

# Utilización de minería de datos en el análisis de ejercicios de una evaluación de Álgebra Lineal.

Hernán C. Ahumada<sup>1</sup>, Carlos G. Herrera<sup>1</sup>

(1) Departamento de Formación Básica, Facultad de Tecnología y Ciencias Aplicadas, Universidad Nacional de Catamarca.  
hcahumada@tecno.unca.edu.ar, cgherrera@tecno.unca.edu.ar

Fecha de recepción del trabajo: 05/10/2015

Fecha de aceptación del trabajo: 10/04/2016

**RESUMEN:** En el trabajo se analizan los resultados del segundo parcial de la asignatura Álgebra, con el objetivo de conocer el grado de incidencia de los ejercicios en el hecho de que los alumnos aprueben o desapruében el parcial. El análisis se realiza sobre cuatro ejercicios sobre un total de diez, que corresponden a diferentes subtemas del tema Espacios Vectoriales, correspondientes a los contenidos de Álgebra Lineal que se dictan en el primer año de las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca, considerando asimismo las habilidades o procedimientos generales matemáticos necesarios para su resolución. Se aplica la técnica de reglas de asociación para identificar las correspondencias más significativas entre el nivel de resolución de los ejercicios y la aprobación o no del parcial. Se calculan tres métricas para las reglas obtenidas para ponderar la relevancia de cada una de ellas y se establecen las conclusiones correspondientes en base a los resultados más destacados.

**PALABRAS CLAVES:** Álgebra Lineal, Minería de Datos, Reglas de Asociación

*USE OF DATA MINING IN THE ANALYSIS OF EXERCISES OF A MID-TERM EXAMINATION OF LINEAR ALGEBRA.*

**ABSTRACT:** Results of the second mid-term examination of Linear Algebra were analyzed in order to determine the degree of incidence of some exercises on the results of the whole examination. The analysis was performed on four exercises of a total of ten, belonging to different topics of the linear algebra course, which are part of the curricula in the first year of Engineering at the Faculty of Technology and Applied Sciences, National University of Catamarca. General skills or mathematical procedures were also considered for their resolution. The technique of association rules was applied to identify the most significant correlation between the level of resolution of the exercises and the passing or failure of the mid-term exam. Three metrics were calculated for the rules obtained to weigh their relevance. Conclusions are set based on the most outstanding results.

**KEYWORDS:** Linear Algebra, Data Mining, Association Rules

## 1 INTRODUCCIÓN

Se presentan resultados preliminares que corresponden a un proyecto de investigación cuyo objetivo es utilizar técnicas Minería de Datos para determinar patrones de comportamiento académico en alumnos de primer año de las carreras de Ingeniería y que permitan establecer medidas didácticas e institucionales a los efectos de diseñar un sistema de alerta temprana de deserción de alumnos. También se utilizará la información disponible en la historia académica como así también la que se logre al evaluar los procedimientos o habilidades generales matemáticos en cada alumno que cursa las asignaturas de primer año de las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

La minería de datos educativos (Educational Data Mining -EDM) (Romero, 2010) es una disciplina relacionada con el desarrollo de métodos para extraer información útil a partir de los datos que se generan en los entornos educativos, y utilizarla para mejorar dicho

entorno. La información así obtenida se convierte en el insumo indispensable para la toma de decisiones.

Publicaciones previas han demostrado que la minería de datos se puede utilizar para detectar alumnos en riesgo de deserción. Luan (2002) aplicó exitosamente técnicas de minería de datos para predecir qué grupos de alumnos podrían abandonar los estudios. En un trabajo relacionado, Lin (2012) usó técnicas de minería de datos para optimizar los esfuerzos para retener estudiantes. Para ello se generaron modelos predictivos basados en datos de los estudiantes, que permitieron detectar con gran precisión aquellos alumnos que debían recibir ayuda de los sistemas de retención de alumnos. Investigadores de la Universidad Estatal de Bowie (Chacón, 2012) desarrollaron un sistema basado en minería de datos que permite a dicha Institución identificar y atender a estudiantes en riesgo de abandono de estudios. Madhyastha, (2009) utiliza minería de datos para analizar la similitud de conceptos en una evaluación de tipo de opciones múltiples. Cetintas (2014) analiza las frases relevantes e irrelevantes de

problemas de matemáticas utilizando técnicas probabilísticas.

En este trabajo se plantea la utilización de técnicas de minería de datos para detectar y analizar patrones frecuentes en una evaluación parcial de contenidos de Álgebra Lineal en alumnos de primer año de las carreras de Ingeniería. Se consideran, del total de diez ejercicios de la evaluación, los cuatro que corresponden a distintos subtemas del concepto de Espacio Vectorial. Las situaciones planteadas, requieren además de habilidades matemáticas de interpretar, identificar objetos matemáticos o graficar de acuerdo a las definiciones de habilidades o procedimientos generales propuestos por Delgado Rubí (1997). Se busca identificar y cuantificar la incidencia que tiene la forma de responder cada ejercicio en la aprobación o no del parcial.

## 2 MARCO TEÓRICO

Reglas de Asociación es una bien conocida técnica de minería de datos, que cobró popularidad desde que Agrawal presentara un algoritmo eficiente para generarlas automáticamente a partir de procesar bases de datos (Agrawal, 1993). Una regla de asociación tiene dos partes, antecedente y consecuente. El antecedente está formado por un atributo o conjunción de atributos encontrados en los datos. El consecuente representa a otro atributo encontrado en combinación con el antecedente. Formalmente, una regla de asociación es una implicación lógica de la forma  $X \Rightarrow Y$ , donde  $X$  es la premisa o antecedente que representa la condición o conjunción de condiciones que deben ser ciertas para que la conclusión o consecuente  $Y$  se cumpla. Por lo tanto, una regla de la forma  $X \Rightarrow Y$ , puede ser interpretada a modo general de la siguiente forma: si sucede  $X$ , entonces sucede  $Y$ .

Las reglas de asociación ayudan a descubrir combinaciones de pares atributo-valor que ocurren con frecuencia en un conjunto de datos e inferir relaciones a partir de ellas que resultan bastante sencillas de interpretar. Entre los algoritmos para encontrar reglas de asociación se destacan los denominados Apriori (Agrawal, 1994) y Eclat (Zaki, 1997). Dichos algoritmos utilizan estrategias bastante diferentes. Apriori, cuenta las frecuencias de las transacciones de manera incremental con un criterio de primero en ancho.

En vez de contar frecuencias, Eclat emplea equivalencia de clases, búsqueda primero en profundidad y conjunto intersección. Ambos algoritmos se usan para identificar eventos frecuentes, con la ventaja de que los resultados obtenidos con Apriori pueden usarse para generar reglas de asociación.

Para cuantificar el grado de validez de una regla de asociación, se utilizan principalmente dos medidas denominadas soporte y confianza (Hasler, 2007). Soporte se puede calcular para cada evento en particular o para una regla.

El soporte de un evento es la frecuencia relativa de ocurrencia de tal evento en la muestra de datos. Puede ser vista como la probabilidad marginal (o no condicional) de que ocurra el evento. En símbolos:

$$\text{sop}(X) = P(X) \quad (1)$$

El soporte de una regla de asociación, se entiende como la proporción de veces que el antecedente  $X$  y el consecuente  $Y$  se presentan de manera simultánea en los datos. Es por lo tanto, la probabilidad conjunta de ambos eventos. En símbolos:

$$\text{sop}(X \Rightarrow Y) = P(X \cap Y) \quad (2)$$

El soporte de una regla se calcula como el cociente del número de casos a los cuales la regla se aplica y predice correctamente entre el número total de casos.

En tanto que, Confianza es la proporción de eventos que contienen a  $X$  en los cuales también se observa la ocurrencia de  $Y$ , puede interpretarse como la probabilidad del consecuente  $Y$  dado que ocurre  $X$ . En símbolos:

$$\text{conf}(X \Rightarrow Y) = \frac{\text{sop}(X \Rightarrow Y)}{\text{sop}(X)} \quad (3)$$

$$\text{conf}(X \Rightarrow Y) = \frac{P(X \cap Y)}{P(X)} \quad (4)$$

$$\text{conf}(X \Rightarrow Y) = P(Y/X) \quad (5)$$

La confianza se calcula como el cociente entre el número de casos a los cuales la regla se aplica y predice correctamente entre el número de instancias a las cuales la regla se aplica. Esto es, la confianza da la probabilidad de que la regla clasifique correctamente una instancia a la cual se aplica (Carreño, 2006).

Dado que un algoritmo de reglas de asociación tiende a generar una gran cantidad de casos es usual establecer valores de umbral mínimos tanto para el soporte como para la confianza. Se debe a que una regla con bajo soporte significa que la misma se cumple en pocos casos, en tanto que si la confianza de la regla es baja implica que la ocurrencia del antecedente influye débilmente sobre el consecuente. Entonces, si el soporte de una regla mide la frecuencia de ella y la confianza representa la fortaleza de la regla, es de interés encontrar reglas frecuentes (soporte alto) y fuertes (confianza alta).

$$\text{lift}(X \Rightarrow Y) = \frac{\text{sop}(X \Rightarrow Y)}{\text{sop}(X) * \text{sop}(Y)} \quad (6)$$

$$\text{lift}(X \Rightarrow Y) = \frac{\text{conf}(X \Rightarrow Y)}{\text{sop}(Y)} \quad (7)$$

El indicador lift de una regla de asociación permite medir la ganancia de información que resulta la probabilidad condicional (confianza de la regla) con respecto a la probabilidad no condicional (soporte del consecuente).

Valores de lift próximos a 1 pueden interpretarse como que la regla establecida prácticamente no reduce la incertidumbre ya que numerador y denominador son valores cercanos.

Cuanto mayor a 1 sea el lift, se deberá al hecho de que la probabilidad condicional del consecuente Y (confianza de la regla) es mayor que la probabilidad no condicional (soporte del consecuente Y). Podemos interpretar esta situación como la reducción en la incertidumbre por el aporte que supone la ocurrencia del antecedente X de la regla. Cuanto mayor sea el lift, diremos que más fuertemente relacionados están el antecedente y el consecuente.

### 3 METODOLOGÍA

Se trata de una investigación de tipo cuantitativa, sobre una muestra de 65 alumnos de la cátedra Álgebra correspondiente a las carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca.

Se analizan la incidencia de cuatro ejercicios, denominados E4, E5, E6, E7 sobre un total de diez en el resultado final de una evaluación parcial de contenidos de Álgebra Lineal correspondiente a primer año de carreras de Ingeniería de la Facultad de Tecnología y Ciencias Aplicadas de la Universidad Nacional de Catamarca utilizando la técnica de Minería de Datos (Data Mining) denominada Reglas de Asociación.

La evaluación corresponde al tema Espacios Vectoriales que corresponde a un curso de Álgebra Lineal y las situaciones seleccionadas para el análisis son las siguientes:

E4: Indique si el conjunto solución del sistema de ecuaciones

$$\begin{aligned} x + y - z &= 0 \\ 2x - y + z &= 0 \end{aligned}$$

Constituye un subespacio de  $(\mathbb{R}^3, +, \mathbb{R}, *)$ . Justifique su respuesta. Interprete geoméricamente el sistema de ecuaciones y su conjunto solución.

Objetivo: el alumno deberá representar las relaciones existentes entre los conceptos matemáticos de sistemas de ecuaciones lineales homogéneos y el espacio

vectorial respectivo y explicar su interpretación geométrica correspondiente.

E5: Indique, justificando sus respuestas, si los siguientes conjuntos de vectores de  $\mathbb{R}^3$  son linealmente independientes o linealmente dependientes.

- $A = \{(1, -1, 2); (0, 1, 1)\}$
- $B = \{(1, 0, 1); (1, 1, 0); (2, 0, 2)\}$
- $C = \{(1, 0, 1); (1, 1, 0); (0, 1, 1)\}$
- $D = \{(1, 0, 1); (1, 1, 0); (0, 1, 1); (1, 1, 1)\}$
- $E = \{(1, 0, 1); (0, 0, 0)\}$

Objetivo: En este ejercicio el alumno pondrá de manifiesto los procedimientos generales matemáticos de Interpretar y de Demostrar, si un conjunto de vectores es Linealmente Independiente o Linealmente Dependiente.

E6: Sea el espacio vectorial  $W = \{(x, y, z) \in \mathbb{R}^3 / x + y + z = 0\}$

- Indique si el conjunto de vectores  $A = \{(1, 0, -1); (0, -1, 1); (2, -1, -1)\}$  genera W.
- Indique, justificando su respuesta si los vectores del conjunto A son Linealmente Independientes o Linealmente Dependientes.

Objetivo: El primer apartado del ejercicio requiere el procedimiento general matemático de Demostrar que un conjunto de vectores es un Conjunto Generador de un Espacio Vectorial. La segunda parte de la situación planteada requiere de la Habilidad de Interpretar, la Dependencia Lineal de los vectores del conjunto A.

E7: Sea el espacio vectorial  $W = \{(x, y, z) \in \mathbb{R}^3 / x + y + z = 0\}$

- Indique una base y dimensión de W. Interprete geoméricamente W.
- Es posible generar W con solo dos vectores del conjunto  $A = \{(1, 0, -1); (0, -1, 1); (2, -1, -1)\}$ . Justifique su respuesta.

Objetivo: en esta situación se evaluaron las habilidades matemáticas de Interpretar y graficar, sobre la base de los conceptos de Base y Dimensión de un espacio vectorial.

En el trabajo se analiza la incidencia de cada uno de estos ejercicios en el resultado final de la evaluación teniendo en cuenta que los ejercicios corresponden a diferentes subtemas de un tema general como el de espacios vectoriales. Estos subtemas son los conceptos de subespacio en E4, dependencia o independencia lineal en E5, conjunto generador y dependencia lineal en el ejercicio E6, base y dimensión en el E7. También se tiene en cuenta, independiente de los subtemas antes mencionados, que los ejercicios E4 y E5 son fundamentalmente conceptuales, es decir que para su realización el alumno debe tener claro los conceptos teóricos de sistemas de ecuaciones, espacios vectoriales,

dependencia e independencia lineal, mientras que en los ejercicios E6 y E6 además de concepto deben tener claro la interpretación geométrica del espacio vectorial W.

Para generar las reglas de asociación se utiliza el lenguaje de programación R (Team, 2014), que brinda las herramientas apropiadas para aplicar técnicas de minería de datos. Se aplicó el algoritmo Apriori implementado en la librería arules (Hahsler, 2009). Las gráficas de las reglas, obtenidas con arules, se generan mediante la librería arulesViz. (Hahsler, 2011).

#### 4 RESULTADOS

En Tabla 1 se presentan los resultados que se obtienen utilizando Reglas de Asociación, considerando sólo aquellas reglas de asociación que tengan al menos un soporte del 15%. Se realiza una descripción de un ejemplo para la regla (E7=NR)⇒ (Aprueba = NO).

Tabla 1: Resultados del análisis de datos utilizando la técnica de Reglas de Asociación.

Regla N°	Antecedente	Consecuente	Soporte	Confianza	Lift
1	{}	{Aprueba=SI}	0,600	0,600	1,000
2	{}	{Aprueba=NO}	0,400	0,400	1,000
3	{E7=NR}	{Aprueba=NO}	0,338	0,688	1,720
4	{E7=BIEN}	{Aprueba=SI}	0,292	1,000	1,000
5	{E4=BIEN}	{Aprueba=SI}	0,262	0,895	1,491
6	{E6=BIEN}	{Aprueba=SI}	0,246	0,800	1,333
7	{E5=REG}	{Aprueba=SI}	0,246	0,667	1,111
8	{E6=NR, E7=NR}	{Aprueba=NO}	0,231	0,750	1,875
9	{E6=NR}	{Aprueba=NO}	0,231	0,652	1,630
10	{E5=BIEN}	{Aprueba=SI}	0,200	1,000	1,667
11	{E6=MAL}	{Aprueba=SI}	0,200	0,684	1,140
12	{E4=BIEN, E6=BIEN}	{Aprueba=SI}	0,154	1,000	1,667
13	{E4=MAL}	{Aprueba=NO}	0,154	0,588	1,471

De la base de datos surge que 32 de los 65 alumnos no respondieron el ejercicio 7, por lo tanto el soporte del evento E7=NR resulta:

$$sop(E7 = NR) = 32/65$$

Para calcular el soporte de la regla (E7 = NR) ⇒ (Aprueba = NO) se contabiliza la cantidad de veces que se dan simultáneamente ambos eventos, es decir alumnos que no respondieron el ejercicio 7 y que no aprobaron el parcial. Esa cantidad se divide en el total de casos. El cociente representa la probabilidad conjunta.

$$sop((E7 = NR) \Rightarrow (Aprueba = NO)) = 22/65$$

$$sop((E7 = NR) \Rightarrow (Aprueba = NO)) = 0.3384615$$

Significa que la regla establecida se verifica en el 33.85% de los casos.

El cálculo de la confianza de la regla se plantea de la siguiente manera:

$$conf(X \Rightarrow Y) = \frac{sop(X \Rightarrow Y)}{sop(X)}$$

$$conf(X \Rightarrow Y) = \frac{22/65}{32/65}$$

$$conf(X \Rightarrow Y) = 0.6875$$

Dado que la confianza de una regla se interpreta como la probabilidad condicional del consecuente dado que ocurre el antecedente. El valor de confianza obtenido representa que existe una probabilidad de 0.6875 de que un alumno desaprobe el parcial si no responde el ejercicio 7.

El valor del parámetro lift está dado por:

$$lift(X \Rightarrow Y) = \frac{sop(X \Rightarrow Y)}{sop(X) * sop(Y)}$$

$$lift(X \Rightarrow Y) = \frac{conf(X \Rightarrow Y)}{sop(Y)}$$

$$lift(X \Rightarrow Y) = \frac{0.6875}{0.4}$$

$$lift(X \Rightarrow Y) = 1.71875$$

Por lo tanto, en base a los valores de soporte, confianza y lift de la regla (E7 = NR) ⇒ (Aprueba = NO), representa que la misma se verifica en el 33.85% de los casos, y que si el ejercicio 7 no es respondido existe una probabilidad de 0.6875 de que se desaprobe el parcial. Como el lift es mayor a uno (1.71875) quiere decir que existe dependencia entre el antecedente y el consecuente, esto es que si no se responde el ejercicio incrementa la probabilidad de desaprobación del parcial.

En la tabla 1 sólo se tienen dos reglas de asociación con el soporte mínimo establecido que cuentan con 2 ítems en el antecedente. Esto se debe a que es baja la proporción de alumnos que responden correctamente un par de ejercicios de los cuatro considerados para este estudio.

En la Figura 1 se pueden observar las reglas de mayor soporte en la condición de Aprobado o Desaprobado en la evaluación de Álgebra. En eje horizontal se representa la posición del ejercicio en el antecedente (1 o 2), mientras que rhs representa el consecuente de la regla. En eje vertical se representan los valores de antecedente y consecuente. El espesor de línea representa el soporte de la regla. De este gráfico

observando los ejes verticales que se conectan mediante las flechas y el grosor de las líneas, se deduce que las reglas de mayor soporte, considerando la condición de desaprobado, se producen cuando no responden los ejercicios E7 y E6 en ese orden. Es decir, la regla  $(E7 = NR) \Rightarrow (Aprueba = NO)$  y la regla  $(E6 = NR) \Rightarrow (Aprueba = NO)$  son las de mayor soporte en la condición de alumno desaprobado. Cuando se analiza la condición de alumnos aprobados las reglas de mayor soporte son  $(E7 = BIEN) \Rightarrow (Aprueba = SI)$  y  $(E4 = BIEN) \Rightarrow (Aprueba = SI)$ .

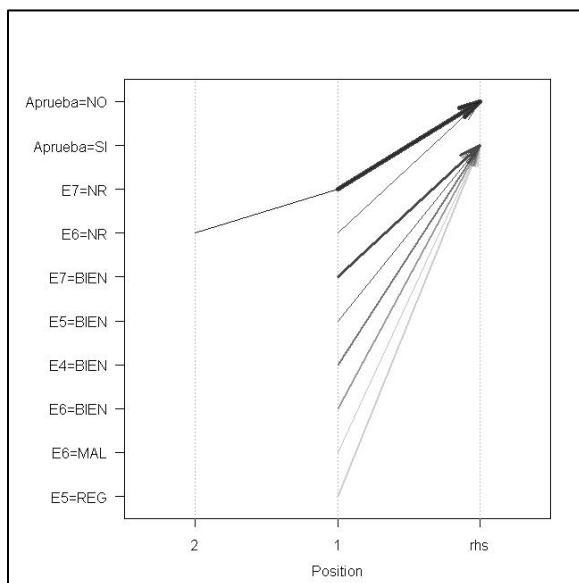


Fig. 1 Reglas de asociación representadas en ejes paralelos.

En Figuras 2 y 3 se presentan esquemáticamente las reglas con mayores valores de soporte y lift para la condición de Aprobado o Desaprobado respectivamente. El tamaño de cada círculo está asociado al soporte de la regla. La tonalidad de relleno depende del lift de la regla. En la Figura 2, se considera como consecuente la condición de Aprobado (APROBADO=SI). Las reglas con mayor soporte son las que tienen como antecedente  $(E7=BIEN)$ ,  $(E4=BIEN)$ ,  $(E6=BIEN)$ ,  $(E5=REG)$  siendo el coeficiente lift de esta última regla muy cercano a la unidad lo que indica la poca asociación de las dos proposiciones  $(E5=REG)$  y  $(APROBADO=SI)$ . De izquierda a derecha, la tonalidad de los círculos disminuye, indicado que tales reglas tienen valores de lift cercanos a 1, y que por lo tanto la relación establecida entre antecedente y consecuente puede deberse más bien a cuestiones del azar.

El círculo del extremo derecho de la Figura 2, representa solamente el soporte del consecuente (APROBADO=SI), y por lo tanto su tamaño sirve de

referencia para comparar con el soporte de las reglas que aparecen graficadas hacia la izquierda.

En Figura 3, se considera como consecuente la condición de Desaprobado (APROBADO=NO). Las reglas con mayor soporte tienen como antecedente  $(E7 = NR)$  y  $(E4 = NR)$  aunque el coeficiente lift de la primera es muy elevado (1.7185), lo cual es indicado por la tonalidad más oscura de la circunferencia. La proposición  $(E4=NR)$  tiene un coeficiente lift 1.3235, lo que indica que tiene menos asociación con la condición de (APROBADO=NO).

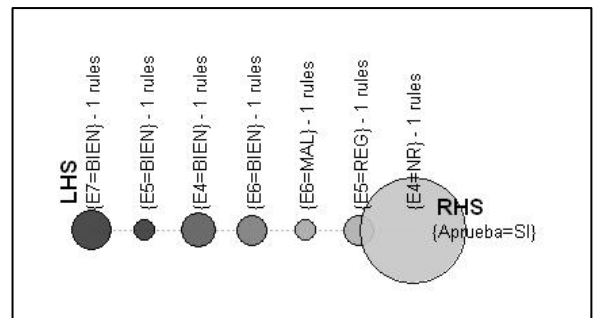


Fig. 2: Soporte y lift de Reglas donde el consecuente es Aprueba=SI.

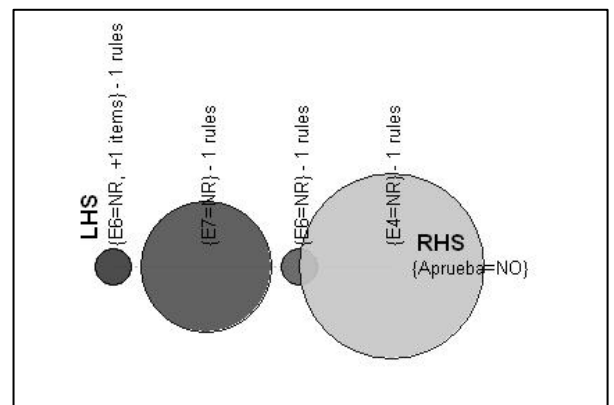


Fig. 3: Soporte y lift de Reglas donde el consecuente es Aprueba=NO.

### 5 CONCLUSIONES

La aplicación de la técnica de Reglas de Asociación permitió realizar un análisis novedoso de los resultados de un grupo de ejercicios de una evaluación parcial de la asignatura Álgebra. Se obtuvieron distintas reglas de asociación cuya relevancia y significancia fue cuantificada por indicadores de soporte, confianza y lift. Las reglas y las métricas calculadas fueron representadas gráficamente para facilitar su conformación y relevancia.

Del análisis de los resultados obtenidos se pueden determinar los ejercicios cuya forma de resolución influyen de manera más determinante en la aprobación o no del parcial.

Para el caso de los alumnos que aprobaron el parcial, se encontró la regla de asociación que establece que si el ejercicio 7 es correctamente resuelto se incrementa en mayor medida la probabilidad de aprobar el parcial.

Analizando los casos de quienes desaprobaron el parcial, las reglas de asociación generadas permiten observar que el motivo más influyente es no responder el ejercicio 7.

Desde el punto de vista cognoscitivo, tales conclusiones se pueden fundamentar por el hecho de que el ejercicio N° 7, que corresponde al subtema base de un espacio vectorial, involucra temas como sistema generador de un espacio vectorial y dependencia lineal, teniendo en cuenta que además requiere de habilidades o procedimientos generales matemáticos de interpretar y graficar.

El tipo de información obtenida mediante reglas de asociación puede ser utilizado para el diseño de ejercicios que favorezcan el desarrollo de competencias fundamentales para una mejor comprensión de los contenidos desarrollados.

## REFERENCIAS

- Agrawal, R., Imieliński, T., & Swami, A. Mining association rules between sets of items in large databases. In *ACM SIGMOD Record* (Vol. 22, No. 2, pp. 207-216). ACM. 1993
- Agrawal, R., & Srikant, R. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499). 1994.
- Carreno, E., & Leguizamón, G. Balance entre confianza, soporte y comprensibilidad en la evolución de reglas de clasificación. In *VIII Workshop de Investigadores en Ciencias de la Computación*. 2006
- Cetintas, S., Si, L.; Xin, Y.; Zhang, D.; Park, J.; Tzur, R.: "A joint probabilistic classification model of relevant and irrelevant sentences in mathematical word problems." *arXiv preprint arXiv:1411.5732*. 2014.
- Chacon, Fabio, Donald Spicer, and A. Valbuena. "Analytics in support of student retention and success." *Research Bulletin* 3. 2012.
- Delgado Rubí. Un sistema de habilidades generales para la enseñanza en la Matemática. La Habana, ISPJAE, 6. 1997.
- Hahsler, M., Grün, B., & Hornik, K. Introduction to arules—mining association rules and frequent item sets. *SIGKDD Explor.* 2007.
- Hahsler, M., Grün, B., Hornik, K., & Buchta, C. Introduction to arules—A computational environment for mining association rules and frequent item sets. The Comprehensive R Archive Network. 2009.
- Hahsler, M., & Chelluboina, S.. Visualizing association rules: Introduction to the R-extension package arulesViz. R project module, 223-238. 2011
- Lin, Tung Yen, Cercone, Nick. Rough sets and data mining: Analysis of imprecise data. *Springer Science & Business Media*, 2012.
- Luan, Jing. "Data mining and its applications in higher education." *New directions for institutional research* 2002.113. (17-36). 2002.
- Madhyastha, Tara, and Earl Hunt. "Mining diagnostic assessment data for concept similarity." *JEDM—Journal of Educational Data Mining* 1.1 : 72-91. 2009
- Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (Eds.). Handbook of educational data mining. CRC Press. 2010.
- Team, R. C.: The R project for statistical computing. R Foundation for Statistical Computing web-site. www. R-project.org. Accessed. 2014
- Zaki, M. J., Parthasarathy, S., Ogihara, M., & Li, W. New Algorithms for Fast Discovery of Association Rules. In *KDD* (Vol. 97, pp. 283-286). 1997.