

# SISTEMA DE MENÚ INTERACTIVO APOYADO POR UN ALGORITMO BASADO EN REGLAS DE ASOCIACIÓN PARA LA ALIMENTACIÓN DE PACIENTES INTERNOS DEL HOSPITAL DEL PACÍFICO

Rosa Maribel Martínez Manzo Ing.<sup>1</sup>, M.T.I. Eloy Cadena Mendoza<sup>2</sup>,  
Dr. Eduardo de la Cruz Gámez<sup>3</sup>, Dra. Miriam Martínez Arroyo<sup>4</sup>

**Resumen** En el presente trabajo se verá el funcionamiento de un algoritmo que genera reglas de asociación, así como su aplicación en el proyecto a desarrollar. Éstas ayudarán a indicar las diferentes posibilidades de cambios entre los elementos de un mismo grupo de alimentos en el menú del Hospital del Pacífico del puerto de Acapulco de Juárez, Guerrero.

**Palabras clave** -- Alimentación, Algoritmos, Reglas, Asociación.

## Introducción

En el hospital la alimentación de los pacientes juega un papel muy importante para la recuperación de los mismos, puesto que en los alimentos se encuentran los nutrientes necesarios para que la curación del cuerpo y la medicación se optimicen. En este sentido la creación de un menú interactivo al que tengan acceso los pacientes hospitalizados a través de una aplicación en sus teléfonos móviles proporciona al mismo la oportunidad de verse involucrado en su alimentación y por tanto en su recuperación de manera más positiva.

Por lo tanto, la utilización de un algoritmo de reglas de asociación aplicado a los diversos grupos de alimentos permitirá la posibilidad de intercambio de elementos del mismo grupo basándose en las necesidades alimenticias de los pacientes del hospital. Se creó una Base de Datos en donde son almacenados tanto los ingredientes por grupos alimenticios como los menús con los que cuenta el hospital, ya que esto propicia que la búsqueda con las reglas sea más centralizada; se busca un elemento del mismo grupo con similitudes nutricionales para que éste sea tomado como sustituto.

Esto es posible ya que el médico al ingresar el paciente y hacer el diagnóstico del mismo le asigna una dieta acorde a su padecimiento o condición, esta instrucción es llevada a cabo por los enfermeros quienes se encargan de que el paciente cumpla con la dieta establecida e indican a la cocina lo que puede y no puede comer el paciente, para que así se preparen los alimentos.

## Descripción del Método

### *Las reglas de asociación*

Las reglas de asociación han sido el objetivo de muchos trabajos de investigación desde que Agrawal et al., propusieran el algoritmo de aprendizaje A priori (Agrawal, R. and Srikant, R., 1994) y su utilización en grandes bases de datos (Agrawal, R., Imielinski, T., Swami, A., 1993). Haciendo uso de su notación, podemos definir una regla de asociación como una implicación de la forma  $X \Rightarrow Y$ , donde X se denomina antecedente e Y consecuente. Tanto X como Y estarán formados por conjuntos de elementos pertenecientes a la tabla de transacciones T que estamos analizando. Una tabla de transacciones consta de un número indeterminado de registros que contienen diferentes secuencias de valores de los atributos que definen un registro. Los atributos que forman cada uno de los registros

<sup>1</sup> Rosa Maribel Martínez Manzo Ing. estudiante de la Maestría en Sistemas Computacionales del Instituto Tecnológico de Acapulco. [rosa.mtz2408@gmail.com](mailto:rosa.mtz2408@gmail.com) (autor correspondiente)

<sup>2</sup> El M.T.I. Eloy Cadena Mendoza es Profesor en la Maestría en Sistemas Computacionales del Instituto Tecnológico de Acapulco. [eloy\\_cadena@yahoo.com](mailto:eloy_cadena@yahoo.com)

<sup>3</sup> El Dr. Eduardo de la Cruz Gámez es Jefe de la División de Estudios de Posgrado e Investigación y Profesor en la Maestría en Sistemas Computacionales del Instituto Tecnológico de Acapulco. [gameduardo@yahoo.com](mailto:gameduardo@yahoo.com)

<sup>4</sup> La Dra. Miriam Martínez Arroyo es Profesora y Coordinadora de la Maestría en Sistemas Computacionales del Instituto Tecnológico de Acapulco. [miriamma\\_ds@hotmail.com](mailto:miriamma_ds@hotmail.com)

dependerán del campo de aplicación. La búsqueda de asociaciones suele dar lugar a la obtención de un número muy elevado de reglas. Para seleccionar las más representativas, el proceso debe ir seguido por una evaluación de las mismas. Las medidas más empleadas para estimar la validez de una regla son las que aparecen descritas a continuación (Ma. Dolores Ruiz, 2010):

**Support (cobertura):** La cobertura expresa el tanto por ciento de registros de T que satisfacen la unión de los elementos del consecuente y del antecedente  $s(X \Rightarrow Y) = s(X \cup Y)$  la fórmula anterior indica que el conjunto s, donde X implica a Y es igual al conjunto s donde X se une con Y.

**Confidence (confianza):** La confianza es una medida de la efectividad de una regla. Representa el porcentaje de casos en los que dado el antecedente se verifica la implicación.  $c(X \Rightarrow Y) = s(X \Rightarrow Y) / s(X)$  dado el conjunto c donde X implica a Y es igual al cociente del conjunto s donde X implica a Y sobre el conjunto de s de X.

Puede utilizarse para estimar la probabilidad condicionada del consecuente dado el antecedente:  $P(Y/X) = P(X \cup Y) / P(X) = c(X \Rightarrow Y)$  P es la probabilidad del cociente de Y sobre X, esto es igual al cociente del conjunto P donde X se une con Y sobre la probabilidad de que ocurra X, que es lo mismo que el conjunto c donde X implica Y.

**Lift:** Cuantifica la relación existente entre X e Y:

- lift > 1: X e Y positivamente correlacionados
- lift < 1: X e Y negativamente correlacionados
- lift = 1: X e E independientes.

$li(X \Rightarrow Y) = s(X \Rightarrow Y) / s(Y)$  esto se lee como el levantamiento donde X implica Y que es igual al cociente del soporte donde X implica Y sobre s de X.

Las reglas de asociación son un algoritmo bien definido, cuyo significado se mide por factores de soporte y confianza. Se componen de una condición y un resultado. Generalmente se representan de la forma: "Si (if) condición entonces (then) resultado". Es una implicación de la forma  $X \rightarrow Y$  que significa que en las tuplas de datos para las cuales valen los atributos de X, también valen los atributos de Y.

#### El algoritmo A priori

El algoritmo A priori tiene como objetivo la extracción de reglas de asociación de una base de datos de transacciones. Puede descomponerse en dos tareas:

- 1.- Encontrar todos los conjuntos de elementos, que tienen una cobertura por encima de la mínima cobertura dada
- 2.- Utilizar los conjuntos de elementos con mayor cobertura de la fijada como umbral para generar reglas que superen un cierto nivel de confianza.

La notación del algoritmo A priori (R. Agrawal and R. Srikant, 1994) se ve en la tabla 1:

k-itemset	Un itemset con k ítems
$L_k$	Conjunto de itemsets frecuentes k (aquellos con soporte mínimo)
$C_k$	Conjunto de itemsets candidatos k (itemsets potencialmente frecuentes)

Tabla 1. Notación del algoritmo A priori (R. Agrawal and R. Srikant, 1994)

A continuación, se muestra el pseudocódigo del algoritmo A priori:

```

L1 = { Conjunto de itemsets frecuentes 1 }
for (k=2; Lk-1≠0; k++) do begin
Ck = apriori-gen(Lk-1) // Nuevos candidatos
forall transacciones t ∈ D do begin
Ct = subconjunto (Ck, t); // Candidatos en t
forall candidatos c ∈ Ct do
c.count++;
end
Lk = { c ∈ Ck | c.count ≥ minsup }
end Answer UkLk
    
```

La primera pasada del pseudocódigo algoritmo cuenta las ocurrencias de los ítems en todo el conjunto de datos para determinar los itemsets frecuentes 1. Los subsiguientes pasos del algoritmo son básicamente dos: primero, los itemsets frecuentes  $L_{k-1}$  encontrados en la pasada (k - 1) son usados para generar los itemsets candidatos  $C_k$ , usando la función apriori-gen descrita en la siguiente subsección. Y segundo se cuenta el soporte de los itemsets candidatos

C<sub>k</sub> a través de un nuevo recorrido a la base de datos. Se realiza el mismo proceso hasta que no se encuentren más itemsets frecuentes. (J. C. Alvarado, 2010).

### Implementación

Para el desarrollo del sistema se utiliza una laptop DELL con procesador Intel Core i5 de 64 bits, 6Gb de RAM. El sistema operativo sobre el cual se está trabajando es un equipo HP, con Windows 10 Pro, Procesador Intel Core 2 Duo de 64 bits, 3Gb de RAM. El lenguaje de programación es Java, con entorno de desarrollo NetBeans 8.1 y Weka, para la parte del móvil se utiliza el entorno de desarrollo Android Studio 3.0, la base de datos está en el SGBD Microsoft SQL Server Management Studio 17.

En la Figura 1 se muestra un diagrama de clases desarrollado con la herramienta Enterprise Architect 12, el cual muestra algunas de las relaciones tanto de las tablas en la base de datos del menú y el inventario, así como las conexiones de la base con las ventanas en la APK que se visualizarán desde el teléfono celular.

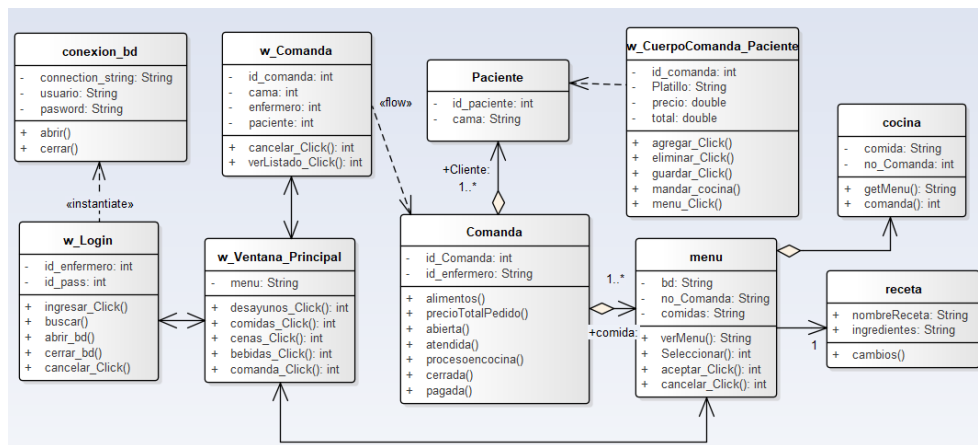


Figura 1 Diagrama de clases (Elaboración propia)

Se creó una base de datos para el sistema, en la cual se muestra la conexión a la base de datos y las relaciones entre las tablas, cada tabla tiene tanto sus atributos específicos como los que comparte con otras tablas. En la Figura 2 se muestra la base de datos. Está conformada de la siguiente manera, como tabla principal está la tabla “Comanda”; en dicha tabla se levanta el pedido del paciente, está ligada a la tabla “Pacientes”, “Enfermero” y “TipoDieta” y esta última está ligada a la tabla “Receta” que a su vez tiene relación a las tablas “Desayuno”, “Comida” y “Cena”

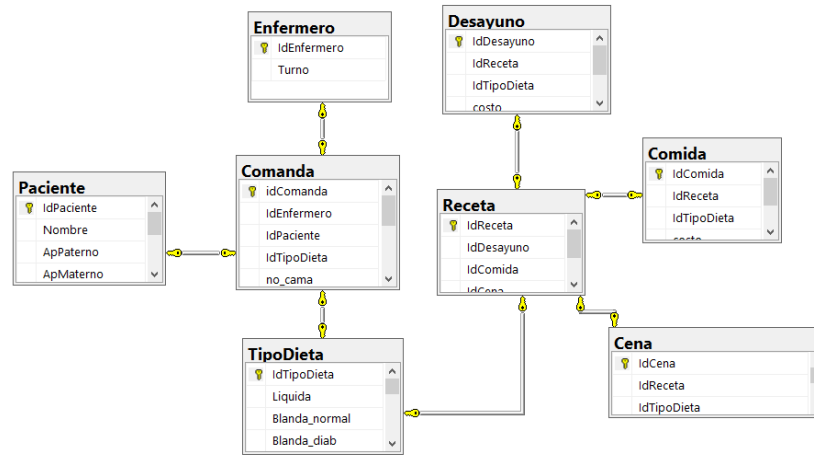


Figura 2. Base de Datos (Elaboración propia)

## Resultados

### Generación de las reglas de asociación

Se utilizó la herramienta Weka. Ésta es un software de código abierto emitido bajo la licencia pública general GNU (Licencia Weka 2019) para la implementación del algoritmo A priori, se realizó una prueba a uno de los grupos “frutas” de alimentos que se dan en el hospital, en este grupo se muestran las frutas y sus propiedades, como las calorías, hidratos, proteínas, grasas y fibras entre otros. Esta información se pasó a un formato que el programa pudiera leer, Figura 4.

El fichero frutas.arff contiene datos sobre las características de las frutas que se dan en el hospital, ya sea para prepararse en jugos o para darse como postres o desayunos. El formato arff (Attribute-Relation file format) consiste, simplemente, en un fichero de texto en el que se almacena una tabla de datos, con una línea por tupla y los valores de una misma tupla separados por comas (en la misma línea del fichero de texto). Adicionalmente, los ficheros arff incluyen una cabecera con información adicional acerca de los nombres y tipos de datos asociados a los distintos atributos de la relación, tal como se muestra en la Figura 4.

```
frutas: Bloc de notas
Archivo Edición Formato Ver Ayuda
@relation frutas
@attribute fruta {manzana, naranja, pera, pinya, sandia}
@attribute hidratos {5,9,10,11,12}
@attribute proteinas {0,1}
@attribute grasas {0,1}
@attribute agua {85,86,87,88,89,94}
@attribute fibra {1,2,3}
@attribute kcal {20,40,42,50}
@data
manzana,10,0,0,87,2,42
naranja,9,1,0,87,3,40
pera,12,1,0,85,2,50
pinia,11,1,0,87,1,50
sandia,5,1,0,94,1,20
```

Figura 4. Archivo “frutas.arff”.

Al ingresar el archivo frutas.arff a la aplicación Weka y en la sección de “Associate” se seleccionó el algoritmo A priori con las características de que generará solo 10 reglas con un mínimo de confianza del 0.9 y un mínimo de soporte de 0.1, y entonces se llegó al siguiente resultado, el cual muestra que hay frutas que comparten la misma cantidad de un atributo, por ejemplo hay tres frutas que contienen 87 por ciento de agua por cada 100 gramos de la fruta y estas son la naranja, la manzana y la piña, hay mucha similitud entre algunas frutas.

```
Best rules found:
1. proteinas=1 4 ==> grasas=0 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. agua=87 3 ==> grasas=0 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. fibra=1 2 ==> proteinas=1 2 <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.4)
4. kcal=50 2 ==> proteinas=1 2 <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.4)
5. fibra=1 2 ==> grasas=0 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. fibra=2 2 ==> grasas=0 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. kcal=50 2 ==> grasas=0 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. proteinas=1 agua=87 2 ==> grasas=0 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. grasas=0 fibra=1 2 ==> proteinas=1 2 <conf:(1)> lift:(1.25) lev:(0.08) [0] conv:(0.4)
10. proteinas=1 fibra=1 2 ==> grasas=0 2 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
```

Figura 5. Reglas generadas con Weka.

## Conclusiones

Entre los múltiples algoritmos para la generación de las reglas de asociación el algoritmo A priori resultó ser el más adecuado puesto que es el más rápido al momento de procesar la información, en este documento se mostró a manera de pseudocódigo su funcionamiento. El uso de las reglas de asociación las cuales corresponde a una implicación del tipo  $if \rightarrow then$  ayudan al desarrollo del programa haciendo posible la relación de intercambio entre elementos. En cuanto al desarrollo del sistema y de aplicación móvil se han tenido que realizar múltiples investigaciones para que su funcionamiento sea óptimo y las conexiones entre ellos no se interfieran.

## Referencias

- Agrawal, R. and Srikant, R. Fast algorithms for mining association rules. Proc. 20th Int. Conf. Very Large Data Bases, (VLDB-94), 1994, pp 487-499.
- Agrawal, R., Imielinski, T., Swami, A. Mining association rules between sets of items in large databases. Proceedings of the International Conference on Management of Data (SIGMOD-1993), Washington, D.C., 1993, pp. 207-216 (ISBN:0-89791-592-5)
- Descripción de las funciones de Android Studio en el link <https://developer.android.com/studio/features.html> (2018).
- Descripción de las funciones de Android Studio en el link: <https://developer.android.com/studio/run> (2019).
- Descripción del algoritmo A priori en WEKA en el link: <http://weka.sourceforge.net/doc.stable-3-8/weka/associations/Apriori.html> (2019).
- Descripción de la herramienta Weka, software de código abierto emitido bajo la Licencia Pública General de GNU en el link: <https://www.cs.waikato.ac.nz/ml/weka/>.(2019)
- Juan Carlos Alvarado Pérez, Implementación de TaryKDD y Evaluación de Rendimiento de Nuevos Algoritmos, DPTOIA-IT, Departamento de Informática y Automática Universidad de Salamanca Junio, 2010
- María Dolores Ruiz Jiménez Modelado formal para representación y evaluación de reglas de asociación, D.L.: GR 2323-Editorial de la Universidad de Granada, 2010, pp. 12-14 (2010 ISBN: 978-84-693-2336-7)
- R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In VLDB Conference, Santiago, Chile, 1994