UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Revista Vínculos
https://revistas.udistrital.edu.co/index.php/vinculos/

I+D Investigación y Desarrollo

Integrando conocimiento mediante reglas de asociación: caso de estudio

Knowledge integration using association rules: case of study

María Alejandra Malberti Riveros[1] Graciela Elida Beguerí[2] Raúl Oscar Klenzi[3]

Para citar este artículo: M.A. Malberti Riveros, G. E. Beguerí, R. O. Klenzi, "Integrando conocimiento mediante Reglas de Asociación: caso de estudio". *Revista Vínculos*, vol. 17, n°. 1, pp. 24-31 enero-junio. 2020. DOI: https://doi.org/10.14483/2322939X.15785

Recibido: 02-01-2020 / Aprobado: 14-02-2020.

Resumen

Mejorar el proceso de enseñanza-aprendizaje es uno de los objetivos que tienen las instituciones de educación superior. En este contexto, la interrelación entre las asignaturas del diseño curricular se torna imprescindible. Sin embargo, muchas veces no se realizan los esfuerzos necesarios para integrarlas efectivamente. En este trabajo se presenta una experiencia para una asignatura troncal, Inteligencia Artificial, la cual permite al alumno integrar conocimientos mediante actividades prácticas y lograr así la articulación con algunos contenidos académicos adquiridos en el transcurso de la carrera.

Palabras clave: educación superior, minería de datos, probabilidad, reglas de asociación.

Abstract

Higher Education institutions mainly aim at improving the teaching-learning process. In this context, the interrelationship among courses in the development of the Curricular Design becomes essential. However, in many cases not all the required efforts are made in order for subjects to be effectively integrated. This paper presents a learning experience designed for a core subject, Artificial Intelligence, with the purpose of allowing students to integrate knowledge through practical activities as well as achieving better articulation with previously acquired academic contents throughout their course of study.

Keywords: association rules, data mining, higher education, probability.

1 Magíster, Departamento e Instituto de Informática, Universidad Nacional de San Juan, Argentina. Correo electrónico: amalberti@gmail.com. ORCID: https://orcid.org/0000-0002-8332-4339
2 Magíster, Departamento e Instituto de Informática, Universidad Nacional de San Juan, Argentina. Correo electrónico: grabeda@gmail.com. ORCID: https://orcid.org/0000-0002-3629-8430
3 Magister, Departamento e Instituto de Informática, Universidad Nacional de San Juan, Argentina. Correo electrónico: rauloscarklenzi@gmail.com. ORCID: https://orcid.org/0000-0003-0061-6094

ISSN 1794-211X • e-ISSN 2322-939X • Vol 17, N° 1 (enero-junio 2020). pp. 24-31. Universidad Distrital Francisco José de Caldas-Facultad Tecnológica.

1. Introducción

En Argentina, la Ley Federal de Educación en su artículo 22 establece que las universidades tienen entre sus funciones la de "Desarrollar el conocimiento en el más alto nivel con sentido crítico, creativo e interdisciplinario, estimulando la permanente búsqueda de la verdad" [1].

En ese sentido, la interdisciplinariedad consiste en:

> "un trabajo común teniendo presente la interacción de las disciplinas científicas, de sus conceptos, directrices, de su metodología, de sus procedimientos, de sus datos y de la organización de la enseñanza y constituye, además, una condición didáctica y una exigencia para el cumplimiento del carácter científico de la enseñanza. Los conocimientos sin vinculación entre sí rompen la asimilación consciente de los conocimientos y habilidades. Lograr una adecuada relación entre las diferentes asignaturas que conforman un Plan de Estudio, influye en el consecuente incremento de la efectividad de la enseñanza tanto en términos cuantitativos como cualitativos" [2].

Por resolución del Ministerio de Educación, está dispuesto que "En el Plan de Estudios los contenidos deben integrarse horizontal y verticalmente. Asimismo, deben existir mecanismos para la integración de docentes en experiencias educacionales comunes" [3].

En el año 2012, y con el propósito expuesto, los autores presentan una primera iniciativa en pos de integrar áreas de conocimientos donde toman "como punto central el tema Minería de Datos, por ser éste un campo multidisciplinar que se ha nutrido de diferentes áreas conceptuales tales como, estadística, aprendizaje automático y bases de datos entre otras" [4]. En el trabajo se pone de manifiesto la importancia que tiene la vinculación entre conceptos, técnicas, herramientas y terminología.

Waldemiro Vélez Cardona resalta "la importancia de crear condiciones de aprendizaje que conduzcan —deliberadamente— a la integración de saberes" y la "vincula a la idea de interdependencia o interrelación de los diferentes elementos que constituyen un todo" [5]. También expresa lo siguiente: "Un error que a mi juicio se ha cometido en la universidad es asumir que la integración del conocimiento la hace el estudiante por su cuenta, sin que sea necesario hacer nada deliberado para propiciarle" [5].

La Unesco entiende que es "vital reconocer y complementar un pensamiento que separa con un pensamiento que reúne los conocimientos separados" [6] y que uno de los problemas al cual la educación se enfrenta es que "existe una inadecuación cada vez más amplia, profunda y grave por un lado entre nuestros saberes desunidos, divididos, compartimentados y por el otro, realidades o problemas cada vez más poli disciplinarios, transversales, multidimensionales, transnacionales, globales, planetarios" [6].

Las estrategias de ensayo o recirculación de la información que más utilizan los estudiantes son aquellas que permiten evocar los aprendizajes cuando los requieren. La mayoría de los alumnos encuentran relación entre los nuevos conocimientos con los ya adquiridos y reconocen que lo aprendido tiene relación con lo que se va a abordar y utilizan diferentes tácticas para integrar y relacionar la nueva información con los conocimientos que ya poseían [7].

Con el fin de responder al desafío de articular conceptos y conocimientos de asignaturas que los alumnos suelen verlas como independientes, se pueden incorporar actividades prácticas como las que se presentan, lo anterior a manera de experiencia educacional.

2. Propuesta

Vincular las asignaturas Probabilidad y Estadística, Estructuras de Datos y Algoritmos e Inteligencia Artificial de las carreras de grado del Departamento de Informática de la Facultad de Ciencias Exactas, Físicas y Naturales en la Universidad Nacional de

ISSN 1794-211X • e-ISSN 2322-939X • Vol 17, N° 1 (enero-junio 2020). pp. 24-31. Universidad Distrital Francisco José de Caldas-Facultad Tecnológica.

San Juan y promover la integración del conocimiento, en la última asignatura mencionada se incluyen temas de minería de datos, entre los cuales están las estrategias de reglas de asociación.

En este marco, se diseña la experiencia educacional y se detallan las dos alternativas para motivar a los alumnos: por un lado, se considera el cálculo de probabilidades; por otro, se exponen algunas métricas para la evaluación de reglas de asociación.

Se enuncian a continuación las definiciones involucradas. Formalmente, a partir de los modelos matemáticos propuestos para dirigir el problema de búsqueda de reglas de asociación [8], [9], [10] y [11], se considera:

$D=\{d_1,d_2,...,d_n\}$ un conjunto de ítems y $T=\{t_1,t_2,\ldots,t_n\}$ un conjunto de transacciones, donde cada transacción t_i es un conjunto de ítems tal que $t_i \subseteq D$ $1 \leq i \leq n$.

La implicación $X \Rightarrow Y$ es una regla de asociación donde $X \subset D$, $Y \subset D$ y $X \cap Y = \phi$. Los conjuntos X e Y son mutuamente excluyentes y $X \cup Y \subseteq t_i$, esto es, el conjunto de ítems formado por aquellos que corresponden al antecedente o al consecuente de la regla de asociación debe estar contenido o ser igual a alguna de las transacciones pertenecientes a T.

Para evaluar las reglas de asociación descubiertas en el conjunto de transacciones, se adoptan los factores soporte y confianza dados en [8] y [12]. Así, la regla $X \Rightarrow Y$ tiene soporte s en el conjunto de transacciones T, $0 \leq s \leq 1$, si el s% de las transacciones de T contienen tanto a X como a Y.

El soporte es definido sobre un conjunto de ítems y es usado como una medida de trascendencia o de importancia del mismo. Este puede ser considerado como la probabilidad de que las transacciones contengan un conjunto de ítems (1).

$$Soporte\ (X) = P(X) \quad (1)$$

Para el caso de las reglas de asociación, el conjunto está formado por los ítems del antecedente y del consecuente (2).

$$Soporte\ (X \Rightarrow Y) = soporte\ (X \cup Y) = P(X \cup Y) \quad (2)$$

La regla $X \Rightarrow Y$ se mantiene en el conjunto de transacciones T, con factor de confianza, $0 \leq c \leq 1$ si el c% de las transacciones de T que satisfacen X también satisfacen Y, es decir, el porcentaje de transacciones que contienen ambos X e Y, esto es $X \cup Y$, respecto al número total de transacciones que contienen X (3).

$$confianza\ (X \Rightarrow Y) = \frac{soporte\ (X \cup Y)}{soporte\ (X)} = \frac{P(X \cup Y)}{P(X)} \quad (3)$$

La confianza es definida como la probabilidad de que las transacciones que contienen el antecedente de la regla también contengan el consecuente, esto es, la probabilidad de que ocurra Y dado que ya ocurrió X. La confianza puede ser considerada, entonces, como un estimador de la probabilidad condicional $P(Y/X)$. Se puede comprobar que el valor de la confianza para la regla $X \Rightarrow Y$ es distinto al valor de la confianza para la regla $Y \Rightarrow X$. La confianza no es simétrica, simbólicamente en (4).

$$confianza\ (X \Rightarrow Y) \neq confianza\ (Y \Rightarrow X) \quad (4)$$

Existen otras medidas para evaluar la importancia de las reglas generadas. Entre estas medidas se considera una muy popular, denominada *lift*, la cual es un tipo de medida de independencia estadística y se define en (5).

$$lift\ (X \Rightarrow Y) = lift\ (Y \Rightarrow X) = \frac{P(Y/X)}{P(Y)} \quad (5)$$

Este factor establece una relación entre la ocurrencia simultánea de X e Y cuando los conjuntos de ítems que conforman el antecedente y el consecuente de la regla son estadísticamente independientes. Reglas de asociación con valores de *lift* menores a 1 no deberían ser tenidas en cuenta para la toma de decisiones [13].

ISSN 1794-211X • e-ISSN 2322-939X • Vol 17, N° 1 (enero-junio 2020). pp. 24-31. Universidad Distrital Francisco José de Caldas-Facultad Tecnológica.

De lo expuesto, se desprende, por un lado, la importancia de la teoría de conjuntos, ya que sienta la base de la formulación del tema y, por el otro, la evaluación de las reglas de asociación descubiertas que recurren a los factores soporte y confianza, los cuales están definidos en base a la teoría de probabilidad.

2.1. Descripción

Como se mencionó, la experiencia se implanta en el marco de la asignatura Inteligencia Artificial de las carreras de grado del Departamento de Informática de la Universidad Nacional de San Juan. Se consideran dos ejemplos, el primero se basa en el conjunto de datos de la canasta de mercado (*market-basket*) [8] y el segundo con los datos del clima (weather. arff) provistos en el *software* para aprendizaje automático Weka, con licencia GPL [9].

Es importante tener presente que el propósito de las reglas de asociación es hallar reglas que permitan predecir la ocurrencia de un ítem basado en la ocurrencia de otros ítems en la transacción, lo anterior a partir del análisis de un conjunto dado de transacciones. En esta oportunidad, siguiendo el primer ejemplo, los ítems son productos y el conjunto de transacciones contienen las compras de productos realizadas por diversos clientes, por lo que cada transacción refiere a los productos comprados conjuntamente por un solo cliente. A continuación, se detallan los conjuntos de ítems y de transacciones *D* y *T*, respectivamente.

D = {Cerveza, Coca, Huevos, Leche, Pan, Pañales}

T = {{Pan, Leche, Pañales}, {Pan, Pañales, Cerveza, Huevos}, {Leche, Pañales, Cerveza, Coca}, {Pan, Leche, Pañales, Cerveza}, {Pan, Leche, Pañales, Coca}}

Al colocar los distintos ítems de las transacciones como columnas de una tabla y en cada celda el símbolo de conteo "/" o los valores *true-false*, dependiendo de la presencia o no del producto en la transacción, se tiene la Tabla 1.

Tabla 1. Disposición y conteo de los ítems de las transacciones

TID	Cerveza	Coca	Huevos	Leche	Pan	Pañales
1				/	/	/
2	/		/		/	/
3	/	/		/		/
4	/			/	/	/
5		/		/	/	/

Fuente: elaboración propia.

El enfoque es el que sigue el *software* R para encontrar reglas de asociación y medidas interesantes de ellas [14]. Se presentan en orden alfabético los productos, dado que la salida es de este modo. El acrónico TID se emplea para la identificación de las transacciones. Por otro lado, la Figura 1 muestra el archivo de extensión csv que contiene a las transacciones.

```
Id,Producto
1,Pan
1,Leche
1,Panales
2,Pan
2,Panales
2,Cerveza
2,Huevos
3,Leche
3,Panales
3,Cerveza
3,Coca
4,Pan
4,Leche
4,Panales
4,Cerveza
...
```

Figura 1. Disposición de las transacciones para el archivo csv
Fuente: elaboración propia

ISSN 1794-211X • e-ISSN 2322-939X • Vol 17, N° 1 (enero-junio 2020). pp. 24-31. Universidad Distrital Francisco José de Caldas-Facultad Tecnológica.

La Figura 2 muestra en la parte superior el código R que procesa el archivo con las transacciones, mientras que en la inferior se puede observar la imagen con los resultados de la ejecución de las tres últimas acciones en la consola de R.

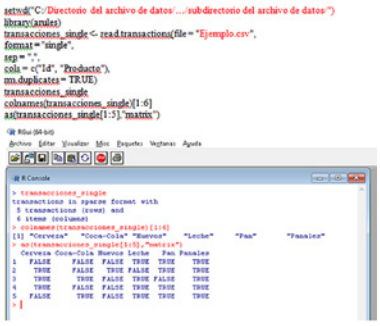

Figura 2. Código R y resultado de la ejecución en la consola de R
Fuente: elaboración propia.

El código R y la vista de la consola, correspondiente a las reglas de asociación con soporte y confianza ≥ 0.6, se proporciona en la Figura 3.

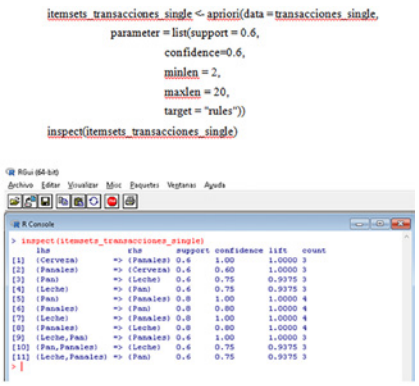

Figura 3. Código R y la vista de las reglas de asociación con soporte y confianza ≥ 0.6
Fuente: elaboración propia

De acuerdo con las definiciones dadas en (2) y (3), las medidas soporte y confianza para la regla {Pañales}⇒ {Cerveza} son:

soporte ({Pañales} ⇒ {Cerveza}) = *soporte* ({Pañales, Cerveza}) = *P*(Pañales y Cerveza) = $\frac{3}{5}$

confianza ({Pañales} ⇒ {Cerveza}) =

$$\text{soporte}(\{\text{Pañales, Cerveza}\}) \Big/ \text{soporte}(\{\text{Pañales}\}) =$$

$$= P(\text{Pañales y Cerveza}) \Big/ P(\text{Pañales}) = \frac{3}{5} \Big/ \frac{5}{5} = \frac{3}{5} = P\left(Y \Big/ X\right)$$

Lo anterior muestra el enfoque estadístico conjuntamente con las definiciones de la teoría de reglas de asociación. Los valores correspondientes al soporte de cada ítem se indican en la Tabla 2.

Tabla 2. Valores correspondientes al soporte de los ítems

TID	Conteo o frecuencia absoluta	Soporte
Cerveza	3	0.6
Coca	2	0.4
Huevos	1	0.2
Leche	4	0.8
Pan	4	0.8
Pañales	5	1

Fuente: elaboración propia

El valor de soporte lo decide el usuario y se elige considerando el mínimo de ocurrencias deseadas en el total de transacciones. Si en el ejemplo se pretende como mínimo 3 ocurrencias en el total de las 5 transacciones, se estará requiriendo reglas con un soporte igual o superior a $\frac{3}{5}$, esto es soporte ≥ 0.6. Si bien se pueden encontrar *software* libres que permiten descubrir reglas de asociación conjuntamente con las métricas más comunes tales como confianza, soporte y *lift*, entre otras, se opta por el uso de R. La elección de R (Project) se debe a que permite realizar la visualización y comparación con los resultados obtenidos manualmente, aplicando solo conceptos de probabilidad.

Cuando las transacciones están conformadas por ítems de variables disjuntas en R se utiliza el formato *basket*. Para este caso, se presenta el segundo ejemplo con los datos del archivo que provee Weka. El correspondiente archivo csv se muestra en la Figura 4.

La Figura 5 proporciona el código R y la correspondiente vista de la consola relativa al ejemplo con formato *basket*.

ISSN 1794-211X • e-ISSN 2322-939X • Vol 17, N° 1 (enero-junio 2020). pp. 24-31. Universidad Distrital Francisco José de Caldas-Facultad Tecnológica.

```
1,Soleado,Alta,Alta-H,FALSO,No
2,Soleado,Alta,Alta-H,Cierto,No
3,Cubierto,Alta,Alta-H,FALSO,Si
4,Lluvioso,Suave,Alta-H,FALSO,Si
5,Lluvioso,Fría,Normal,FALSO,Si
6,Lluvioso,Fría,Normal,Cierto,No
7,Cubierto,Fría,Normal,Cierto,Si
8,Soleado,Suave,Alta-H,FALSO,No
9,Soleado,Fría,Normal,FALSO,Si
10,Lluvioso,Suave,Normal,FALSO,Si
11,Soleado,Suave,Normal,Cierto,Si
12,Cubierto,Suave,Alta-H,Cierto,Si
13,Cubierto,Alta,Normal,FALSO,Si
14,Lluvioso,Suave,Alta-H,Cierto,No
```

Figura 4. Contenido del archivo csv, con los datos traducidos y rescatados de weather.nominal.arff
Fuente: elaboración propia

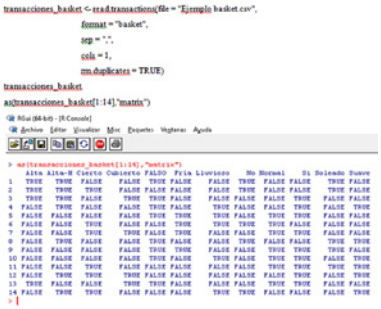

Figura 5. Código R y vista de la consola con el ejemplo con formato *basket*
Fuente: elaboración propia.

El valor "AltaH" de la variable humedad se distingue así dado que de ser tomado solo como "Alta", como lo presentan algunos documentos de la web, habría un mismo valor correspondiente a distintas variables (columnas) y el *software* no podría realizar dicha diferenciación. Los valores "*hot*" para la variable *temperature* y "*high*" para *humidity* están traducidos como "Alta" en ambos casos. Lo que el algoritmo realiza internamente es colocar los ítems, independientemente de las variables a la que pertenecen, como rótulo de columnas de una tabla y luego completarla con los valores *true-false*, tal como lo expuesto en la Tabla 1 y Figura 2.

Prosiguiendo con el propósito de que los alumnos empleen los conceptos adquiridos, se calculan los valores de soporte, confianza y *lift* correspondientes a la regla {Suave}⇒ {AltaH} y se comparan con los obtenidos a través del software R.

Para calcular la medida *lift* y apreciar la conexión entre asignaturas, se torna interesante mostrar la tabla de contingencia para las variables citadas y de ahí realizar el cálculo (Tabla 3).

Tabla 3. Tabla de contingencia para las variables temperatura y humedad

		Humedad		
		AltaH	**Normal**	**Total**
Temperatura	Alta/cálida	3	1	4
	Fría	0	4	4
	Suave	4	2	6
Total		7	7	14

Fuente: elaboración propia

Luego, los valores de las medidas son:

$$lift\left(\{\text{Suave}\}\Rightarrow\{\text{AltaH}\}\right)=\frac{P\left(\text{AltaH}/\text{Suave}\right)}{P\left(\text{AltaH}\right)}=\frac{4/6}{7/14}=1.33$$

$4/14$ = *P*(AltaH y Suave) = *soporte* ({Suave}⇒{AltaH})

$$P(\text{AltaH})=\frac{7}{14} \qquad \text{y} \qquad P(\text{Suave})=\frac{6}{14}$$

Es importante recalcar al estudiante que el numerador, para el cálculo de *lift*, corresponde a la definición de probabilidad condicional, donde los casos se restringen a solo los que contienen suave, lo que hace que el denominador sea 6 y de estos 6 solo 4 tienen temperatura alta.

Para el cálculo de la confianza de la regla citada puede ser provechoso realizar consultas SQL o bien filtros mediante una hoja de cálculo. En el caso de filtros, primero por temperatura "Suave", donde esta cantidad es el denominador y luego contar cuántas de ellas tienen humedad "AltaH" para obtener así el numerador (Figura 6).

ISSN 1794-211X • e-ISSN 2322-939X • Vol 17, N° 1 (enero-junio 2020). pp. 24-31. Universidad Distrital Francisco José de Caldas-Facultad Tecnológica.

	A	B	C	D	E
1	Cielo	Temperat	Humedad	Viento	Jugar
5	Lluvioso	Suave	AltaH	FALSO	Si
9	Soleado	Suave	AltaH	FALSO	No
11	Lluvioso	Suave	Normal	FALSO	Si
12	Soleado	Suave	Normal	Cierto	Si
13	Cubierto	Suave	AltaH	Cierto	Si
15	Lluvioso	Suave	AltaH	Cierto	No
16					
17					
18					

Figura 6. Captura de pantalla luego de realizar filtro por temperatura "Suave"
Fuente: elaboración propia.

De esta forma, el valor de la confianza es:

$$confianza\ (\{Suave\} \Rightarrow \{AltaH\}) = \frac{4}{6} = 1.3 = P\left(AltaH \,/_{Suave}\right)$$

3. Conclusiones

Se considera que impartir este tema en la asignatura Inteligencia Artificial, o en la que correspondiese, resulta beneficioso tanto para el alumno como para los docentes involucrados, ya que es más factible que resulte así una integración exitosa. Los alumnos pueden apreciar una aplicación concreta del campo laboral con el empleo de conceptos adquiridos en otras asignaturas.

Un docente podría diseñar una unidad didáctica con lo presentado en el cuerpo del artículo, por ejemplo, construir actividades como las detalladas en el acápite "Descripción". Se juzga conveniente un tiempo medio de cinco horas áulicas para la realización de la experiencia y que la misma se lleve a cabo en un gabinete con el *software* ya instalado, o bien que los estudiantes lo instalen previamente en sus computadoras personales.

La experiencia presentada se considera interesante y positiva, ya que, con ejercicios simples que no suponen complejidad alguna y que no requieren de muchos conceptos, el alumno podrá integrar temas previamente impartidos.

Cabe mencionar que los ejemplos trabajados son los que aparecen en mucha de la literatura de ciencia de datos. Se sugiere implementar este tipo de práctica en distintas asignaturas para ir combinando una integración de conceptos asistida por los docentes con la integración que deben realizar los alumnos por ellos mismos, ya que, si bien algunos estudiantes lo pueden asimilar de manera natural, otros parecieran ver los contenidos por primera vez o no le es fácil trasladar lo visto en otras asignaturas y volcarlo al nuevo conocimiento cuando es requerido.

Es recomendable que los docentes intenten mencionar posibles aplicaciones relativas a la carrera y disponer o facilitar lecturas relativas a ello.

Agradecimientos

Agradecemos la valiosa colaboración prestada por María José Marcovecchio, docente e investigadora de la Universidad Nacional de San Juan, en la revisión y confección del abstract.

Referencias

[1] Seguimos Educando, "Ley de Educación Superior". [En línea]. Disponible en: https://www.educ.ar/recursos/91820/ley-de-educacion-superior

[2] D. Pérez, C. M. Rodríguez, L. Padrón, J. Padrón y M. V. Velázquez, "La interdisciplinariedad en el proceso de enseñanza aprendizaje", *Odiseo, Revista Electrónica de Pedagogía*. [En línea]. Disponible en: https://odiseo.com.mx/marcatexto/la-interdisciplinariedad-en-el-proceso-de-ensenanza-aprendizaje/

[3] Ministerio de Educación, Ciencia y Tecnología, "Resolución 1610/2004". [En línea]. Disponible en: https://www.coneau.gob.ar/archivos/resoluciones/RESME1610-04.pdf

ISSN 1794-211X • e-ISSN 2322-939X • Vol 17, N° 1 (enero-junio 2020). pp. 24-31. Universidad Distrital Francisco José de Caldas-Facultad Tecnológica.

[4] G. Beguerí, A. Malberti, y R. O. Klenzi, "Integrando áreas disciplinares en un diseño curricular", en *VII Congreso de Tecnología en Educación y Educación en Tecnología*, Buenos Aires, junio 2012. http://sedici.unlp.edu.ar/handle/10915/18342

[5] W. Vélez, "La integración del conocimiento como fundamento de los estudios generales", *Revista Ciencia y Sociedad*, vol. 38, n.° 4, pp. 643-658, 2013. https://doi.org/10.22206/cys.2013.v38i4.pp643-658

[6] L. Medina y L. L. Guzmán, *Innovación curricular en instituciones de educación superior. Pautas y procesos para su diseño y gestión*. Ciudad de México: Asociación Nacional de Universidades e Instituciones de Educación Superior, 2011.

[7] A. P. León, E. Risco y C. Alarcón, "Estrategias de aprendizaje en educación superior en un modelo curricular por competencias", *Revista de la Educación Superior*, vol. 43, n.° 172, pp. 123-144, 2014. https://doi.org/10.1016/j.resu.2015.03.012

[8] R. Agrawal, T. Imieliński y A. Swami, "Mining association rules between sets of items in large databases", en *ACM SIGMOD international conference on Management of data*, Washington D.C., 1993. https://doi.org/10.1145/170035.170072

[9] M. Syan, J. Han y P. S Yu, "Data mining: an overview from a database", *IEEE Transactions on Knowledge and data Engineering*, vol. 8, n.° 6, pp. 866--883,1996. https://doi.org/10.1109/69.553155

[10] J. Hipp, U. Güntzer y G. Nakhaeizadeh, "Algorithms for association rule mining. A general survey and comparison", ACM SIGKDD Explorations Newsletter, vol. 2, n.° 1, pp. 58-64, 2000. https://doi.org/10.1145/360402.360421

[11] M. Hahsler, K. Hornik y T. Reutterer, "Implications of probabilistic data modeling for mining association rules", en *Data and Information Analysis to Knowledge Engineering*, Berlin, 2006. https://doi.org/10.1007/3-540-31314-1_73

[12] C. Silverstein, S. Brin, R. Motwani, "Beyond market baskets: Generalizing association rules to dependence rules", *Data Mining and Knowledge Discovery*, vol. 2, n.° 1, pp. 39-68, 1998. https://doi.org/10.1145/253262.253327

[13] M. A. Malberti y G. Elida, "Reglas de Asociación con los datos de una biblioteca universitaria", *Revista Cubana de Ciencias Informáticas*, vol. 9, n.° 4, pp. 30-45, 2015. https://doi.org/10.29019/enfoqueute.v6n2.62

[14] R Foundation, "The R Project for Statistical Computing". [En línea]. Disponible en: https://www.r-project.org/

ISSN 1794-211X • e-ISSN 2322-939X • Vol 17, N° 1 (enero-junio 2020). pp. 24-31. Universidad Distrital Francisco José de Caldas-Facultad Tecnológica.