



FACULTAD
DE CIENCIAS
ECONÓMICAS



Universidad
Nacional
de Córdoba

REPOSITORIO DIGITAL UNIVERSITARIO (RDU-UNC)

Explotación de información aplicada a la caracterización de patrones socio-económicos de la población estudiantil de carreras de ciencias económicas

Laura Cecilia Díaz, Sebastian Martins, José María Las Heras,
Ramón García-Martínez

Ponencia presentada en XI Congreso de Tecnología en Educación y Educación en Tecnología
(TE&ET) realizado en 2016 en la Facultad de Informática, Ciencias de la Comunicación y
Técnicas Especiales – Universidad de Morón. Buenos Aires, Argentina



Esta obra está bajo una [Licencia Creative Commons Atribución 4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Explotación de Información Aplicada a la Caracterización de Patrones Socio-Económicos de la Población Estudiantil de Carreras de Ciencias Económicas

Laura Cecilia Díaz^{1,2}, Sebastian Martins^{2,3}, José María Las Heras⁴, Ramón García-Martínez²

1. Facultad de Ciencias Exactas, Físicas y Naturales. Universidad Nacional de Córdoba

2. Grupo de Investigación en Sistemas de Información. Universidad Nacional de Lanús

3. Doctorado en Ciencias Informáticas Facultad de Informática. Universidad Nacional de La Plata

4. Facultad de Ciencias Económicas. Universidad Nacional de Córdoba

laura.diaz@unc.edu.ar, smartins089@gmail.com, lasherasasociados@hotmail.com, rgm1960@yahoo.com

Resumen

El conocimiento de características del estudiante es un factor significativo para la mejora de los procesos de aprendizaje, constituyéndose en un tema de relevancia en Educación Superior en contextos de masividad, en tanto contribuye al establecimiento de políticas de mejora de la calidad educativa. En esta comunicación se presentan los resultados de la utilización de procesos de explotación de información orientados al descubrimiento de patrones de rendimiento académico en la carrera de Contador de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba.

Palabras claves: Gestión de la Educación Superior, Carreras de Ciencias Económicas, Estudiantes, Explotación de Información.

1. Introducción

La Explotación de Información es la sub-disciplina de los Sistemas de Información que aporta las herramientas para la transformación de información en conocimiento [García-Martínez et al., 2015]. Ha sido definida como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información [Britos et al., 2005]. Un Proceso de Explotación de Información se define, como un grupo de tareas relacionadas lógicamente [Curtis et al., 1992] que, a partir de un conjunto de información con un cierto grado de valor para la organización, se ejecuta

para lograr otro, con un grado de valor mayor que el inicial [Ferreira et al., 2005]. Adicionalmente, existe una variedad de técnicas de minería de datos, en su mayoría provenientes del campo del Aprendizaje Automático [García-Martínez et al., 2013], susceptibles de ser utilizadas en cada uno de estos procesos. La Ingeniería de Explotación de Información (IEI) entiende en los procesos y las metodologías utilizadas para: ordenar, controlar y gestionar la tarea de encontrar patrones de conocimiento en masas de información [Martins, 2014].

En el multidisciplinar escenario de la Educación Superior en contextos de masividad se debaten cuestiones axiológicas, epistemológicas y metodológicas asociadas con la accesibilidad a los más altos niveles del conocimiento, con la construcción compleja de saberes, en un momento histórico caracterizado por vertiginosos cambios tecnológicos que impactan en las sociedades actuales [Juarros, 2006]. Es en este escenario que el uso de IEI ofrece la oportunidad de descubrir comportamientos socioeconómicos, académicos, cognitivos, entre otros, de los sujetos en procesos de aprendizaje, que con otras metodologías no serían necesariamente detectados [Kuna et al., 2010].

En este contexto, en esta comunicación se presentan las preguntas de investigación que se formularon los autores sobre rendimiento académico de alumnos de los primeros años de cursos de ciencias económicas (Sección 2), se describen los materiales y métodos utilizados para el descubrimiento de patrones de comportamiento (Sección 3), se muestran

los resultados obtenidos y una interpretación tentativa (Sección 4), y se formulan conclusiones preliminares sobre los hallazgos y se plantean futuras líneas de investigación (Sección 5).

2. Preguntas de Investigación

En el contexto descrito en la Sección 1 se plantea la pregunta general:

¿Cómo se caracteriza el rendimiento académico de los estudiantes de la carrera de Contador de la Facultad de Ciencias Económicas de la Universidad Nacional de Córdoba, tomando a la asignatura Administración y Sistemas de Información Gubernamental como eje del análisis?

Se ha desglosado la pregunta general en las siguientes preguntas específicas:

- *¿Qué similitudes socioeconómicas hay entre estos estudiantes? ¿Cómo se caracterizan?*
- *¿Qué similitudes en relación a su procedencia geográfica? ¿Qué características se encuentran en estos grupos?*

3. Materiales y Métodos

Las variables escogidas para el desarrollo del presente trabajo tienen en cuenta dos aspectos: el protocolo desarrollado y aplicado para el caso de estudio de los estudiantes de las carreras de Ingeniería de la Facultad de Ciencias Exactas, Físicas y Naturales [Díaz et al., 2015] y los conocimientos a priori del experto del dominio, Profesor Titular de la asignatura tomada como eje para el presente análisis.

Se tomaron trece variables para el análisis como representativas de los aspectos involucrados en las preguntas de investigación. Tres de ellas hacen referencia a la procedencia del estudiantes: [i] si es o no Argentino, [ii] si procede o no de la provincia de Córdoba, y [iii] si es o no Córdoba capital su localidad de procedencia. Seis variables se relacionan con sus aspectos personales,

familiares y socio económicos: [iv] su género, [v] si costea o no su estudios con aportes de su familia, [vi] si costea o no su estudios con su trabajo, [vii] si es o no beneficiario de beca de ayuda económica y [viii] el máximo nivel de estudios alcanzado por su madre, e [ix] ídem por su padre. Por último, cuatro son las variables relacionadas con su rendimiento académico: [x] si aprobó o no la asignatura al finalizar la cursada, [xi] si demoró en cursar la asignatura, esto es si la cursó conforme al plan de carreras o si lo hizo con alguna demora y en tal caso, cuánto, [xii] en qué medida cumple el plan de carrera actualmente y [xiii] cuál fue su desempeño en el primer cuatrimestre de su carrera. Se decidió no tomar como variables representativas del rendimiento académico ni a las calificaciones de la asignatura, ni al promedio de calificaciones con y sin aplazos, en razón del sesgo proveniente de las subjetividades de los evaluadores al generar esas calificaciones, y de las diversas normativas vigentes en las distintas unidades académicas. Las escogidas resultan más permeables al momento de realizar comparaciones o generar estándares. En esta sección se describe la base de datos utilizada en la explotación de información (Sección 3.1), se presentan los procesos de explotación de información elegidos (Sección 3.2), y las tecnologías de minería de datos aplicadas en los procesos (Sección 3.3).

3.1. Descripción de la Base de Datos

Se dispone de una base de datos con más de 6500 registros de alumnos registrados en el sistema SIU_Guaraní para la carrera Contador en la Facultad de Ciencias Económicas (FCE) de la Universidad Nacional de Córdoba. Del total de alumnos, se extrajo la información de aquellos estudiantes que se inscribieron a la materia Administración y Sistemas de Información Gubernamental en los años 2012 a 2014 tanto del tipo académico, como socio-económico y de situación geográfica. A continuación se describen detalladamente las variables consideradas para la resolución de las preguntas/problemas:

- Procedencia del sustento económico del alumno: trabajo propio, familia y/o de beca. Cada una representada como una variable booleana donde 1 indica que su fuente de ingreso proviene de dicha variable.
- Últimos estudios alcanzados por su padre y madre, representados por una escala de 0 a 4, en la cual 0 indica que el padre o la madre no tienen estudios, 1 que tienen estudios primarios completos o secundarios incompletos, 2 estudios secundarios completos o superiores incompletos, 3 superiores completos y 4 posee estudios de posgrado.
- Género del estudiante (1 indica que es Masculino, 0 Femenino).
- Ubicación de procedencia expresada mediante tres variables booleanas: si es argentino, si es de la Provincia de Córdoba, y si es de Córdoba Capital. En caso afirmativo su valor es 1, 0 en caso contrario.
- Si el alumno aprobó la materia durante el mismo año que realizó la cursada (asignando 1 en caso afirmativo, 0 en caso contrario).
- Si el alumno realizó la cursada de la materia acorde a lo establecido en el plan de estudios (asignando 1 en caso afirmativo, 0 en caso contrario).
- Dos variables que determinan el rendimiento del alumno en su primer año de ingreso (categorizados en 4 niveles incrementales desde 0 a 3 acorde a la cantidad de materias que realizan) y su desempeño en el total de años cursados respecto al plan de estudios (categorizados en 5 niveles incrementales de 0 a 4 de acuerdo al grado de materias aprobadas respecto del plan de estudios y del año en el que ingresaron).

Es relevante destacar que a partir de la base de datos original (con los datos en crudo) y la selección de las variables representativas para el dominio de interés, fueron realizadas distintas tareas de pre-procesado para adecuar

la información a las necesidades y requerimientos específicos del proyecto.

3.2. Descripción de la Base de Datos

Los procesos de explotación de información definen las técnicas o algoritmos a utilizar en base a las características del problema de explotación. En [García-Martínez et al., 2013] se definen 5 tipos de procesos: descubrimiento de reglas de comportamiento, descubrimiento de grupos, descubrimiento de atributos significativos, descubrimiento de reglas de pertenencia a grupos y ponderación de reglas de comportamiento o de pertenencia a grupos. Acorde a los intereses de este trabajo, es relevante describir los siguientes dos procesos:

a) Descubrimiento de reglas de comportamiento:

El proceso de descubrimiento de reglas de comportamiento (figura 1) aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema.

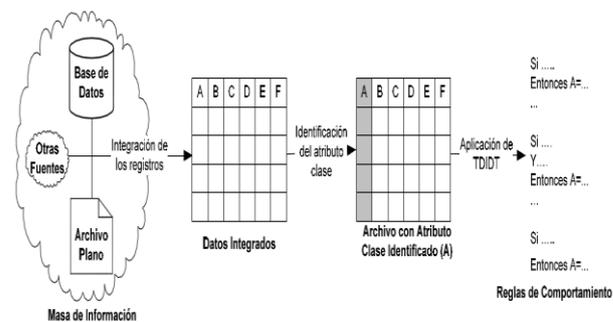


Fig. 1. Proceso Descubrimiento de reglas de comportamiento

Para este proceso se propone la utilización de algoritmos de inducción TDIDT para descubrir las reglas de comportamiento de cada atributo clase. Como resultado de la aplicación del algoritmo de inducción TDIDT al atributo clase se obtiene un conjunto de reglas que definen el comportamiento de dicha clase.

b) Descubrimiento de reglas de pertenencia a grupos:

El proceso de descubrimiento de reglas de pertenencia a grupos (figura 2) aplica cuando

se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida “a priori”, pero presente en la masa de información disponible sobre el dominio de problema.

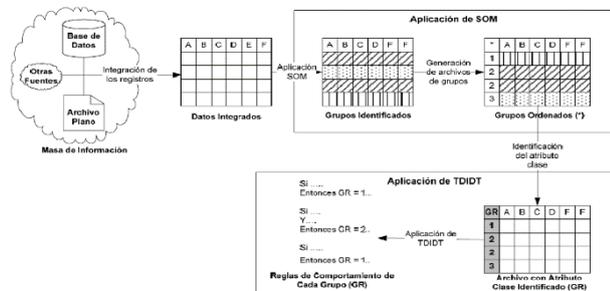


Fig. 2. Proceso Descubrimiento de reglas de pertenencia a grupos

Para el descubrimiento de reglas de pertenencia a grupos se propone la utilización de algoritmos de agrupamiento (por ej.: SOM, K-MEANS) para el hallazgo de los mismos y; una vez identificados, la utilización de algoritmos de inducción (por ej.: de la familia TDIDT) para establecer las reglas de pertenencia a cada uno.

3.3. Tecnologías de Minería de Datos Aplicadas en los Procesos

Los tipos de algoritmos de minería de datos relevantes para el caso de aplicación, se identifican a partir de los procesos de explotación de información aplicados. De dicho análisis se identifican dos tipos de algoritmos: de la familia TDIDT y de clustering.

El algoritmo de clasificación mediante árboles de decisión utilizado es C4.5 [Quinlan, 1993], el cual es descendiente de los algoritmos CLS e ID3. El algoritmo c4.5 clasifica el conjunto de datos mediante la generación de árboles de decisiones, los cuales consisten en una lista de reglas de la forma “si A y B y ... entonces clase X” a partir de las cuales se pueden identificar todas las reglas que describen a una clase.

Se utilizaron distintos algoritmos de clustering pertenecientes a distintas familias o tipología de algoritmos, con el objetivo de poder identificar distintas características

complementarias a partir del modo en el cual cada tipología de algoritmo comprende el conjunto de datos. Los algoritmos utilizados son: Mapas Auto-Organizados (SOM), K-Means y Hierarchical Agglomerative Clustering (HAC).

El algoritmo SOM [Kohonen, 1995], perteneciente a la familia de algoritmos de agrupamiento basado en modelos, es un tipo de red neuronal la cual utiliza una función de cercanía (o vecindario) de registros con el objetivo de determinar las propiedades topológicas del espacio de entrenamiento.

El algoritmo K-means [MacQueen, 1967], perteneciente a la familia de algoritmos de particionamiento, es un método iterativo simple para particionar un conjunto de datos en un número K de clusters, donde K es un valor definido por el usuario. La idea principal del algoritmo es definir un conjunto K de centroides (uno por cada cluster a identificar), a los cuales se les asocia los puntos (un punto representa un registro) más cercanos. Una vez asignada toda la base de datos, se recalculan la ubicación de los centroides como baricentros de los registros asignados a cada uno, y se vuelve a repetir el proceso de agrupamiento hasta que no se produzcan más cambios en la ubicación de los centroides.

El algoritmo HAC, perteneciente a la familia de algoritmos Jerárquicos, cuyo método de agrupamiento es Bottom-Up (de abajo hacia arriba) donde cada registro representa un cluster por sí mismo. Posteriormente, cada cluster es agrupado en clusters más generales de forma sucesiva hasta el nivel deseado, generando un dendrograma cuyas agrupaciones se encuentran a una altura similar.

4. Resultados e Interpretación

El objetivo general en esta instancia consiste en realizar un análisis exploratorio de las características socioeconómicas de los estudiantes de ciencias económicas, representativa de asignaturas en contextos de

masividad en la UNC, en relación a su rendimiento académico.

En la sección 4.1 se presenta una descripción estadística de las variables consideradas para las preguntas de investigación previamente identificadas. Posteriormente se presentan los resultados obtenidos de aplicar los procesos de explotación de información para determinar patrones respecto al desempeño académico a partir de las variables socioeconómicas y de procedencia geográfica (sección 4.2) de la población analizada. Adicionalmente, se considera relevante destacar que se han utilizado distintos tipos de algoritmos de agrupamiento con el objetivo de incorporar visualizaciones complementarias de los datos a partir de las características distintivas de cada algoritmo considerado.

4.1. Análisis Estadístico del Dominio

Inicialmente se proporcionan los resultados estadísticos generales (tabla 1), para todas las variables involucradas en el desarrollo, con el fin de facilitar la lectura posterior de los resultados obtenidos.

Tabla 1. Resultados del procesamiento estadístico de las Variables usadas para la descripción del problema

Denominación de la variable	Naturaleza de la variable	Media	DesvSt
'Aprobó Asignatura en cursada'	Booleana	0,04	0,19
'Ritmo Inicial'	Ordinal discretizada	1,78	1,45
'Cumple Plan'	Ordinal discretizada	0,29	0,91
'Demora en Cursarla'	Booleana	1,36	0,75
'Beca'	Booleana	0,02	0,15
'Trabaja'	Booleana	0,61	0,49
'Familia'	Booleana	0,58	0,49
'Sexo'	booleana	0,43	0,50
'Madre últimos Estudios'	Ordinal discretizada	2,17	0,86
'Padre últimos Estudios'	Ordinal discretizada	1,96	0,84
'Argentino'	Booleana	1	0,06
'Córdoba'	Booleana	0,81	0,39
'Capital Córdoba'	booleana	0,50	0,50

En la tabla 1, se presentan la media, en adelante Media del Universo (MdU), y el

desvió estándar para cada variable seleccionada. Se observa que, del total de la población estudiantil y atendiendo a las escalas de las variables expresadas en la sección 3.1, aproximadamente: cursaron la mayoría de las materias en el primer cuatrimestre de carrera sin embargo cuando cursan la asignatura (tercer año para el plan nuevo vigente desde 2009 y quinto año para el plan viejo) han acumulado una demora en su plan de carrera de dos años como mínimo, el 2% accede a beca, el 61% trabaja, son casi inexistente los extranjeros, el 81% es de la provincia de Córdoba y el 50% de Córdoba Capital, el 58% costea sus estudios con aportes de la familia, el 57 % es de sexo femenino, el nivel de estudios de la madre es levemente superior al del padre, ambos cercanos a estudios secundarios completos o universitarios incompletos.

El experto del dominio percibe que los estudiantes no aprueban inmediatamente al finalizar la cursada (4%) porque la asignatura les resulta 'ajena' a sus estudios, no posee correlativas superiores y además, es muy extensa. Su sugerencia consiste en reprocesar los datos tomando un año como mínimo para el plazo de aprobación.

4.2. Respuesta académica en relación con variables socioeconómicas y de procedencia geográfica

En este apartado se describen las reglas de pertenencia a los grupos identificadas para cada uno de los algoritmos utilizados con el fin de caracterizar los estudiantes por sus similitudes. Además se proporciona una descripción de la composición del grupo, para aquellos casos en los que los hallazgos son considerados significativos por el experto del dominio ya sea porque confirman sus percepciones, porque las refutan o por constituir un aporte novedoso. Por último, se describen las percepciones del experto del dominio.

El algoritmo KMeans identifica cinco grupos. Los clasifica eficientemente, e invita a profundizar en su composición. Las variables

que participan en la caracterización son: 'Beca' (si/no), 'Aprobó en cursada' (si/no), 'Córdoba' (si/no), 'Familia' (si/no), 'Trabaja' (si/no), 'Ritmo Inicial' (de 0 a 3) y 'Cumple Plan' (de 0 a 4).

A continuación, para cada grupo se declara la regla de pertenencia y se proporciona una descripción de su composición en términos de las trece variables usadas en el estudio.

K1 (2120 individuos): Proceden de Córdoba, costean sus estudios con aporte familiar, no poseen beca de ayuda económica y no aprobaron la asignatura al finalizar la cursada.

Iniciaron sus estudios con un buen ritmo en la carrera, aunque levemente inferior a los que aprobaron la asignatura (K4) y superior a los becarios (K5), el nivel de estudios de ambos padres es similar a la MdU, el 53% son de Capital, el 33% trabaja, demoraron un poco menos que la MdU en cursar la materia, actualmente cumplen el plan de carreras, aunque muy lentos, algo mejor que la MdU. Según las percepciones del experto: hay una correlación directa entre nivel económico, mayor disponibilidad de tiempo porque no trabajan y mejor integración con el medio, son cordobeses.

K2 (2200 individuos): Proceden de la provincia de Córdoba, no aprobaron la asignatura en la cursada y no poseen beca. Todos trabajan, el 71% proceden de Capital, sólo el 10% recibe aporte económico familiar, éstos también trabajan. Todos iniciaron su carrera con el más bajo rendimiento académico. El nivel de estudios del padre y de la madre es bajo, aunque ligeramente superior al de los beneficiarios de beca (K5). Son actualmente los más demorados en su plan de carrera. Según las percepciones del experto, los resultados son muy consistentes y hay una correlación directa entre nivel económico, nivel de estudios formales de sus padres y rendimiento académico

K3 (1018 individuos): Proceden de otras provincias, no poseen beca y no aprobaron la asignatura al finalizar la cursada. Ambos padres poseen un alto nivel de estudios, aunque inferior al grupo K4, el 74% recibe

apoyo económico de la familia, la mitad son mujeres, y el 48% trabaja para costear sus estudios. Este grupo incluye a todos los extranjeros. Comenzaron con un ritmo inicial levemente inferior a la MdU y lo sostienen actualmente en el cumplimiento del plan de carrera y demoran más que la MdU en cursar esta asignatura. Según la percepción del experto, son estudiantes de nivel socioeconómico medio alto que no han logrado adecuarse al ritmo de estudios universitario, que se ven debilitados en un medio extraño y las características 'impersonales de la masividad'.

K4 (212 individuos): Aprobaron la asignatura al finalizar la cursada. Iniciaron la carrera cursando prácticamente todas las materias correspondientes al plan siendo el grupo de mayor 'Ritmo Inicial', cursan esta asignatura en el cuatrimestre que corresponde, sin embargo llevan muy lento su plan de carrera, aunque levemente mejor en relación a la MdU (0,75 vs 0,29). El nivel de estudios de ambos padres es el más alto de la población, el 84 % recibe apoyo económico de su familia, el 34% trabaja, ninguno posee beca, el 79% es de Córdoba y el 42% de Capital, el 54% son mujeres. La percepción del experto en relación a la característica que distingue a este grupo: 'los que aprobaron al finalizar la cursada' tienen relación con aspectos socio-económicos que podría verse materializado en el alto nivel de estudios alcanzado por ambos padres.

K5 (125 individuos): Poseen becas y no aprobaron al finalizar la cursada. En su gran mayoría son mujeres (86%), el nivel de estudios de ambos padres es el más bajo del universo, a lo sumo estudios incompletos de nivel secundario, el 94% son de Córdoba, reciben ayuda económica familiar, muy pocos trabajan (20%), su ritmo inicial de carrera fue decayendo con el tiempo y demoran en cursar la asignatura. La percepción del experto es que son estudiantes de nivel socio-económico más bajo, materializado en el nivel de estudios de ambos padres, que reciben un importante apoyo de sus familias, pero que no es suficiente. En una primera mirada,

sorprende que sean mujeres en su gran mayoría, tal vez asociado a las condiciones del mercado laboral. Que en casi su totalidad sean de Córdoba, tal vez explicaría que gente de otras provincias en las mismas condiciones, prefiere elegir otras opciones menos adversas, con costos de transporte más accesibles, inmediatez en materia emotiva y sentimental, etc. No sorprende que el rendimiento académico sea bajo, en su gran mayoría los estudiantes no alcanzan el ritmo académico propuesto por el plan de carrera.

Los resultados de la aplicación de HAC clasifican al universo de estudiantes en ocho conjuntos (C1 a C8). Las principales variables que participan en la caracterización de los grupos son: 'Beca' (si/no), 'Aprobó en cursada' (si/no), 'Córdoba' (si/no), 'Trabaja' (si/no), 'Ritmo Inicial' (de 0 a 3) y 'Cumple Plan' (de 0 a 4), los atributos que difieren del anterior son los relativos a la forma en que el estudiante costea su carrera. No se muestran resultados ya que no agregan nuevos patrones sobre el comportamiento poblacional.

El análisis de los datos utilizando el algoritmo SOM, permite caracterizar seis grupos. No se encontraron hallazgos novedosos que contribuyan significativamente al descubrimiento de nuevos patrones de comportamiento.

Se puede observar que en las reglas para la clasificación de las poblaciones identificadas, los atributos que tienen mayor presencia son: 'Aprobó Asignatura al finalizar la cursada', 'Beca', 'Trabaja', 'Córdoba', 'Cumple Plan' y 'Ritmo Inicial'. Otros atributos aparecen raramente, como 'nivel de estudios de la madre' y 'Familia'. Son descartados: 'Sexo', 'nivel de estudios del padre', 'Argentino' y 'Córdoba Capital' por todos los algoritmos en el proceso de descubrimiento de grupos a partir de sus similitudes. Este hallazgo sugiere la necesidad de indagar acerca de cuáles son los atributos que se interrelacionan más significativamente en la definición de estos grupos.

5. Conclusiones

Las interpretaciones que se realizan en esta presentación responden al objetivo de una primera caracterización de la población en estudio, a instancias de obtener los primeros resultados que dan luz a los procesos decisionales de las políticas públicas destinadas a este universo de estudiantes. Por otra parte, ellas están acotadas y en cierta medida, determinadas, por las profundidades de los árboles de inducción.

Se observa que es necesario contar con conocimiento a priori del experto del dominio y de sus percepciones en las instancias de preparación de la base de datos y selección de variables como a la hora de leer los resultados e interpretarlos para construir conclusiones que contribuyan a mejorar el conocimiento que se tiene acerca de los estudiantes de Educación Superior en contextos de masividad. En este sentido, durante el desarrollo del trabajo se sostuvo el diálogo en forma permanente entre el equipo de Ingeniería de Explotación de Información y los Expertos del Dominio, co-autores en esta comunicación.

Todos los algoritmos contribuyeron a dar respuesta a la primera pregunta específica de investigación relacionada con la cuestión de similitudes, quedando acreditado que la metodología utilizada puede seguir siendo utilizada en la profundización y extensión de las indagaciones en curso.

Los resultados muestran que: i) Los estudiantes extranjeros fueron incorporados en el cluster de alumnos procedentes de otras provincias, es relevante destacar que la cantidad de registros existentes son escasos para realizar un análisis detallado sobre los mismos, ii) que en esta carrera son muy escasas las becas de ayuda económica y no es posible evaluar su impacto en el rendimiento académico a diferencia de lo que fue posible hacer en el primer caso en estudio, de Ingeniería, ya mencionado, iii) el nivel de estudios de la madre es levemente superior al del padre en todos los casos y parece ser un factor no significativo en el rendimiento académico de los estudiantes, iv) es relevante el porcentaje de estudiantes que trabaja y su

impacto en el rendimiento académico, v) hay un grupo de estudiantes que, siendo de Córdoba, costea sus estudios con su trabajo y no recibe aportes de su familia, vi) se aprecia diferencias relacionadas con el género sólo en la obtención de becas de ayuda económica, vii) la única variable de procedencia que está presente en la caracterización de los grupos es 'Córdoba', siendo significativa en más de una ocasión para explicar comportamientos como pudo apreciarse en los resultados y viii) el plan de carreras es de muy difícil cumplimiento.

En apretada síntesis, para todos los grupos descubiertos, el nivel de estudios de la madre es levemente superior al del padre, el género de los estudiantes no parece tener mayor relevancia en su desempeño, a los estudiantes que trabajan, importante cantidad, se les dificulta más sostener el plan de carrera, como así también a los de nivel económico bajo y a los que no proceden de Córdoba.

Las líneas de trabajo sugeridas se relacionan con la necesidad de identificar los atributos de índole socioeconómicos que mayor impactan en el rendimiento académico de los estudiantes e indagar sobre otros aspectos relativos a sus formas de vida.

6. Financiamiento

Las investigaciones que se reportan en esta comunicación han sido financiadas parcialmente por el proyecto SECyT 05/M257 de la Universidad Nacional de Córdoba; por el proyecto 33A205 de la Universidad Nacional de Lanús y por beca PROMINF-UNLa-2015-2017 Secretaria de Políticas Universitarias, Ministerio de Educación de Argentina.

7. Referencias

Britos, P., Hossian, A., García Martínez, R. y Sierra, E. 2005. *Minería de Datos Basada en Sistemas Inteligentes*. 876 páginas. Editorial Nueva Librería. ISBN 987-1104-30-8.

Curtis, B., Kellner, M., Over, J. 1992. *Process Modelling. Communications of the ACM*, 35(9): 75-90.

Díaz, L., Martins, S., García-Martínez, R. 2015. *Descubrimiento de Patrones Socio-económicos de Población Estudiantil de Carreras de Ingeniería Basado en Tecnologías de Explotación de Información*. Proceedings X Congreso de Tecnología en Educación y Educación en Tecnología. Pág. 306-315. ISBN 978-950-656-154-3.

Ferreira, J., Takai, O., Pu, C. 2005. *Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services*. Proceedings Seventh IEEE International Conference on E-Commerce Technology. Pág. 471-474.

García-Martínez, R., Britos, P., Martins, S., Baldizzoni, E. 2015. *Ingeniería de Proyectos de Explotación de Información*. Nueva Librería. ISBN 987-1871-34-1.

García-Martínez, R., Britos, P., Rodríguez, D. 2013. *Information Mining Processes Based on Intelligent Systems*. Lecture Notes on Artificial Intelligence, 7906: 402-410. ISSN 0302-9743.

Juarros, M. 2006. *¿Educación superior como derecho o como privilegio?: Las políticas de admisión a la universidad en el contexto de los países de la región*. Andamios, 3(5): 69-90. ISSN 1870-0063.

Kohonen, T. 1982. *Self-Organized Formation of Topologically Correct Feature Maps*. Biological Cybernetics. Vol 43.

Kuna, H., García Martínez, R., Villatoro, F. 2010. *Pattern Discovery in University Students Desertion Based on Data Mining*. Advances and Applications in Statistical Sc. J., 2(2): 275-286. ISSN 0974-6811.

MacQueen, J. 1967. *Some methods for classification and analysis of multivariate observations*. 5th Berkeley Symposium on mathematics, Statistics and Probability, 1, S. 281-298.

Martins, S. 2014. *Derivación del Proceso de Explotación de Información Desde el Modelado del Negocio*. Revista

9 y 10 de junio de 2016

Latinoamericana de Ingeniería de Software, 2(1): 53-76. ISSN 2314-2642.

Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*. San Mateo: Morgan Kaufmann.