En cuanto a información sobre la estructura familiar del alumno, se seleccionó “Grupo\_familiar” que indica el número de personas que componen el grupo familiar. “Ingreso\_bruto\_familiar” por su parte muestra el tramo en el que se encuentra la familia, pudiendo clasificarse en doce tramos, con una diferencia de \$144.000 entre los tramos<sup>4</sup>. El atributo “¿Vive\_sus\_padres?” indica si ambos padres viven (1), sólo lo hace la madre (2), sólo lo hace el padre (3) o ambos están fallecidos (4). Por último, se incluyó la educación que tenían los padres, teniendo trece categorías<sup>5</sup>.

**Etapa 3:** En esta etapa, de pre-procesamiento de datos, se decidió eliminar los registros con datos del DEMRE no encontrados, ya que eran un número de alumnos pequeño en relación a la muestra de datos total que se tenía..

Por otro lado, al encontrar datos “Sin Información” se dejó como un valor nulo.

---

<sup>4</sup> Anexo 4: Detalle Tramos de Ingreso Bruto Familiar

<sup>5</sup> Anexo 5: Detalle Educación de los Padres

Finalmente se eliminaron 63 datos del periodo otoño 2015 quedando 2696 registros, 45 de primavera 2014 quedando 2344 registros, 52 de otoño 2014 quedando 2520 registros, 44 de primavera 2013 quedando 2160 registros, 46 de otoño 2013 quedando 2312 registros, 41 de primavera 2012 quedando 2039 registros y 45 de otoño 2012 quedando 2198 registros.

**Etapa 4:** En la transformación de datos, se modificaron atributos de los seleccionados para lograr una mejor obtención de los objetivos establecidos en un comienzo.

Para la variable “Tramo”, sólo se dejó el número del tramo al cual correspondía, por ejemplo, si un alumno pertenecía al tramo 2, este sería reemplazado por el número 2. Se realizó esto, ya que los tramos aumentaban de forma ascendente y lo hacían en la misma proporción.

Para la variable “Sem\_Avance” se realizó el mismo procedimiento anterior, reemplazando sólo por el número el semestre de avance en el que el alumno se encontraba.

Por otro lado, el atributo “Grupo\_Dependencia” se decidió cambiar la asignación entregada por el DEMRE y representar las categorías con otros números, Particular Pagado cambió de 1 a 20, el Particular Subvencionado cambió de 2 a 10 y el Municipal de 3 a 5. Con ello se intentó representar las similitudes y diferencias entre los tres tipos de colegios. En definitiva consideramos que el colegio municipal y particular subvencionado se asemeja más que a los colegios Particulares Pagados.

Dentro de las variables relacionadas con los padres, específicamente, “¿Viven\_sus\_Padres?” se le asignó a las opciones un número que pudiera representar la diferencia entre ellas, es decir, que vivan los dos, que viva el padre o la madre, y que no viva ninguno. Asignándole el número 5 a ambos viven, el número 2 a sólo padre y a sólo la madre, y el 0 a ninguno vive. En este mismo punto, para la “Educación de los padres”, se decidió separarla, dejando cada educación en una columna, y además de esto, se reemplazó el estudio que tenían por la cantidad de años que en promedio representa cada una de las opciones. Por ejemplo, si el padre tiene educación universitaria completa, se reemplazó por el número 15, lo que quiere decir que en promedio estudió 15 años para a dicha categoría. Todas las asignaciones de años a la educación se calcularon como un promedio y aproximado. No es un cálculo exacto, ya que depende de la carrera, de la universidad o del instituto, sin embargo, se acerca bastante a la duración real de cada una de las carreras.

La variable “Tiene\_trabajo\_rem” se modificó con el objetivo de que la opción Sin Información pasara a ser dato nulo. Entonces la opción No trabaja quedo como 0, Trabaja Ocasionalmente quedó como 1 y Trabaja Permanentemente quedo como 2.

Finalmente se optó por normalizar los datos de la muestra, para evitar los problemas de diferencia de rangos entre las variables seleccionadas. Cabe destacar que para interpretar de manera más intuitiva y acertada los resultados se desnormalizaron, volviendo a sus magnitudes originales.

**Etapas 5:** La tarea de DM elegida para alcanzar los objetivos planteados en este estudio es clusterización no supervisada. Es de carácter no supervisada debido a que se quiere lograr es aprender de los datos sin conocer los patrones previamente establecido, en nuestro caso específico, encontrar patrones de atributos sin conocer los perfiles resultantes de estudiantes con anterioridad. Se puede encontrar un mayor detalle de este método, denominado Análisis de Conglomerado en el capítulo 2 (página 19).

**Etapas 6:** El algoritmo elegido para encontrar los patrones buscados es el de Distancias Euclidianas. La elección del algoritmo fue exclusivamente por la compatibilidad de éste con herramienta de minería de datos que se utilizó para analizar la muestra y encontrar los clústeres, SPSS Modeler.

**Etapas 7:** Para obtener patrones más representativos antes de incorporar los datos en el programa SPSS Modeler, se dividió la muestra por semestre de avance (semestre dos al seis) y por escuela. Es decir, para cada semestre de avance realizó el mismo proceso diferenciando la muestra para los estudiantes de la Escuela de Economía y Administración, la que incluye la carrera de Ingeniería Comercial, y por otro lado, la Escuela de Sistemas de Información y Auditoría, la que incluye por su parte la carrera de Contador Auditor e Ingeniería de Información y Control de Gestión.

Se obtuvo distintos clústeres por semestre de avance y escuela. Se obtuvo, la media, la desviación estándar, el error estándar y la cantidad de estudiantes pertenecientes a cada cluster obtenido. También se calculó el test de media.

**Etapa 8 y Etapa 9:** La interpretación de patrones y la consolidación de conocimiento obtenido se desarrollará en el capítulo 4: Resultados del Estudio.

### 3.3. Pregunta de Investigación

Continuando con el objetivo del estudio, la pregunta que se busca responder con la investigación es: ¿Cuáles son los distintos grupos de estudiantes que existen en FEN diferenciados por semestre de avance al momento de inscribir sus cursos?

Sumado a lo anterior y de manera complementaria, también se busca responder lo siguiente: ¿Cuáles son las características más relevantes que identifican y diferencian a los grupos de alumnos encontrados?

## 4. Resultados del Estudio

La Facultad de Economía y Negocios de la Universidad de Chile está constituida por dos escuelas, la Escuela de Economía y Administración que imparte la carrera de Ingeniería Comercial y la Escuela de Sistemas de Información y Auditoría conformada por la carrera de Contador Auditor e Ingeniería de Información y Control de Gestión. Se decidió separar estas escuelas con el objetivo de encontrar patrones más representativos y personalizados para la facultad.

Siempre teniendo como meta lograr los objetivos planteados en este estudio, se tomó la muestra de datos y se dividió por semestre de avance y por escuela según la carrera.

A continuación se detallaran por semestre de avance, los distintos clústeres que se formaron para las dos escuelas que conforman FEN en la actualidad. El detalle de los resultados obtenidos se encuentra resumido en el anexo 5, test de medias realizado para cada semestre de avance (del 2 al 6) según las dos escuelas existentes.

## **Semestre de Avance 2**

- **Escuela de Economía y Administración**

**Clúster 1 (202):** Poseen el mayor factor de rendimiento alcanzando un 91,7%, además poseen el menor porcentaje de reprobación del semestre anterior con un 9,3%. Este grupo está conformado sólo por mujeres, las que provienen generalmente de colegios coeducacionales y particulares pagados. Poseen el segundo ingreso bruto familiar más alto, encontrándose en promedio entre el tramo 10 y 11. La educación de ambos padres es en promedio alta, superando los 15 años de educación.

**Clúster 2 (90) y Clúster 3 (328):** Ambos clúster son bastante similar al clúster 1, sin embargo poseen un factor de rendimiento un poco menor (89% y 87,2% respectivamente) y un porcentaje de reprobación del semestre anterior un poco más alto (13% y 14% respectivamente). Sin embargo, la principal diferencia es que el clúster 2 se compone en su totalidad de hombres los cuales provienen de colegios masculinos, mientras que el clúster 3 está compuesto 100% de hombres quienes provienen exclusivamente de colegios coeducacionales.

**Clúster 4 (257):** Este grupo de alumnos es el que posee menor factor de rendimiento (87,3%) similar al del clúster 2. A la vez poseen la mayor tasa de reprobación del semestre anterior (15% aproximadamente). Su principal diferencia con los otros clústeres es que generalmente estos alumnos provienen de colegios municipales. Por otro lado, son el grupo que en promedio tienen el menor ingreso bruto ubicándose en el tramo 3. Además sus padres poseen los menores años de educación en promedio, alcanzando los 12 años. Este grupo está compuesto por hombres y mujeres, sobresaliendo la presencia masculina.

Existen variables que poseen promedios muy similares para los cuatro clúster que se formaron, las más homogéneas son: Tramo que se mueve entre 4,6 y 4,7, Grupo familiar que varía entre 4 y 5, la variable de Viven sus padres la cual se encuentra entre 4 y 5.

*Tabla 9. Variables diferenciadoras más relevantes, EEA-2SEM*

	Clúster 1		Clúster 2		Clúster 3		Clúster 4	
	Media	Desv. Stand						
Factor_Rend sem anterior	0.917	0.113	0.873	0.156	0.890	0.127	0.872	0.134
%_REPROB SEM ANTERIOR	0.093	0.151	0.144	0.202	0.131	0.172	0.146	0.179
GRUPO_DEPENDENCIA	18.639	3.492	17.333	4.983	19.223	2.807	6.887	2.428
GRUPO_FAMILIAR	4.851	2.531	4.711	2.781	4.598	2.641	4.054	1.568
INGRESO_BRUTO_FAM	10.510	2.660	9.889	3.272	10.793	2.479	3.261	1.934
EDU_PADRE	16.179	1.707	16.260	1.491	16.312	1.591	12.234	3.239
EDU_MADRE	15.860	1.840	15.493	2.075	15.770	1.832	11.963	3.167
SEXO	2.000	0.000	1.000	0.000	1.000	0.000	1.358	0.480
REGIMEN	2.827	0.379	1.000	0.000	3.000	0.000	2.549	0.739

- **Escuela de Sistemas de Información y Auditoría**

**Clúster 1 (144):** Está conformado sólo por mujeres, las que generalmente provienen colegio coeducacionales. Con un factor de rendimiento del 84,4% y una tasa de reprobación entorno al 20%.

**Clúster 2 (51):** Grupo que en general proviene de colegio municipal. Pertenece al tramo 2 de ingreso bruto familiar. Además sus padres son los que tienen en promedio la menor cantidad de años en la educación (10 años). Está compuesto tanto por hombre como por mujeres, los que generalmente provienen colegios coeducacionales. Además, muestra un factor de avance y tasa de reprobación del semestre anterior levemente menor al cluster 1 sin ser muy distinto a los otros grupos.

**Clúster 3 (162):** Grupo que posee el mayor factor de rendimiento (87%). Generalmente provienen de colegios particulares subvencionados o particulares pagados. Sumado a lo anterior, son el grupo que posee en promedio el mayor ingreso bruto familiar (Tramo 7). Está compuesto únicamente por hombres, los que provienen en su totalidad de colegios coeducacionales.

**Clúster 4 (64):** Es el segundo grupo con menor factor de rendimiento (83,7%) y el que posee la mayor tasa reprobación del semestre anterior (20%). Generalmente provienen de colegios particulares subvencionados. Son los únicos que no trabajan. Está compuesto sólo hombres, los que vienen en un 100% de colegios masculinos.

Todos los clústeres en general poseen un factor de rendimiento entre el 82% y el 87% y una tasa de reprobación del semestre anterior sobre el 15%, estando todos bastante cercanos en los resultados académicos.

*Tabla 10. Variables diferenciadoras más relevantes, ESIA-2SEM*

	Clúster 1		Clúster 2		Clúster 3		Clúster 4	
	Media	Desv. Stand						
Factor_Rend sem anterior	0.844	0.154	0.821	0.142	0.870	0.156	0.837	0.163
%_REPROB SEM ANTERIOR	0.194	0.220	0.184	0.209	0.161	0.215	0.204	0.239
GRUPO_DEPENDENCIA	12.465	6.173	7.255	2.513	14.012	5.706	10.625	5.669
INGRESO_BRUTO_FAM	6.243	3.820	2.647	1.180	7.340	3.916	6.719	3.574
EDU_PADRE	14.246	3.032	10.652	2.877	14.624	2.476	14.054	2.706
EDU_MADRE	13.786	2.930	10.936	3.010	14.280	2.636	14.263	2.539
SEXO	2.000	0.000	1.549	0.503	1.000	0.000	1.000	0.000
REGIMEN	2.701	0.489	2.804	0.401	3.000	0.000	1.000	0.000

### **Semestre de Avance 3**

- **Escuela de Economía y Administración**

**Clúster 1 (437):** Grupo proveniente en su mayoría de colegios particulares pagados. Además poseen el mayor ingreso bruto acercándose al tramo 11. Este grupo está compuesto por hombres, los que provienen en su totalidad de colegios coeducacionales. Con un factor de rendimiento de 92% y una tasa de reprobación del semestre anterior del 7%

**Clúster 2 (338):** Grupo de rendimiento académico más alto (95%). Además, el porcentaje de reprobación del semestre anterior es el más bajo (4,9%). En general provienen de colegio particulares pagados y no trabajan. Este grupo está compuesto por solo por mujeres, las que generalmente provienen de colegios coeducacionales.

**Clúster 3 (327):** Proviene generalmente de colegio municipales. Son los que poseen menores ingresos brutos (tramo 3). Por otro lado, sus padres son los que cuentan con la menor educación. Lo compone alumnos de ambos géneros, los que generalmente provienen de colegios coeducacionales.

**Clúster 4 (130):** Grupo de alumnos con mayor porcentaje de reprobación promedio (9,5%). No trabajan. Son en su totalidad hombres que provienen en un 100% de colegios masculinos.

En general todos los clúster poseen un Factor rendimiento sobre 90% y a la vez una tasa de reprobación menor al 10%. Por otro lado, en promedio todos los grupos se acercan al tramo 5, lo que refleja que toman una carga académica similar y poseen resultados similares.

Las variables diferenciadoras de este semestre, se encuentra la educación de los padres, la que es muy similar entre los clúster 1, 2 y 4, teniendo estos en promedio un alto nivel de educación (16 años aproximadamente), pero los padres de los alumnos del clúster 3 poseen una educación en promedio cuatro años menor. Además, existe una diferencia similar en el ingreso bruto familiar, donde el clúster 3 se aleja considerablemente de los otros tres grupos ya que se encuentra casi seis tramos bajo los otros tres cluster. Finalmente, otra diferencia del clúster 3 es que generalmente provienen de colegios municipales, en cambio los otros cluster tienden a acercarse más a los particulares pagados.



Tabla 11. Variables diferenciadoras más relevantes, EEA-3SEM

	Clúster 1		Clúster 2		Clúster 3		Clúster 4	
	Media	Desv. Stand						
Factor_Rend sem anterior	0.923	0.116	0.952	0.097	0.908	0.134	0.913	0.132
%_REPROB SEM ANTERIOR	0.077	0.148	0.049	0.116	0.086	0.166	0.095	0.179
TRAMOS	4.741	0.670	4.669	0.687	4.752	0.711	4.731	0.680
GRUPO_DEPENDENCIA	19.325	2.613	19.009	3.030	7.416	2.502	16.885	5.401
INGRESO_BRUTO_FAM	10.801	2.412	10.666	2.445	3.450	2.005	9.908	3.097
EDU_PADRE	16.371	1.570	16.264	1.835	12.473	3.113	16.009	2.040
EDU_MADRE	15.980	1.734	15.934	1.795	12.620	2.823	15.355	2.236
SEXO	1.000	0.000	2.000	0.000	1.385	0.487	1.000	0.000
REGIMEN	3.000	0.000	2.805	0.397	2.514	0.747	1.000	0.000

- **Escuela de Sistemas de Información y Auditoría**

**Clúster 1 (177):** Grupo que posee un porcentaje de reprobación del 13%. Este clúster está compuesto únicamente por hombres, los que provienen de colegio coeducacionales y mayoritariamente de colegios subvencionados.

**Clúster 2 (53):** Grupo que posee el menor porcentaje de reprobación (8%). Además son los que poseen menores ingresos brutos familiares (tramo 2). Sus padres son los que en promedio poseen menores años de educación (11 años). Los integrantes son hombres y mujeres, provienen generalmente de un colegio municipal y coeducacional.

**Clúster 3 (74):** Grupo que posee el mejor factor rendimiento (89,9%). Generalmente los integrantes de este grupo provienen de colegios particulares pagados. Está compuesto en su totalidad por mujeres, las que en su mayoría proviene de colegios coeducacionales.

**Clúster 4 (86):** Grupo compuesto únicamente por hombres, los que generalmente proviene de colegio particular subvencionado y masculino.

**Clúster 5 (98):** Grupo conformado por mujeres, las que en general provienen de colegios municipales o particulares subvencionados.

Dentro de las similitudes existente entre los clústeres, todos poseen un factor de rendimiento entorno al 88%, más aún, todos varían entre el 87% y el 89%, existiendo una mínima diferencia. A su vez, el porcentaje de reprobación también es bastante similar entre los grupos, moviéndose entre el 9% y el 13% aproximadamente. Sumado a lo anterior, los cinco clúster piden una carga similar, encontrándose entre el tramo 4 y 5 principalmente.

Las principales diferencias entre los clústeres obtenidos están en las variables de grupo de dependencia e ingreso bruto familiar, siendo variables socioeconómicas anteriores al ingreso a la universidad.

*Tabla 12. Variables diferenciadoras más relevantes, ESIA-3SEM*

	Clúster 1		Clúster 2		Clúster 3		Clúster 4		Clúster 5	
	Media	Desv. Stand								
Factor_Rend sem anterior	0.881	0.144	0.880	0.117	0.899	0.121	0.873	0.157	0.892	0.131
%_REPROB SEM ANTERIOR	0.128	0.187	0.087	0.143	0.113	0.200	0.116	0.195	0.117	0.177
GRUPO_DEPENDENCIA	13.475	5.710	8.019	2.469	17.162	4.539	10.407	6.194	7.500	2.513
INGRESO_BRUTO_FAM	6.825	3.796	2.849	1.183	9.095	3.286	6.163	3.626	3.714	1.856
EDU_PADRE	14.688	2.668	11.360	2.776	15.152	2.200	13.560	2.703	12.789	3.760
EDU_MADRE	14.194	2.571	11.735	2.871	14.641	2.065	14.080	2.437	12.642	2.675
SEXO	1.000	0.000	1.623	0.489	2.000	0.000	1.000	0.000	2.000	0.000
REGIMEN	3.000	0.000	2.811	0.395	2.905	0.295	1.012	0.108	2.510	0.502

## **Semestre de Avance 4**

- **Escuela de Economía y Administración**

**Clúster 1 (870):** Grupo con mayor factor de rendimiento (93,5%), y además poseen la menor tasa de reprobación del semestre anterior (6,6%). Se diferencia de los otros clúster debido poseen un mayor ingreso bruto familiar (tramo 10 aproximadamente). Por otro lado, este grupo de alumnos tiene los padres con mayores años de educación, obteniendo en promedio 16 años. La composición es mixta, y en general los alumnos que lo integran provienen de colegios particulares pagados y coeducacionales.

**Clúster 2 (194):** Grupo con resultados académicos intermedios entre el clúster 1 y el clúster 3. Posee un ingreso bruto familiar relativamente bajo, específicamente se encuentran en promedio en el tramo 4. Los padres de estos estudiantes poseen una educación en promedio de 13 años, al igual que el clúster 3. Este grupo está compuesto únicamente por mujeres, que por lo general provienen de colegios municipales o particulares subvencionados.

**Clúster 3 (271):** Grupo con los resultados académicos más bajos, ya que poseen el menor factor de rendimiento (88,9%) y la tasa de reprobación del semestre anterior mayor (11%). Al igual que el clúster 2, su ingreso bruto tiende a ser relativamente bajo, ubicándose en promedio en el tramo 4. Este clúster está compuesto únicamente por hombres, los que por lo general provienen de colegios municipales o particulares subvencionados.

Dentro de las variables similares destaca la carga académica que toman en promedio (Tramo), obteniendo los tres clústeres un promedio de 4,7.

Es importante mencionar que los clúster 2 y 3 son bastantes similares en distintas variables, como por ejemplo el ingreso bruto familiar, el grupo de dependencia y los años de educación de los padres, donde la principal diferencia que existe entre ellos es el género que compone a los grupos, estando el clúster 2 compuesto puramente de mujeres y el clúster 3 solamente de hombres.

*Tabla 13. Variables diferenciadoras más relevantes, EEA-4SEM*

	Clúster 1		Clúster 2		Clúster 3	
	Media	Dev. Stand	Media	Dev. Stand	Media	Dev. Stand
Factor_Rend sem anterior	0.935	0.094	0.907	0.118	0.889	0.138
%_REPROB SEM ANTERIOR	0.066	0.130	0.098	0.156	0.111	0.175
GRUPO_DEPENDENCIA	19.828	1.302	9.562	4.955	8.303	3.810
INGRESO_BRUTO_FAM	10.909	2.242	4.418	2.851	4.609	2.777
EDU_PADRE	16.472	1.300	13.459	3.216	13.393	2.884
EDU_MADRE	15.957	1.704	13.491	2.529	12.972	2.988
SEXO	1.370	0.483	2.000	0.000	1.000	0.000
REGIMEN	2.706	0.658	2.608	0.489	2.173	0.987

- **Escuela de Sistemas de Información y Auditoría**

**Clúster 1 (131):** Este grupo está compuesto únicamente por hombres los que en promedio vienen de colegios municipales o particulares subvencionados. Además poseen un tramo de ingreso familiar bajo, encontrándose en promedio en el tramo 4. La educación de sus padres tiende a no ser universitaria, ya que su promedio no supera en promedio los 13 años.

**Clúster 2 (78):** Es el grupo que posee el factor de rendimiento mayor (90,2%). Está compuesto únicamente por mujeres, las que provienen generalmente de un colegio particular pagado y coeducacional. Es el grupo con el segundo ingreso bruto familiar más alto, acercándose al tramo 9. La educación de los padres es generalmente es universitaria completa, ya que su promedio alcanza los 15 años de estudio.

**Clúster 3 (132):** Este es el grupo que posee la menor tasa de reprobación del semestre anterior (9,6%). Lo integran únicamente mujeres, las que en general proviene de un colegio municipal. La educación de los padres, al igual que en el clúster 1 no alcanza a ser universitaria, ya que su promedio llega a los 12 años.

**Clúster 4 (76):** Grupo que posee el menor factor de rendimiento (85,3%) y a la vez posee la mayor tasa de reprobación del semestre anterior (16,2%). Los integrantes son solo hombres, los que provienen generalmente de colegios particulares pagados y coeducacionales. Por otro lado son los que poseen mayores ingresos brutos familiares, encontrándose en promedio en el tramo 9. Los padres de estos alumnos poseen educación completa, ya que su promedio es de 15 años.

Podemos ver que existen semejanzas entre los clústeres 1 y 3, y el 2 y 4. Estos clústeres comparten principalmente las variables socio-demográficas, como lo es la educación de los padres y el grupo de dependencia. Podemos ver que la principal diferencia entre estos grupos es la composición de género, donde el clúster 1 está compuesto únicamente por hombres versus

el clúster 3 que lo hace únicamente por mujeres. Lo mismo ocurre con los otros dos clústeres.

*Tabla 14. Variables diferenciadoras más relevantes, ESIA-4SEM*

	Clúster 1		Clúster 2		Clúster 3		Clúster 4	
	Media	Desv. Stand						
Factor_Rend sem anterior	0.878	0.137	0.902	0.118	0.887	0.107	0.853	0.143
%_REPROB SEM ANTERIOR	0.127	0.175	0.114	0.180	0.096	0.144	0.162	0.213
GRUPO_DEPENDENCIA	7.901	2.477	16.282	4.864	7.121	2.481	18.947	3.089
INGRESO_BRUTO_FAM	4.176	2.332	8.987	3.379	3.159	1.693	9.132	3.226
EDU_PADRE	12.917	2.767	15.953	1.578	12.387	3.250	15.485	2.335
EDU_MADRE	12.816	3.214	15.078	2.073	12.148	3.252	15.045	2.123
SEXO	1.000	0.000	2.000	0.000	2.000	0.000	1.000	0.000
REGIMEN	2.374	0.931	2.872	0.336	2.485	0.502	2.368	0.936

## **Semestre de Avance 5**

- **Escuela de Economía y Administración**

**Clúster 1 (292):** Grupo de estudiantes que mejor desempeño académico tienen en promedio, ya que son los que poseen el factor de rendimiento mayor (94,2%) y la tasa de reprobación del semestre anterior menor (7,3%). Este grupo está conformado únicamente por mujeres, las que en general provienen de colegios particulares pagados y coeducacionales. Por otro lado, el ingreso bruto familiar promedio que poseen es relativamente alto, encontrándose aproximadamente en el tramo 10. Este grupo posee los padres con mayores años de educación promedio, siendo 16 años.

**Clúster 2 (292):** Grupo con un factor de rendimiento promedio y una tasa de reprobación intermedia. Este grupo está compuesto solo por hombres quienes en su gran mayoría provienen de colegios particulares pagados y en su 100% de colegios coeducacionales. El ingreso bruto familiar promedio se

encuentra cercano al tramo 11. Los padres de este clúster, tienen un comportamiento similar al del clúster anterior, ya que poseen en promedio 16 años de estudios.

**Clúster 3 (95):** Grupo con un factor de rendimiento promedio y una tasa de reprobación intermedia. Este grupo está compuesto en su mayoría por alumnos del género masculino, los que provienen tanto de los colegios particulares pagados como de los subvencionados, aunque es importante mencionar que prevalecen los primeros. Además el colegio tiende a ser de tipo masculino. Existe una gran similitud en los promedios de las variables de ingreso y de educación de padres con el clúster anterior.

**Clúster 4 (300):** Grupo de alumnos el rendimiento académico más bajo de este semestre (90,3%) y la mayor tasa de reprobación del semestre anterior (10,4%). Este clúster lo componen tanto alumnos hombres como mujeres, prevaleciendo los primeros. Destaca la gran diferencia que existe en el ingreso bruto familiar que posee en comparación con los otros clústeres, estando ubicado aproximadamente siete tramos más abajo.

Se puede apreciar claramente que en general todos poseen un buen factor de rendimiento, sobre el 90% en promedio y por otro lado una baja tasa de reprobación, ya que el promedio más alto de esta variable alcanza un 10%. Destaca el hecho de que ningún grupo le dedica tiempo al trabajo.

Por el lado de las diferencias, se encontró que los primeros tres clústeres tienen un ingreso bruto promedio ubicado en el tramo 10, versus el clúster 4

que se encuentra en el tramo 3. Ocurre algo similar con la variable educación de los padres, ya que en los primeros tres clústeres los promedios están muy cercanos, pero en el último clúster hay un descenso en la cantidad de años de educación que poseen los padres de ese grupo.

*Tabla 15. Variables diferenciadoras más relevantes, EEA-5SEM*

	Clúster 1		Clúster 2		Clúster 3		Clúster 4	
	Media	Desv. Stand						
Factor_Rend sem anterior	0.942	0.081	0.927	0.098	0.931	0.082	0.903	0.120
%_REPROB SEM ANTERIOR	0.073	0.147	0.075	0.136	0.098	0.169	0.104	0.165
GRUPO_DEPENDENCIA	18.699	3.446	19.452	2.280	16.000	6.341	8.750	4.444
INGRESO_BRUTO_FAM	10.171	2.738	10.664	2.489	10.558	2.500	3.733	2.178
EDU_PADRE	16.473	1.211	16.524	1.144	16.189	1.606	12.563	3.194
EDU_MADRE	16.062	1.653	15.993	1.675	15.926	1.770	12.553	2.505
SEXO	2.000	0.000	1.000	0.000	1.021	0.144	1.350	0.478
REGIMEN	2.795	0.405	3.000	0.000	1.011	0.103	2.463	0.803

- **Escuela de Sistemas de Información y Auditoría**

**Clúster 1 (155):** Es el grupo de alumnos que en promedio poseen el mejor factor de rendimiento (91,3%) y a la vez poseen la menor tasa de reprobación del semestre anterior (8,6%). Formado únicamente por mujeres, las que provienen generalmente de un colegio particular subvencionado. Su ingreso bruto familiar es medio-bajo, ya que se encuentra en promedio en el tramo 5. Por último, la educación de los padres alcanza los 13 años promedio, por lo que refleja que no poseen educación universitaria completa.

**Clúster 2 (24):** Poseen un rendimiento promedio en relación a los demás grupos. Está compuesto por hombres y mujeres, los que tienden a provenir de colegios municipales y coeducacionales. Son los que en promedio tienen el ingreso bruto familiar menor, acercándose al tramo 3. Por último, los

padres de estos alumnos poseen el menor promedio en años de educación, alcanzando los 10 años.

**Clúster 3 (6):** Poseen un rendimiento promedio en relación a los demás grupos. Son en su plenitud hombres, los que provienen únicamente de colegios subvencionados y generalmente estos son de carácter masculinos. Destaca entre estos alumnos el hecho de que tienden a trabajar permanentemente y a la vez bastantes horas. Son los segundos con ingresos familiares menores, encontrándose en promedio en el tramo 3. Sumado a lo anterior, son los segundos que poseen los padres con los años de educación menor, comparado con el resto de los grupos. Sin embargo, es un grupo en el cual existen muy pocos estudiantes.

**Clúster 4 (115):** Poseen un rendimiento promedio en relación a los demás grupos. Este grupo está compuesto únicamente por hombres, los que generalmente provienen de colegios particulares subvencionados y está compuesto solamente por hombres. Es el grupo que posee mayores ingresos familiares, tomando el tramo 6. Lo anterior no indica que se encuentren en un tramo alto, ya que los tramos de ingresos van del 1 al 12. Son a la vez, los que en promedio poseen los padres con mayor educación. Sin embargo, se debe mencionar que en promedio no alcanzan a tener educación universitaria completa.

**Clúster 5 (53):** Son los que poseen los resultados académicos menores, ya que tienen el factor de rendimiento más bajo (84,9%) y la tasa de reprobación mayor (14,3%). Este grupo está compuesto en su totalidad por

hombres, los que generalmente provienen de colegios particulares subvencionados. La educación de los padres es similar clúster 4.

Las principales diferencias siguen viniendo de las variables socio-demográficas, específicamente, de las variables grupo dependencia, ingreso bruto y educación de los padres.

Dentro de las variables que son bastante similares y no generan conclusiones significativas son el tramo que toman, el grupo familiar y si viven sus padres.

*Tabla 16. Variables diferenciadoras más relevantes, ESIA-5SEM*

	Clúster 1		Clúster 2		Clúster 3		Clúster 4		Clúster 5	
	Media	Desv. Stand								
Factor_Rend sem anterior	0.913	0.090	0.888	0.139	0.888	0.112	0.883	0.124	0.849	0.145
%_REPROB SEM ANTERIOR	0.086	0.133	0.092	0.172	0.136	0.158	0.146	0.188	0.143	0.204
GRUPO_DEPENDENCIA	10.677	5.924	7.292	2.545	10.000	0.000	13.130	6.014	11.415	6.749
TIENE_TRABAJO_REM	0.000	0.000	0.000	0.000	1500	0.548	0.019	0.136	0.000	0.000
HORAS_QUE_DEDICA_TRABAJO	0.000	0.000	0.000	0.000	49167	24.580	0.087	0.656	0.000	0.000
INGRESO_BRUTO_FAM	5.129	3.710	2.792	1.351	3.167	0.983	6.478	3.648	5.943	3.342
EDU_PADRE	13.435	3.219	10.417	3.175	13.667	2.582	14.874	2.581	14.167	2.282
EDU_MADRE	13.448	2.775	10.958	1.367	11.667	0.816	14.131	2.978	14.000	2.102
SEXO	2.000	0.000	1.417	0.504	1.000	0.000	1.000	0.000	1.000	0.000
REGIMEN	2.594	0.493	2.667	0.637	1.333	0.816	3.000	0.000	1.000	0.000

## **Semestre de Avance 6**

- **Escuela de Economía y Administración**

**Clúster 1 (172):** Su desempeño académico es promedio. Este grupo está compuesto sólo por mujeres las que tienden a provenir de colegios particulares subvencionados o municipales. Son uno de los grupos que poseen los menores ingresos familiares promedio (tramo 3), junto al clúster 5. La cantidad de años promedio de educación de los padres no alcanza la universitaria completa, teniendo como resultado un promedio de 12 años.

**Clúster 2 (395):** Desempeño académico promedio. Este grupo está compuesto únicamente por hombres, quienes mayoritariamente provienen de colegios particulares pagados y en un 100% son colegios con régimen masculinos. Son uno de los grupos que posee el ingreso familiar mayor, ubicándose en el tramo 10. Sumado a lo anterior, poseen los padres con la mayor educación junto al clúster 3, superando en promedio los 15 años.

**Clúster 3 (298):** Grupo que posee el mejor desempeño académico, ya que tienen el mayor factor de rendimiento (94%) y además poseen la menor tasa de reprobación del semestre anterior (6,9%). Este grupo está compuesto solo por mujeres, las que generalmente provienen de colegios particulares pagados y de tipo coeducacional. Poseen los más altos ingresos brutos (Tramo 10) y a la vez tiene a los padres con la mayor cantidad de años educativos, alcanzando los 16 años promedio.

**Clúster 4 (195):** Poseen un desempeño académico promedio. Este grupo está compuesto sólo por hombres, los cuales provienen generalmente de colegios particulares subvencionados con régimen masculinos. El ingreso familiar está en el promedio y la educación de los padres tiende a ser media-alta, ya que alcanza los 14 años.

**Clúster 5 (123):** Grupo con el menor desempeño académico, a pesar de que sigue siendo relativamente bueno. Su factor de rendimiento es de un 89,7% y la tasa de reprobación alcanza un 11,3%. Este grupo está compuesto solamente por hombres, los cuales provienen generalmente de colegios subvencionados o municipales y a la vez tienden a ser coeducacionales.

En el sexto semestre podemos apreciar clúster muy similares tanto en el factor rendimiento como en la tasa de reprobación. La variable Tramo arroja promedios muy similares para todos los clústeres, variando de 4,7 a 4,8.

Los clústeres siguen diferenciándose principalmente por las variables socio-demográficas, como lo son el ingreso, el grupo de dependencia y la educación de los padres.

Es importante mencionar la similitud entre clústeres, los cuales tienden a diferenciarse únicamente por el género y el grupo de dependencia.

*Tabla 17. Variables diferenciadoras más relevantes, EEA-6SEM*

	Clúster 1		Clúster 2		Clúster 3		Clúster 4		Clúster 5	
	Media	Desv. Stand								
Factor_Rend sem anterior	0.901	0.098	0.922	0.085	0.941	0.082	0.920	0.083	0.897	0.099
%_REPROB SEM ANTERIOR	0.111	0.165	0.078	0.134	0.069	0.125	0.075	0.127	0.113	0.173
GRUPO_DEPENDENCIA	8.750	4.228	19.519	2.143	19.664	1.804	13.359	7.007	8.780	3.093
INGRESO_BRUTO_FAM	3.971	2.739	10.322	2.748	10.379	2.679	7.744	4.070	3.675	2.125
EDU_PADRE	13.812	3.063	16.349	1.383	16.408	1.915	14.743	2.874	12.619	3.455
EDU_MADRE	13.430	2.656	15.773	1.850	16.071	1.636	14.168	3.055	12.595	2.984
SEXO	2.000	0.000	1.000	0.000	2.000	0.000	1.000	0.000	1.000	0.000
REGIMEN	2.587	0.494	3.000	0.000	2.826	0.389	1.000	0.000	2.992	0.090

- **Escuela de Sistemas de Información y Auditoría**

Formación de tres clústeres

**Clúster 1 (103):** Poseen la menor tasa de reprobación (11%) y un factor de rendimiento de 90%. Son únicamente hombres, los que generalmente provienen de colegios particulares subvencionados pero de tipo coeducacional. Son los que poseen mayor cantidad de integrantes en su familia y a la vez el mayor ingreso bruto familiar.

**Clúster 2 (169):** Poseen el mayor factor de rendimiento, cercano al clúster anterior (90,7%). Está formado sólo por mujeres, las que en su mayoría provienen de colegios particulares subvencionados. Son el grupo que posee menor ingreso bruto familiar, acercándose al tramo 5.

**Clúster 3 (48):** Grupo con el menor rendimiento académico, ya que son los que poseen un factor de rendimiento menor (86,9%) y una tasa de reprobación mayor (16,1%). Este grupo está compuesto solo por hombres, los cuales tienden a provenir de colegios particulares subvencionados con régimen masculinos.

Las variables Tramo y educación de padres (13 años aproximadamente) son similares en los tres clústeres que se formaron.

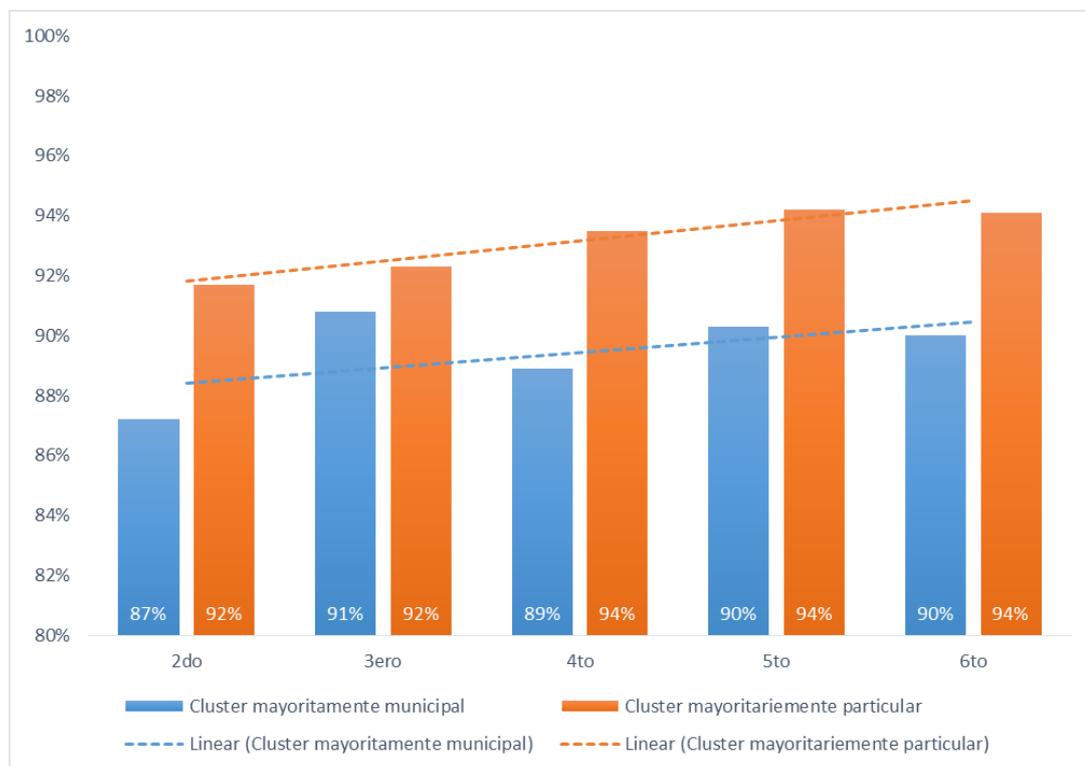
Un resultado a destacar es que lo que diferencia al clúster 1 del clúster 3 es el factor de rendimiento del semestre anterior. Además es importante mencionar que las diferencias socio-demográficas que se tendían a repetir en los clústeres anteriores, como lo eran las variables grupo de dependencia y educación de los padres.

*Tabla 18. Variables diferenciadoras más relevantes, ESIA-6SEM*

	Clúster 1		Clúster 2		Clúster 3	
	Media	Desv. Stand	Media	Desv. Stand	Media	Desv. Stand
Factor_Rend sem anterior	0.900	0.105	0.907	0.095	0.869	0.105
%_REPROB SEM ANTERIOR	0.111	0.156	0.128	0.182	0.161	0.178
GRUPO_DEPENDENCIA	12.864	6.007	10.562	5.861	10.521	6.461
GRUPO_FAMILIAR	4.583	1.332	3.834	1.782	3.688	1.764
INGRESO_BRUTO_FAM	6.155	3.816	4.882	3.592	5.292	3.087
EDU_PADRE	13.879	3.008	13.182	3.803	13.625	3.295
EDU_MADRE	13.768	3.047	13.336	2.955	13.714	2.472
SEXO	1.000	0.000	2.000	0.000	1.000	0.000
REGIMEN	3.000	0.000	2.586	0.494	1.000	0.000

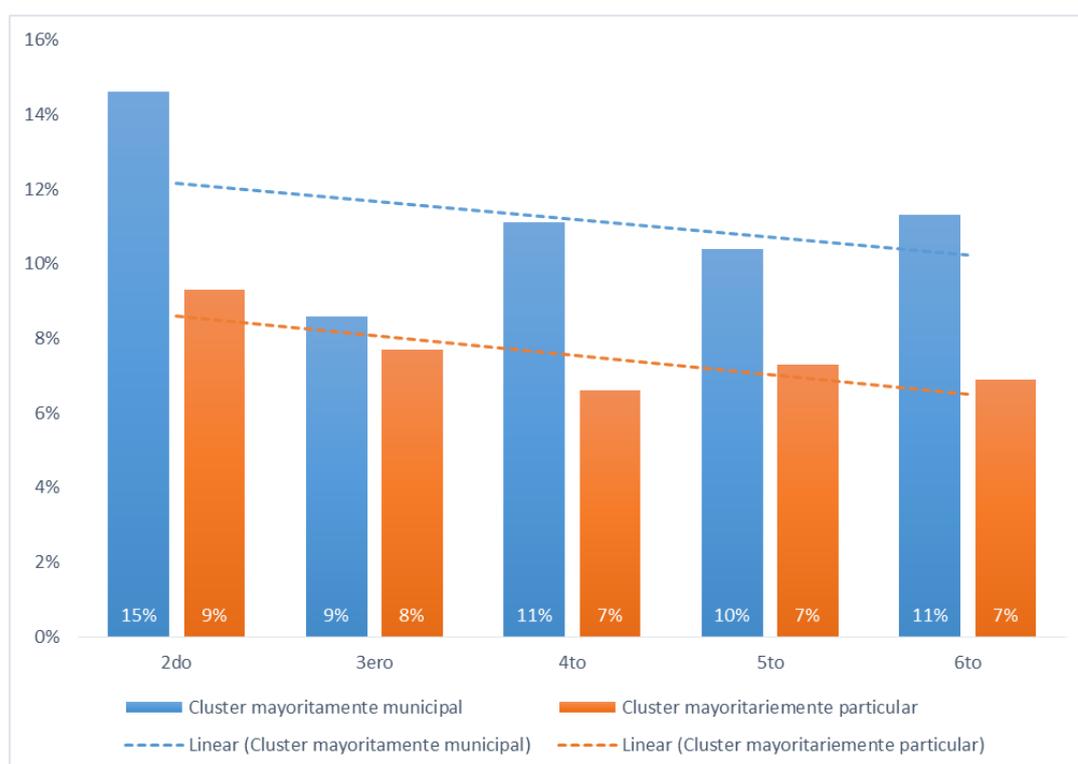
Revisando los resultados de manera transversal a través de los semestres y considerando que las reprobaciones son situaciones no deseadas para los estudiantes, se observa que el porcentaje de reprobación, en la Escuela de Economía y Administración, en los cinco semestres de los cluster de colegios mayoritariamente provenientes de colegios municipal y mayoritariamente provenientes privado van a la baja a medida que avanzan los semestres. Además, se puede observar que la mayor disminución en la tasa de reprobación se encuentra en el grupo compuesto mayoritariamente de estudiantes provenientes de colegios municipales, el cual es un 15% en el segundo semestre y que en los últimos 4 semestres se estabiliza entorno al 10%.

**Ilustración 8.** Evolución del porcentaje de reprobaciones según grupo de dependencia a través de los cinco semestres - EEA



Se puede observar que el factor de rendimiento, en la Escuela de Economía y Administración, mejora para ambos clúster a medida que avanza el tiempo de estadía en la Facultad. Para el clúster compuesto mayoritariamente de estudiantes provenientes de colegio particular este valor se estabiliza entorno al 94% en los últimos tres semestres. Mientras que para el clúster compuesto mayoritariamente de estudiantes provenientes de colegio municipal este valor se estabiliza entorno al 90%.

*Ilustración 9. Evolución del Factor de Avance según grupo de dependencia*

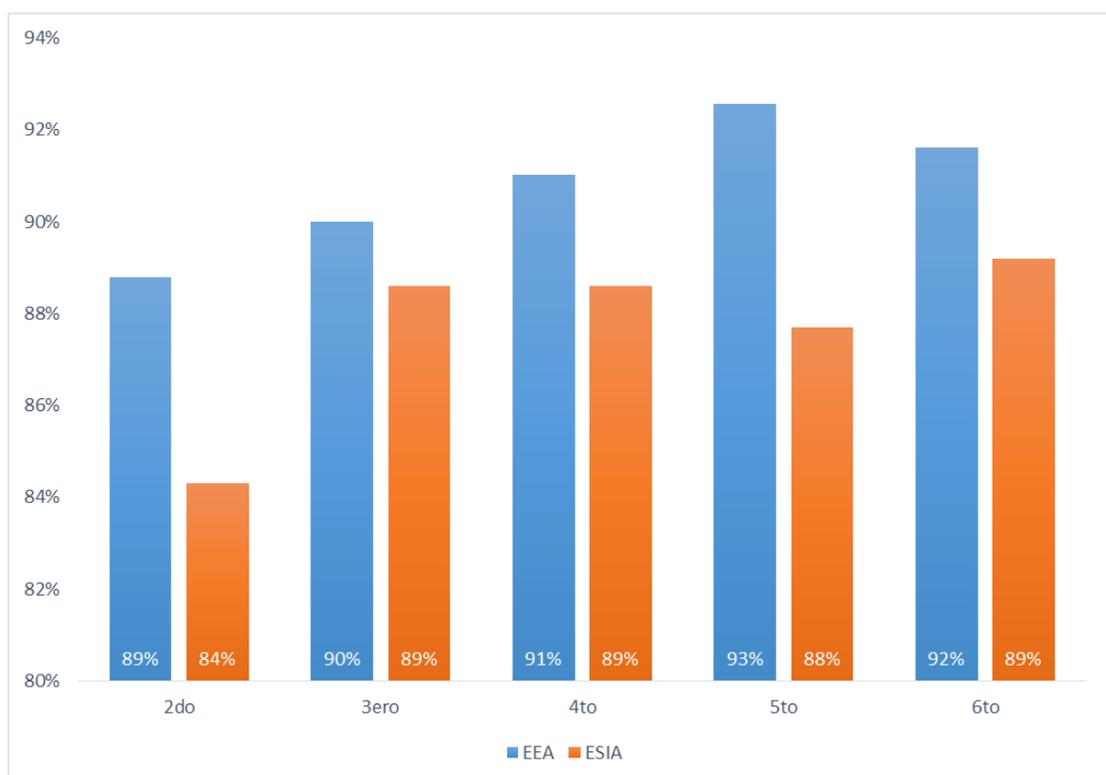


Sobre la Escuela de Sistemas de Información y Auditoría se puede mencionar que los clúster son bastante más homogéneos en sus variable socioeconómicas que los clúster generados para la Escuela de Economía y administración, sin embargo en todos los semestre existe un clúster el cual está compuesto por estudiantes que provienen de colegios particulares

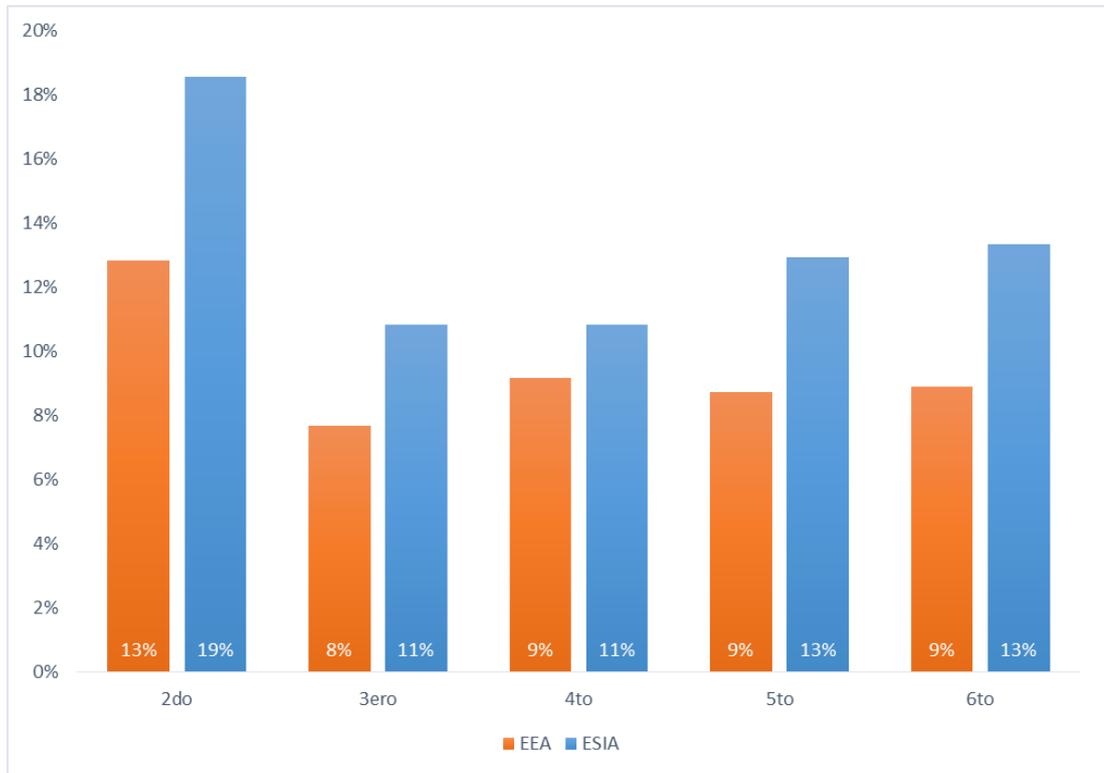
subvencionados y particulares. Sin embargo, para todos los clúster las variables de rendimiento académico son bastante similares en todos los semestres analizados.

La principal diferencia entre los clúster de las Escuelas es que los estudiantes de la Escuela de Sistemas de Información y Auditoría tiene un valor consistentemente más bajo en la variable factor de rendimiento y un valor consistentemente más alto en el porcentaje de reprobación, para todos los semestres.

*Ilustración 10. Factor de Rendimiento entre las Escuelas a través de los semestres*



**Ilustración 11.** Porcentaje de reprobación entre las Escuelas a través de los semestres



## 5. Conclusiones y Propuestas

Las principales conclusiones que nos entregó el análisis de los resultados obtenidos, fue que en la Escuela de Economía y Administración, las variables que tienden a diferenciar a los estudiantes en los primeros semestres son atributos del tipo socio-demográficos y académicas, mientras que para los últimos semestres las diferencias entre los grupos de estudiantes se generan únicamente por las variables socio-demográficas. Esta Escuela posee generalmente las variables de desempeño académico bastante similares en todos los semestres analizados. Los resultados académicos tienden a ser levemente mejores para los estudiantes que provienen mayoritariamente de colegios particulares en los primeros semestres. Sin embargo, esta diferencia comienza a desaparecer a medida que avanzan los semestres, llegando a diferencias académicas muy pequeñas en el sexto semestre manteniendo la diversidad en las variables de socio-demográficas.

Por otra parte, en la Escuela de Sistemas de Información y Auditoría se puede observar que las variables socio-demográficas son similares entre los grupos, en especial en los últimos semestres analizados. Las principales diferencias se pueden ver por el lado del desempeño académico en los primeros semestres, las cual desaparecen con el avance de los estos.

Por otra parte, podemos observar vemos que es importante identificar y conocer los distintos grupos de alumnos existentes en las Escuelas de Pregrado, ya que en el futuro se puedan aplicar medidas particulares a un

determinado grupo de alumnos dependiendo del grupo al que pertenecen con el fin de potenciar y continuar el apoyo entregado a los estudiantes por la Facultad.

Finalmente, se debe destacar que frente a una gran diversidad socio-demográficas en los grupos de estudiantes, se logra disminuir la tasa de reprobación y el factor de rendimiento en la Facultad para todos los grupos de estudiantes a medida que avanzan los semestres. Lo anterior, evidencia el compromiso de la Facultad con la diversidad y la excelencia académica al interior de la Facultad de Economía y Negocios de la Universidad de Chile.

## Bibliografía

[1] Gartner Group 1994. Data mining: The next generation of business intelligence? ATG Research Note T-517-246, Gartner Group Inc., Stamford, CT.

[2] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). "From Data Mining to Knowledge Discovery in Databases", American Association for Artificial Intelligence.

[3] Dunham, M. (2002). "Data Mining, Introductory and Advanced Topics", Prentice Hall.

[4] Pérez López, César. Minería de datos: técnicas y herramientas. Editorial Parninfo, 2007.

[5] Brachman, R., and Anand, T. (1996). "The Process of Knowledge Discovery in Databases: A Human-Centered Approach". In Advances in Knowledge Discovery and Data Mining, 37–58, eds. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. Menlo Park, Calif.: AAAI Press.

[6] Venkatadri, M., and Lokanatha, C. Reddy (2011). "A Review on Data mining from Past to the Future". International Journal of Computer Applications (0975 – 8887) Volume 15– No.7.

[7] Jiménez, À., Álvarez, H. (2010). "Minería de Datos en la Educación". Universidad Carlos III de Madrid.

[8] Baker, R.S.; Yacef, K (2009). "The state of educational data mining in 2009: A review and future visions". JEDM-Journal of Educational Data Mining 1 (1): 2017

[9] Romero, C., Ventura, S. (2010). "Educational Data Mining: A Review of the State of the Art". IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews.

[10] Romero, C., Ventura, S. (JAN-FEB 2013). "WIREs Data Mining Knowl Discov". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.

[11] Huebner, Richard A. "A survey of educational data-mining research". Research in Higher Education Journal. Retrieved 30 March 2014.

- [12] Marbán O. “Modelo matemático paramétrico de estimación para proyectos de Data Mining (DMCOMO)”, Tesis Doctoral, Facultad de Informática, Universidad Politécnica de Madrid, 2003.
- [13] Sharma S., et al. “Framework for formal implementation of the business understanding phase of data mining projects”, *Expert Systems with Applications* 36 (2009) 4114–4124.
- [14] Moine, J., Silva, A., Gordillo, S. “Estudio comparativo de metodologías para minería de datos”. Universidad Tecnológica Nacional, Facultad Regional Rosario.
- [15] Chapman, P., Clinton, J., Keber y otros (2000). “CRISP-DM 1.0 Step by step guide”. SPSS ([www.crisp-dm.org/CRISPWP-0800.pdf](http://www.crisp-dm.org/CRISPWP-0800.pdf)).
- [16] Jiménez, A., Álvarez, H. (2010).”Minería de Datos en la Educación”
- [17] Romero, C., Ventura, S. (2007). “Educational data mining: A survey from 1995 to 2005”. Department of computer Sciences, University of Cordoba, Cordoba, Spain.
- [18] Baker, R. (2010) “Data Mining for Education”. In McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education* (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier
- [19] C. Vialardi, J. Bravo, L. Shafti & A. Ortigosa, (2009) “Recommendation in Higher Education Using Data Mining Techniques”. Universidad Autónoma de Madrid, pp.190-199.
- [20] Ortega, C., Lee, M., Silva, D. & Vásquez, J., (2009) “Carga Académica: Identificación de factores claves en una escuela de Economía y Negocios”, Universidad de Chile, Santiago, Chile

## Anexos

### Anexo 1

Tabla que muestra la lista de referencias de EDM, agrupada de acuerdo al tipo de datos utilizado. Extraído de Romero, C., y Ventura, S., en su artículo de 2010 “Educational Data Mining: A Review of the State of the Art”. (Pág. 603)

TABLE II  
LIST OF EDM REFERENCES GROUPED ACCORDING TO TYPES OF DATA USED.

Type of Data/ Environment	References
Traditional Education	[32], [42], [66], [68], [79], [95], [98], [103], [119], [120], [123], [130], [133], [141], [142], [147], [148], [164], [165], [169], [175], [197], [198], [212], [217], [238], [239], [241], [254], [260], [263], [271], [273], [280], [292], [306].
Web-based Education/ E-learning	[11], [45], [49], [50], [63], [64], [86], [92], [97], [100], [102], [104], [118], [122], [129], [132], [146], [149], [153], [155], [156], [157], [158], [159], [177], [181], [182], [183], [190], [193], [199], [201], [214], [216], [227], [240], [242], [248], [255], [261], [265], [274], [277], [278], [286], [287], [288], [290], [291], [294], [295], [297], [300], [302].
Learning Management Systems	[28], [46], [48], [59], [67], [76], [101], [111], [112], [134], [161], [166], [170], [173], [180], [184], [185], [210], [211], [225], [226], [234], [244], [256], [268], [269], [276], [293], [305].
Intelligent Tutoring Systems	[9], [15], [16], [18], [26], [29], [31], [47], [61], [65], [84], [99], [108], [116], [126], [136], [145], [176], [179], [187], [202], [205], [215], [219], [220], [236], [251], [267], [282], [289], [296].
Adaptive Educational Systems	[4], [23], [37], [38], [69], [93], [94], [107], [125], [127], [135], [138], [140], [150], [162], [163], [189], [221], [229], [247], [259], [262], [270], [279], [281], [303].
Tests/ Questionnaires/	[7], [12], [14], [25], [41], [43], [51], [54], [57], [80], [89], [128], [167], [196], [203], [204], [206], [207], [250], [272], [283], [285], [304].
Texts/ Contents	[1], [3], [40], [73], [109], [143], [152], [160], [237], [249], [253], [266], [285], [299].
Others	[2], [13], [44], [53], [55], [71], [74], [77], [110], [124], [139], [144], [154], [192], [200], [208], [218], [233], [235], [252], [264], [301].

## Anexo 2

Tabla con la cantidad de Ud/Créditos por Periodo de Toma de Ramos, verificando la acumulación de reprobación en los primeros seis semestres de la carrera, utilizando los periodos de otoño 2012 a otoño 2015. Datos obtenidos de la base de Secretaria de Estudios. Elaboración Propia.

<b>Periodo</b>	<b>Total Créditos Reprobados</b>	<b>Suma del semestre 1 al 6</b>	<b>Suma del semestre 1 al 6 / Total</b>
20151	9201	7574	82%
20142	6971	5280	76%
20141	8225	6428	78%
20132	7518	6067	81%
20131	7595	5800	76%
20122	11310	9401	83%
20121	9906	8201	83%

## Anexo 3

Tabla con la cantidad promedio de Ud/Créditos que los alumnos convalidan calculado por Semestre de Ingreso, de otoño 2010 a otoño 2015. Datos obtenidos de la base de Secretaria de Estudios. Elaboración Propia.

<b>Ud/Créditos reconocidos promedio</b>	
<b>Sem_Ingreso Decreto</b>	<b>PromedioDeUd</b>
20101	5,4
20111	4,8
20121	5,0
20131	4,0
20141	3,3
20151	3,2

#### Anexo 4

Detalle sobre los tramos de Ingreso Bruto Familiar que posee el alumno en cuestión. Información entregada por el DEMRE.

INGRESO BRUTO FAMILIAR		
TRAMO	DESDE	HASTA
1	0	144000
2	144001	288000
3	288001	432000
4	432001	576000
5	576001	720000
6	720001	864000
7	864001	1008000
8	1008001	1152000
9	1152001	1296000
10	1296001	1440000
11	1440001	1584000
12	1584001	o más

#### Anexo 5

Detalle sobre la Educación que posee el padre y/o madre del alumno en cuestión. Información entregada por el DEMRE.

EDUCACIÓN DE LOS PADRES	
1	Sin estudios
2	Básica incompleta
3	Básica completa
4	Media incompleta
5	Media completa
6	Centro de formación técnica incompleta
7	Centro de formación técnica completa
8	Universitaria incompleta
9	Universitaria completa
10	Otros estudios
11	Instituto Profesional incompleta
12	Instituto Profesional completa
13	Desconocida o sin información

## Anexo 6.

### Tablas Clúster y Test de Medias

#### A. Semestre de Avance 2

#### Escuela de Economía y Administración

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	Mean (cluster-4)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.917	0.113	0.008	202	0.873	0.156	0.016	90	0.890	0.127	0.007	328	0.872	0.134	0.008	257	4.944	3	873	0.998
%_REPROB SEM ANTERIOR	0.093	0.151	0.011	202	0.144	0.202	0.021	90	0.131	0.172	0.009	328	0.146	0.179	0.011	257	4.014	3	873	0.992
TRAMOS	4.629	0.750	0.053	202	4.700	0.694	0.073	90	4.659	0.690	0.038	328	4.658	0.701	0.044	257	0.218	3	873	0.116
GRUPO_DEPENDENCIA	18.639	3.492	0.246	202	17.333	4.983	0.525	90	19.223	2.807	0.155	328	6.887	2.428	0.151	257	860.297	3	873	1.000
TIENE_TRABAJO_REM	0.011	0.103	0.007	189	0.000	0.000	0.000	80	0.007	0.118	0.007	289	0.012	0.109	0.007	251	0.309	3	805	0.181
HORAS_QUE_DEDICA_TRABAJO	0.099	1.274	0.090	202	0.000	0.000	0.000	90	0.146	2.650	0.146	328	0.226	3.138	0.196	257	0.226	3	873	0.122
GRUPO_FAMILIAR	4.851	2.531	0.178	202	4.711	2.781	0.293	90	4.598	2.641	0.146	328	4.054	1.568	0.098	257	4.945	3	873	0.998
INGRESO_BRUTO_FAM	10.510	2.660	0.187	202	9.889	3.272	0.345	90	10.793	2.479	0.137	328	3.261	1.934	0.121	257	531.796	3	873	1.000
y VIVEN_SUS_PADRES?	4.899	0.541	0.040	179	4.922	0.480	0.055	77	4.755	0.867	0.052	273	4.360	1.310	0.083	250	14.813	3	775	1.000
EDU_PADRE	16.179	1.707	0.132	168	16.260	1.491	0.175	73	16.312	1.591	0.098	266	12.234	3.239	0.221	214	164.295	3	717	1.000
EDU_MADRE	15.860	1.840	0.140	172	15.493	2.075	0.240	75	15.770	1.832	0.112	269	11.963	3.167	0.203	244	142.551	3	756	1.000
SEXO	2.000	0.000	0.000	202	1.000	0.000	0.000	90	1.000	0.000	0.000	328	1.358	0.480	0.030	257	671.877	3	873	1.000
REGIMEN	2.827	0.379	0.027	202	1.000	0.000	0.000	90	3.000	0.000	0.000	328	2.549	0.739	0.046	257	506.634	3	873	1.000

#### Escuela de Sistemas de Información y Auditoría

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	Mean (cluster-4)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.844	0.154	0.013	143	0.821	0.142	0.020	51	0.870	0.156	0.012	162	0.837	0.163	0.020	64	1.707	3	416	0.835
%_REPROB SEM ANTERIOR	0.194	0.220	0.018	143	0.184	0.209	0.029	51	0.161	0.215	0.017	162	0.204	0.239	0.030	64	0.863	3	416	0.539
TRAMOS	4.597	0.796	0.066	144	4.392	0.603	0.084	51	4.586	0.745	0.059	162	4.391	0.633	0.079	64	2.091	3	417	0.899
GRUPO_DEPENDENCIA	12.465	6.173	0.514	144	7.255	2.513	0.352	51	14.012	5.706	0.448	162	10.625	5.669	0.709	64	20.767	3	417	1.000
TIENE_TRABAJO_REM	0.023	0.194	0.017	132	0.020	0.143	0.020	49	0.052	0.250	0.020	154	0.000	0.000	0.000	59	1.164	3	390	0.677
HORAS_QUE_DEDICA_TRABAJO	0.222	1.938	0.161	144	0.392	2.801	0.392	51	0.889	5.467	0.430	162	0.000	0.000	0.000	64	1.252	3	417	0.709
GRUPO_FAMILIAR	4.146	1.843	0.154	144	4.059	1.475	0.207	51	4.043	1.843	0.145	162	3.875	1.732	0.217	64	0.343	3	417	0.206
INGRESO_BRUTO_FAM	6.243	3.820	0.318	144	2.647	1.180	0.165	51	7.340	3.916	0.308	162	6.719	3.574	0.447	64	22.204	3	417	1.000
y VIVEN_SUS_PADRES?	4.600	1.097	0.096	130	4.061	1.547	0.221	49	4.400	1.204	0.100	145	4.702	0.999	0.132	57	3.327	3	377	0.980
EDU_PADRE	14.246	3.032	0.279	118	10.652	2.877	0.424	46	14.624	2.476	0.215	133	14.054	2.706	0.362	56	24.940	3	349	1.000
EDU_MADRE	13.786	2.930	0.261	126	10.936	3.010	0.439	47	14.280	2.636	0.220	143	14.263	2.539	0.336	57	18.465	3	369	1.000
SEXO	2.000	0.000	0.000	144	1.549	0.503	0.070	51	1.000	0.000	0.000	162	1.000	0.000	0.000	64	980.811	3	417	1.000
REGIMEN	2.701	0.489	0.041	144	2.804	0.401	0.056	51	3.000	0.000	0.000	162	1.000	0.000	0.000	64	635.674	3	417	1.000

## B. Semestre de Avance 3

### Escuela de Economía y Administración

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	Mean (cluster-4)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.923	0.116	0.006	437	0.952	0.097	0.005	338	0.908	0.134	0.007	327	0.913	0.132	0.012	130	8.489	3	1228	1.000
%_REPROB SEM ANTERIOR	0.077	0.148	0.007	437	0.049	0.116	0.006	338	0.086	0.166	0.009	327	0.095	0.179	0.016	130	4.980	3	1228	0.998
TRAMOS	4.741	0.670	0.032	437	4.669	0.687	0.037	338	4.752	0.711	0.039	327	4.731	0.680	0.060	130	1.015	3	1228	0.615
GRUPO_DEPENDENCIA	19.325	2.613	0.125	437	19.009	3.030	0.165	338	7.416	2.502	0.138	327	16.885	5.401	0.474	130	1.104.852	3	1228	1.000
TIENE_TRABAJO_REM	0.013	0.153	0.008	384	0.000	0.000	0.000	315	0.013	0.114	0.006	307	0.000	0.000	0.000	122	1.321	3	1124	0.734
HORAS_QUE_DEDICA_TRABAJO	0.192	2.719	0.130	437	0.000	0.000	0.000	338	0.324	3.910	0.216	327	0.000	0.000	0.000	130	1.058	3	1228	0.634
GRUPO_FAMILIAR	4.604	2.679	0.128	437	4.855	2.572	0.140	338	3.859	1.787	0.099	327	4.762	2.493	0.219	130	10.821	3	1228	1.000
INGRESO_BRUTO_FAM	10.801	2.412	0.115	437	10.666	2.445	0.133	338	3.450	2.005	0.111	327	9.908	3.097	0.272	130	717.245	3	1228	1.000
y VIVEN_SUS_PADRES?	4.792	0.763	0.040	361	4.890	0.566	0.033	299	4.425	1.197	0.068	306	4.868	0.617	0.058	114	18.125	3	1076	1.000
EDU_PADRE	16.371	1.570	0.084	350	16.264	1.835	0.108	288	12.473	3.113	0.194	258	16.009	2.040	0.198	106	194.914	3	998	1.000
EDU_MADRE	15.980	1.734	0.093	348	15.934	1.795	0.106	289	12.620	2.823	0.164	295	15.355	2.236	0.213	110	161.502	3	1038	1.000
SEXO	1.000	0.000	0.000	437	2.000	0.000	0.000	338	1.385	0.487	0.027	327	1.000	0.000	0.000	130	1.119.382	3	1228	1.000
REGIMEN	3.000	0.000	0.000	437	2.805	0.397	0.022	338	2.514	0.747	0.041	327	1.000	0.000	0.000	130	730.924	3	1228	1.000

### Escuela de Sistemas de Información y Auditoría

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	Mean (cluster-4)	Std. Dev.	Std. Err.	Count	Mean (cluster-5)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.881	0.144	0.011	177	0.880	0.117	0.016	53	0.899	0.121	0.014	74	0.873	0.157	0.017	86	0.892	0.131	0.013	98	0.484	4	483	0.252
%_REPROB SEM ANTERIOR	0.128	0.187	0.014	177	0.087	0.143	0.020	53	0.113	0.200	0.023	74	0.116	0.195	0.021	86	0.117	0.177	0.018	98	0.503	4	483	0.266
TRAMOS	4.497	0.732	0.055	177	4.642	0.653	0.090	53	4.514	0.726	0.084	74	4.593	0.742	0.080	86	4.541	0.720	0.073	98	0.558	4	483	0.307
GRUPO_DEPENDENCIA	13.475	5.710	0.429	177	8.019	2.469	0.339	53	17.162	4.539	0.528	74	10.407	6.194	0.668	86	7.500	2.513	0.254	98	56.569	4	483	1.000
TIENE_TRABAJO_REM	0.036	0.187	0.014	167	0.000	0.000	0.000	50	0.000	0.000	0.000	68	0.013	0.114	0.013	77	0.036	0.188	0.021	83	1.278	4	440	0.722
HORAS_QUE_DEDICA_TRABAJO	0.644	3.730	0.280	177	0.000	0.000	0.000	53	0.000	0.000	0.000	74	0.233	2.157	0.233	86	0.510	3.341	0.337	98	1.052	4	483	0.620
GRUPO_FAMILIAR	4.006	1.711	0.129	177	3.943	1.524	0.209	53	4.419	1.775	0.206	74	3.884	1.862	0.201	86	3.531	1.906	0.193	98	2.747	4	483	0.972
INGRESO_BRUTO_FAM	6.825	3.796	0.285	177	2.849	1.183	0.163	53	9.095	3.286	0.382	74	6.163	3.626	0.391	86	3.714	1.856	0.187	98	46.562	4	483	1.000
y VIVEN_SUS_PADRES?	4.630	0.990	0.078	162	4.520	1.111	0.157	50	4.912	0.511	0.062	68	4.893	0.669	0.077	75	4.570	1.173	0.132	79	2.740	4	429	0.972
EDU_PADRE	14.688	2.668	0.215	154	11.360	2.776	0.393	50	15.152	2.200	0.271	66	13.560	2.703	0.312	75	12.789	3.760	0.446	71	19.120	4	411	1.000
EDU_MADRE	14.194	2.571	0.203	160	11.735	2.871	0.410	49	14.641	2.065	0.258	64	14.080	2.437	0.281	75	12.642	2.675	0.297	81	14.999	4	424	1.000
SEXO	1.000	0.000	0.000	177	1.623	0.489	0.067	53	2.000	0.000	0.000	74	1.000	0.000	0.000	86	2.000	0.000	0.000	98	1.032.012	4	483	1.000
REGIMEN	3.000	0.000	0.000	177	2.811	0.395	0.054	53	2.905	0.295	0.034	74	1.012	0.108	0.012	86	2.510	0.502	0.051	98	761.891	4	483	1.000

## C. Semestre de Avance 4

### Escuela de Economía y Administración

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.935	0.094	0.003	870	0.907	0.118	0.009	193	0.889	0.138	0.008	271	21.045	2	1331	1.000
%_REPROB SEM ANTERIOR	0.066	0.130	0.004	870	0.098	0.156	0.011	193	0.111	0.175	0.011	271	11.513	2	1331	1.000
TRAMOS	4.749	0.693	0.023	870	4.778	0.718	0.052	194	4.793	0.808	0.049	271	0.438	2	1332	0.354
GRUPO_DEPENDENCIA	19.828	1.302	0.044	870	9.562	4.955	0.356	194	8.303	3.810	0.231	271	2.422.054	2	1332	1.000
TIENE_TRABAJO_REM	0.010	0.143	0.005	780	0.006	0.075	0.006	177	0.034	0.238	0.015	261	2.538	2	1215	0.921
HORAS_QUE_DEDICA_TRABAJO	0.083	1.695	0.057	870	0.041	0.574	0.041	194	0.317	3.215	0.195	271	1.611	2	1332	0.800
GRUPO_FAMILIAR	4.749	2.620	0.089	870	3.644	1.836	0.132	194	4.059	1.529	0.093	271	22.541	2	1332	1.000
INGRESO_BRUTO_FAM	10.909	2.242	0.076	870	4.418	2.851	0.205	194	4.609	2.777	0.169	271	1.024.766	2	1332	1.000
¿VIVEN_SUS_PADRES?	4.853	0.648	0.024	735	4.503	1.119	0.085	175	4.484	1.134	0.071	256	24.162	2	1163	1.000
EDU_PADRE	16.472	1.300	0.048	724	13.459	3.216	0.264	148	13.393	2.884	0.191	229	273.508	2	1098	1.000
EDU_MADRE	15.957	1.704	0.063	723	13.491	2.529	0.196	167	12.972	2.988	0.189	249	218.128	2	1136	1.000
SEXO	1.370	0.483	0.016	870	2.000	0.000	0.000	194	1.000	0.000	0.000	271	373.464	2	1332	1.000
REGIMEN	2.706	0.658	0.022	870	2.608	0.489	0.035	194	2.173	0.987	0.060	271	57.001	2	1332	1.000

### Escuela de Sistemas de Información y Auditoría

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	Mean (cluster-4)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.878	0.137	0.012	130	0.902	0.118	0.013	78	0.887	0.107	0.009	132	0.853	0.143	0.016	76	2.074	3	412	0.897
%_REPROB SEM ANTERIOR	0.127	0.175	0.015	130	0.114	0.180	0.020	78	0.096	0.144	0.012	132	0.162	0.213	0.024	76	2.433	3	412	0.935
TRAMOS	4.763	0.753	0.066	131	4.641	0.837	0.095	78	4.727	0.792	0.069	132	4.526	0.774	0.089	76	1.678	3	413	0.829
GRUPO_DEPENDENCIA	7.901	2.477	0.216	131	16.282	4.864	0.551	78	7.121	2.481	0.216	132	18.947	3.089	0.354	76	338.988	3	413	1.000
TIENE_TRABAJO_REM	0.023	0.151	0.013	129	0.000	0.000	0.000	70	0.017	0.131	0.012	115	0.000	0.000	0.000	67	0.999	3	377	0.607
HORAS_QUE_DEDICA_TRABAJO	0.840	5.790	0.506	131	0.000	0.000	0.000	78	0.455	3.679	0.320	132	0.000	0.000	0.000	76	1.130	3	413	0.664
GRUPO_FAMILIAR	4.176	1.406	0.123	131	4.192	2.052	0.232	78	3.712	2.044	0.178	132	4.250	2.149	0.246	76	2.031	3	413	0.891
INGRESO_BRUTO_FAM	4.176	2.332	0.204	131	8.987	3.379	0.383	78	3.159	1.693	0.147	132	9.132	3.226	0.370	76	145.266	3	413	1.000
yVIVEN_SUS_PADRES?	4.640	0.979	0.088	125	4.731	0.863	0.105	67	4.540	1.165	0.110	113	4.910	0.514	0.063	67	2.233	3	368	0.916
EDU_PADRE	12.917	2.767	0.253	120	15.953	1.578	0.197	64	12.387	3.250	0.316	106	15.485	2.335	0.287	66	36.579	3	352	1.000
EDU_MADRE	12.816	3.214	0.287	125	15.078	2.073	0.259	64	12.148	3.252	0.313	108	15.045	2.123	0.261	66	22.947	3	359	1.000
SEXO	1.000	0.000	0.000	131	2.000	0.000	0.000	78	2.000	0.000	0.000	132	1.000	0.000	0.000	76	\$null\$	\$null\$	\$null\$	\$null\$
REGIMEN	2.374	0.931	0.081	131	2.872	0.336	0.038	78	2.485	0.502	0.044	132	2.368	0.936	0.107	76	8.888	3	413	1.000

## D. Semestre de Avance 5

### Escuela de Economía y Administración

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	Mean (cluster-4)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.942	0.081	0.005	292	0.927	0.098	0.006	292	0.931	0.082	0.008	95	0.903	0.120	0.007	300	7.734	3	975	1.000
%_REPROB SEM ANTERIOR	0.073	0.147	0.009	292	0.075	0.136	0.008	292	0.098	0.169	0.017	95	0.104	0.165	0.010	300	2.891	3	975	0.965
TRAMOS	4.623	0.738	0.043	292	4.575	0.794	0.046	292	4.800	0.846	0.087	95	4.830	0.764	0.044	300	6.778	3	975	1.000
GRUPO_DEPENDENCIA	18.699	3.446	0.202	292	19.452	2.280	0.133	292	16.000	6.341	0.651	95	8.750	4.444	0.257	300	470.803	3	975	1.000
TIENE_TRABAJO_REM	0.000	0.000	0.000	292	0.000	0.000	0.000	292	0.000	0.000	0.000	95	0.000	0.000	0.000	300	\$null\$	\$null\$	\$null\$	\$null\$
HORAS_QUE_DEDICA_TRABAJO	0.000	0.000	0.000	292	0.000	0.000	0.000	292	0.000	0.000	0.000	95	0.000	0.000	0.000	300	\$null\$	\$null\$	\$null\$	\$null\$
GRUPO_FAMILIAR	5.510	1.724	0.101	292	5.421	1.690	0.099	292	5.853	1.902	0.195	95	4.260	1.229	0.071	300	44.894	3	975	1.000
INGRESO_BRUTO_FAM	10.171	2.738	0.160	292	10.664	2.489	0.146	292	10.558	2.500	0.256	95	3.733	2.178	0.126	300	509.349	3	975	1.000
yVIVEN_SUS_PADRES?	4.887	0.572	0.033	292	4.949	0.390	0.023	292	4.842	0.673	0.069	95	4.517	1.135	0.066	300	18.776	3	975	1.000
EDU_PADRE	16.473	1.211	0.071	292	16.524	1.144	0.067	292	16.189	1.606	0.165	95	12.563	3.194	0.184	300	250.451	3	975	1.000
EDU_MADRE	16.062	1.653	0.097	292	15.993	1.675	0.098	292	15.926	1.770	0.182	95	12.553	2.505	0.145	300	214.202	3	975	1.000
SEXO	2.000	0.000	0.000	292	1.000	0.000	0.000	292	1.021	0.144	0.015	95	1.350	0.478	0.028	300	769.248	3	975	1.000
REGIMEN	2.795	0.405	0.024	292	3.000	0.000	0.000	292	1.011	0.103	0.011	95	2.463	0.803	0.046	300	408.224	3	975	1.000

### Escuela de Sistemas de Información y Auditoría

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	Mean (cluster-4)	Std. Dev.	Std. Err.	Count	Mean (cluster-5)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.913	0.090	0.007	155	0.888	0.139	0.028	24	0.888	0.112	0.046	6	0.883	0.124	0.012	115	0.849	0.145	0.020	52	3.318	4	347	0.989
%_REPROB SEM ANTERIOR	0.086	0.133	0.011	155	0.092	0.172	0.035	24	0.136	0.158	0.064	6	0.146	0.188	0.018	115	0.143	0.204	0.028	52	2.697	4	347	0.969
TRAMOS	4.729	0.658	0.053	155	4.917	0.717	0.146	24	4.167	0.408	0.167	6	4.548	0.830	0.077	115	4.453	0.822	0.113	53	3.256	4	348	0.988
GRUPO_DEPENDENCIA	10.677	5.924	0.476	155	7.292	2.545	0.519	24	10.000	0.000	0.000	6	13.130	6.014	0.561	115	11.415	6.749	0.927	53	6.075	4	348	1.000
TIENE_TRABAJO_REM	0.000	0.000	0.000	135	0.000	0.000	0.000	24	1.500	0.548	0.224	6	0.019	0.136	0.013	107	0.000	0.000	0.000	49	300.257	4	316	1.000
HORAS_QUE_DEDICA_TRABAJO	0.000	0.000	0.000	155	0.000	0.000	0.000	24	49.167	24.580	10.035	6	0.087	0.656	0.061	115	0.000	0.000	0.000	53	403.592	4	348	1.000
GRUPO_FAMILIAR	3.568	1.886	0.152	155	4.208	0.932	0.190	24	4.500	1.975	0.806	6	4.348	1.747	0.163	115	4.264	1.799	0.247	53	3.865	4	348	0.996
INGRESO_BRUTO_FAM	5.129	3.710	0.298	155	2.792	1.351	0.276	24	3.167	0.983	0.401	6	6.478	3.648	0.340	115	5.943	3.342	0.459	53	7.152	4	348	1.000
yVIVEN_SUS_PADRES?	4.588	1.143	0.100	131	4.750	0.847	0.173	24	5.000	0.000	0.000	6	4.748	0.837	0.081	107	4.510	1.120	0.160	49	0.842	4	312	0.500
EDU_PADRE	13.435	3.219	0.289	124	10.417	3.175	0.648	24	13.667	2.582	1.054	6	14.874	2.581	0.254	103	14.167	2.282	0.329	48	12.680	4	300	1.000
EDU_MADRE	13.448	2.775	0.248	125	10.958	1.367	0.279	24	11.667	0.816	0.333	6	14.131	2.978	0.288	107	14.000	2.102	0.300	49	8.164	4	306	1.000
SEXO	2.000	0.000	0.000	155	1.417	0.504	0.103	24	1.000	0.000	0.000	6	1.000	0.000	0.000	115	1.000	0.000	0.000	53	1.223.598	4	348	1.000
REGIMEN	2.594	0.493	0.040	155	2.667	0.637	0.130	24	1.333	0.816	0.333	6	3.000	0.000	0.000	115	1.000	0.000	0.000	53	274.376	4	348	1.000

## E. Semestre de Avance 6

### Escuela de Economía y Administración

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	Mean (cluster-4)	Std. Dev.	Std. Err.	Count	Mean (cluster-5)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.901	0.098	0.007	172	0.922	0.085	0.004	395	0.941	0.082	0.005	298	0.920	0.083	0.006	195	0.897	0.099	0.009	123	8.290	4	1178	1.000
%_REPROB SEM ANTERIOR	0.111	0.165	0.013	172	0.078	0.134	0.007	395	0.069	0.125	0.007	298	0.075	0.127	0.009	195	0.113	0.173	0.016	123	4.139	4	1178	0.998
TRAMOS	4.750	0.879	0.067	172	4.701	0.769	0.039	395	4.822	0.728	0.042	298	4.882	0.761	0.054	195	4.886	0.791	0.071	123	2.682	4	1178	0.970
GRUPO_DEPENDENCIA	8.750	4.228	0.322	172	19.519	2.143	0.108	395	19.664	1.804	0.104	298	13.359	7.007	0.502	195	8.780	3.093	0.279	123	457.720	4	1178	1.000
TIENE_TRABAJO_REM	0.013	0.114	0.009	153	0.021	0.179	0.010	340	0.018	0.177	0.011	284	0.024	0.217	0.017	169	0.000	0.000	0.000	123	0.448	4	1064	0.226
HORAS_QUE_DEDICA_TRABAJO	0.035	0.323	0.025	172	0.111	1.070	0.054	395	0.178	2.626	0.152	298	0.097	0.961	0.069	195	0.000	0.000	0.000	123	0.416	4	1178	0.203
GRUPO_FAMILIAR	3.698	2.041	0.156	172	4.253	2.600	0.131	395	4.899	2.420	0.140	298	4.179	2.666	0.191	195	4.309	1.505	0.136	123	7.491	4	1178	1.000
INGRESO_BRUTO_FAM	3.971	2.739	0.209	172	10.322	2.748	0.138	395	10.379	2.679	0.155	298	7.744	4.070	0.291	195	3.675	2.125	0.192	123	255.704	4	1178	1.000
¿VIVEN_SUS_PADRES?	4.463	1.183	0.097	149	4.838	0.679	0.038	315	4.900	0.539	0.033	271	4.688	0.953	0.075	160	4.695	0.911	0.084	118	8.157	4	1008	1.000
EDU_PADRE	13.812	3.063	0.266	133	16.349	1.383	0.078	312	16.408	1.915	0.118	265	14.743	2.874	0.233	152	12.619	3.455	0.325	113	81.999	4	970	1.000
EDU_MADRE	13.430	2.656	0.223	142	15.773	1.850	0.105	309	16.071	1.636	0.100	266	14.168	3.055	0.245	155	12.595	2.984	0.277	116	75.664	4	983	1.000
SEXO	2.000	0.000	0.000	172	1.000	0.000	0.000	395	2.000	0.000	0.000	298	1.000	0.000	0.000	195	1.000	0.000	0.000	123	\$null\$	\$null\$	\$null\$	\$null\$
REGIMEN	2.587	0.494	0.038	172	3.000	0.000	0.000	395	2.826	0.389	0.023	298	1.000	0.000	0.000	195	2.992	0.090	0.008	123	2.000.299	4	1178	1.000

### Escuela de Sistemas de Información y Auditoría

Field	Mean (cluster-1)	Std. Dev.	Std. Err.	Count	Mean (cluster-2)	Std. Dev.	Std. Err.	Count	Mean (cluster-3)	Std. Dev.	Std. Err.	Count	F-Test	df1	df2	Importance
Factor_Rend sem anterior	0.900	0.105	0.010	103	0.907	0.095	0.007	169	0.869	0.105	0.015	48	2.674	2	317	0.929
%_REPROB SEM ANTERIOR	0.111	0.156	0.015	103	0.128	0.182	0.014	169	0.161	0.178	0.026	48	1.366	2	317	0.743
TRAMOS	4.583	0.799	0.079	103	4.615	0.824	0.063	169	4.646	0.956	0.138	48	0.103	2	317	0.098
GRUPO_DEPENDENCIA	12.864	6.007	0.592	103	10.562	5.861	0.451	169	10.521	6.461	0.933	48	5.182	2	317	0.994
TIENE_TRABAJO_REM	0.020	0.140	0.014	101	0.000	0.000	0.000	154	0.143	0.417	0.064	42	11.001	2	294	1.000
HORAS_QUE_DEDICA_TRABAJO	0.485	4.456	0.439	103	0.000	0.000	0.000	169	2.667	10.769	1.554	48	5.652	2	317	0.996
GRUPO_FAMILIAR	4.583	1.332	0.131	103	3.834	1.782	0.137	169	3.688	1.764	0.255	48	7.987	2	317	1.000
INGRESO_BRUTO_FAM	6.155	3.816	0.376	103	4.882	3.592	0.276	169	5.292	3.087	0.446	48	4.021	2	317	0.981
¿VIVEN_SUS_PADRES?	4.820	0.716	0.072	100	4.653	1.030	0.084	150	4.302	1.282	0.196	43	4.211	2	290	0.984
EDU_PADRE	13.879	3.008	0.302	99	13.182	3.803	0.325	137	13.625	3.295	0.521	40	1.193	2	273	0.695
EDU_MADRE	13.768	3.047	0.306	99	13.336	2.955	0.247	143	13.714	2.472	0.381	42	0.725	2	281	0.515
SEXO	1.000	0.000	0.000	103	2.000	0.000	0.000	169	1.000	0.000	0.000	48	\$null\$	\$null\$	\$null\$	\$null\$
REGIMEN	3.000	0.000	0.000	103	2.586	0.494	0.038	169	1.000	0.000	0.000	48	521.357	2	317	1.000