

**INTEGRACIÓN DE LOS  
ALGORITMOS KNN Y  
FP-GROWTH PARA APOYAR  
LA GESTIÓN DE RELACIONES  
CON CLIENTES**

**INTEGRATION OF  
ALGORITHMS KNN AND  
FP-GROWTH TO SUPPORT  
THE MANAGEMENT OF  
RELATIONS WITH CLIENTS**

**INTEGRAÇÃO DOS  
ALGORITMOS KNN E  
FP-GROWTH PARA APOIAR  
A GESTÃO DE RELAÇÕES  
COM CLIENTES**

**Juan Sebastián Mejía Goyeneche <sup>a</sup>,  
Royer David Estrada Esponda <sup>b</sup>**

<sup>a</sup> Facultad de Ingeniería,  
Ingeniería de Sistemas,  
Universidad del Valle, sede Tuluá,  
**sebastian.mejia.710@gmail.com**

<sup>b</sup> Facultad de Ingeniería,  
Ingeniería de Sistemas,  
Universidad del Valle, sede Tuluá,  
**royer.estrada@correounivalle.edu.co**

Fecha de recepción: 30 de junio 2017  
Fecha de aprobación: 24 de abril 2018

## Resumen

Actualmente las organizaciones se encuentran un entorno sumamente competitivo, en el que la administración de las relaciones con los clientes cobra un valor significativo, especialmente en lo relacionado a la fidelización, el aumento de las ventas y el incremento de la rentabilidad. En consecuencia, las organizaciones están llamadas a administrar las relaciones con sus clientes desde el punto de vista estratégico, pero también desde el punto de vista táctico y operativo. Por tal motivo, la tecnología es uno de esos componentes que apoya tal gestión y que potencializa su efectiva implementación. El presente artículo expone los resultados de un proyecto de grado<sup>28</sup>, que abordó la problemática concerniente a la gestión de relaciones con los clientes en un restaurante de comidas rápidas de Tuluá, basado en la metodología SCRUM y técnicas de minería de datos, el resultado fue un sistema informático para el apoyo a la toma de decisiones en el contexto de la inteligencia de negocios. Se destaca la integración de los algoritmos de vecinos cercanos (KNN) y FP-GROWTH, con el fin de emitir recomendaciones en relación con las preferencias de los consumidores con base en la composición de los productos más consumidos por ellos. Además, una vez integrados los algoritmos se probaron en cuatro escenarios diferentes, lo que permitió concluir que tal integración no tiene asociado un costo computacional elevado.

**Palabras clave:** CRM, FP-GROWTH, Inteligencia de Negocios, SCRUM, KNN.

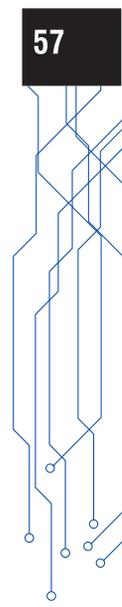
## Abstract

The organizations extremely are competitive surroundings, in which the administration of the relations with the clients acquires a significant, especially in the related thing to the loyalty, the increase of the sales and the increase of the yield. Consequently, the organizations are called to administer the relations with their clients from the strategic point of view, but also from the tactical and operative point of view. By such reason, the technology is one of those components that such management supports and that its effective implementation potentialises. The present article sets out the results of a degree project, that approached problematic concerning the management of relations with the clients in a restaurant of fast meals of Tuluá, based on the methodology the SCRUM and techniques of mining of data, the result was a computer science system for the support to the decision making in the context of the intelligence of businesses. The integration of the algorithms of near neighbors (KNN) and FP-GROWTH stands out, with the purpose of emitting recommendations in relation to the preferences of the consumers with base in the composition of products more consumed by them. In addition, once integrated the algorithms they tried on in four different scenes, which allowed to conclude that such integration does not have associate a high computer cost.

**Key words.** CRM, FP-GROWTH, Intelligence of Businesses, SCRUM, KNN.

## Resumo

As organizações são encontradas resumo-atualmente um ambiente superiormente competitivo, onde a administração das relações com os clientes percebe um valor significativo, especialmente que está relatório fidelisation, o aumento das vendas e o crescimento da rentabilidade. Por conseguinte, as organizações



são chamadas de administrar as relações com os seus clientes do ponto de vista estratégico, mas também o ponto de vista tático e operacional. Por tal motivo, a tecnologia é um destes componentes que apoia tal gestão e que potencialize a sua aposta em obra efetiva. O presente artigo expõe os resultados de um projeto de grau [1], que abordou a problemática relativa à gestão de relações com os clientes num restaurante de restaurações rápidas de Tuluá, baseado a metodologia SCRUM e técnicas de indústria mineira de dados, o resultado foi um sistema informático para o apoio à tomada de decisão no contexto da inteligência de negócios. Sublinha-se a integração dos algoritmos de vizinhos próximos (KNN) e de FP-GROWTH, a fim de emitir recomendações em relação às preferências dos consumidores com base na composição dos produtos mais consumidos por eles. Além disso, uma vez integrados os algoritmos foram provados em quatro cenas diferentes, que permitiu concluir que tal integração não associou um custo computacional que importa.

**Palavras chave.** CRM, FP-GROWTH, Inteligência de Negócios, SCRUM, KNN.

## I. INTRODUCCIÓN

Las organizaciones consideran la información como un recurso significativo para su operación y para la toma de decisiones, no solo porque esta es producto de sus procesos sino también una fuente que alimenta los negocios y que puede ser determinante para lograr el éxito. Sin embargo, para encaminarse hacia la vía del aprovechamiento de tal información, y obtener de ella soporte para la toma de decisiones, es necesario aplicar métodos para la producción, distribución, almacenamiento y recuperación de dicha información, de tal manera que de ésta se pueda obtener conocimiento nuevo acerca del negocio.

En consecuencia, es relevante comentar cómo los sistemas de información se encargan de facilitar dichas tareas, incluso Gunasekaran et al. (2001, citados en Meroño, 2004) comentan que los sistemas de información son un concepto amplio que abarca el diseño de flujos de procesos, de manera que se posibilite la satisfacción de las necesidades de información que las organizaciones manifiestan; así pues, estas

últimas están llamadas a comprender la marcada diferencia entre las Tecnologías de la Información y la Comunicación (TIC), y propiamente los sistemas de información (SI); de hecho, se encuentra que las primeras hacen referencia a “un término genérico para la convergencia de ordenadores, programas y telecomunicaciones” (Meroño, 2004, p.108), mientras que los Sistemas de Información, hacen referencia precisamente a la gestión de los flujos de procesos que se desarrollan al interior de las organizaciones y que son significativos para la operación de las mismas.

Por otra parte, la Gestión de Relaciones con los Clientes (Customer Relationship Management, CRM), es un término que cobra diferentes significados dependiendo de quién lo aborde. Un ingeniero o profesional de las TIC seguramente manifestará que CRM hace referencia a los programas informáticos que se usan para la gestión de sus procesos y que están orientados a sus clientes, mientras que un gerente probablemente comentará que tal término hace referencia a una estrategia organizacional, orientada al logro de una ventaja competitiva que

busca obtener una diferenciación en relación con sus competidores. Por lo anterior, es necesario comentar que CRM es una integración de personas, procesos y tecnologías, que buscan comprender a los clientes de una compañía; esa integración ha venido evolucionando por medio de avances en las tecnologías de la información y por medio de la incorporación de procesos centrados a los clientes (Chen y Popovich, 2003). La Fig. 1 da cuenta de dicha integración.

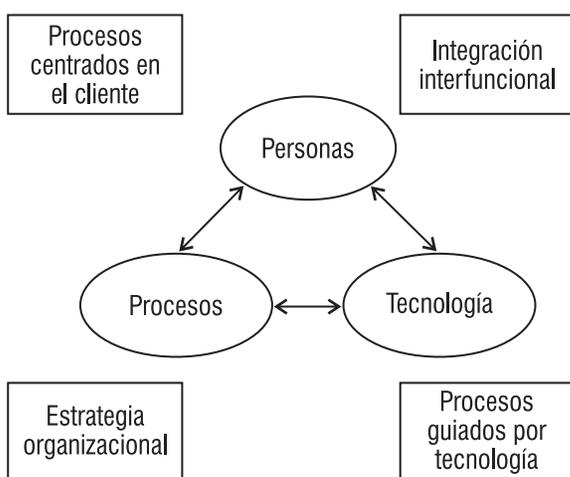


Figura 1: Modelo CRM

Fuente: Elaborado a partir de Chen y Popovich (2003)

Teniendo en cuenta lo anterior, es necesario que las soluciones informáticas no sólo se ubiquen en una labor simplemente operativa o transaccional, puesto que también deben ser aprovechadas para otorgar valor por medio de la entrega de información en contexto, que se traduzca entonces en conocimiento organizacional. De hecho, Chen y Popovich (2003) comentan que “sistemas CRM acumulan, almacenan, mantienen y distribuyen el conocimiento de los clientes a través de toda la organización” (p 677).

Así pues, y teniendo en cuenta que los sistemas informáticos contribuyen de manera significativa con la operación de las compañías, es indispensable extender tales sistemas en

verdaderos elementos diferenciadores de las organizaciones; dicho de otra manera, la ventaja competitiva no se materializa por la cantidad de información almacenada en grandes bases de datos, esta se obtiene gracias a la capacidad de las organizaciones por analizar e interpretar dicha información con el fin de adoptar estrategias a la luz del conocimiento que se encuentra oculto en ella, incluso las aplicaciones CRM aprovechan al máximo las innovaciones tecnológicas con su capacidad de recolectar y analizar datos de los patrones de sus clientes, interpretar sus comportamientos, desarrollar modelos predictivos, responder a tiempo y efectivamente a comunicaciones personalizadas y entregar productos y servicios que añadan valor a sus clientes individuales (Chen y Popovich, 2003).

Paralelamente, Brusco (2011) comenta que las organizaciones transitan por etapas de informatización que están representadas por los siguientes tres niveles:

- A. Difusión de la información al interior de las empresas, en el cual se busca dotar a las organizaciones de mecanismos e infraestructura que posibilite compartir información en su interior.
- B. Mejora de gestión interna, el cual busca que las organizaciones añadan eficiencia a sus procesos internos por medio de la incorporación de sistemas informáticos.
- C. Intercambio de información con clientes y proveedores, el cual hace referencia al establecimiento de relaciones que permitan el intercambio de información que añada valor a la organización y posibilite, para el caso de los clientes, establecer vínculos con el fin de personalizar la atención y reorientar procesos de marketing.

En consecuencia con lo anterior, la Inteligencia de Negocios (Business Intelligence, BI), permite a las

organizaciones llevar a cabo procesos e iniciativas que redundan en el establecimiento de relaciones más fuertes y personalizadas con sus clientes por medio de la aplicación de técnicas como la minería de datos, que se encarga de detectar patrones y comportamientos no muy evidentes en grandes volúmenes de datos o información.

La investigación realizada evidenció que un restaurante ubicado en la ciudad de Tuluá, no contaba con la implementación de un mecanismo que garantizara la información oportuna, permitiéndole coordinar y responder eficientemente a las demandas del mercado y a las necesidades presentes en cada uno de sus procesos, los cuales involucran tanto a clientes como a proveedores.

La información de estos procesos se manejaba de forma manual y se almacenaba en archivos físicos, dificultando el aprovechamiento que se podía obtener de ella, ya que al no estar organizada, no se podía consultar de manera rápida ni se podían determinar patrones y tendencias de comportamiento en los procesos.

Por lo anterior, resultó necesario contar con una herramienta para aprovechar la información manejada generando escenarios, pronósticos y reportes que apoyaran la toma de decisiones. Como solución a esta situación, se implementó un sistema de información WEB que permitió la administración de la información con el apoyo de la minería de datos, en los procesos relacionados con proveedores (compras) y clientes (ventas).

El artículo está estructurado de la siguiente forma: inicialmente se presenta el marco conceptual que permitirá a los lectores identificar los conceptos que sirvieron para el desarrollo de la investigación, así como para la interpretación de sus resultados; luego se presenta la metodología que guio el proceso investigativo y después se

presentan los resultados obtenidos. Finalmente, se consignan las conclusiones a partir de los resultados obtenidos y la interpretación de los autores.

## II. MARCO CONCEPTUAL

El marco conceptual que aquí se presenta busca ilustrar al lector de manera clara los conceptos que sirvieron como fundamento para el desarrollo de la investigación y busca facilitar la comprensión del desarrollo y resultados de la misma.

### A. *Inteligencia de negocios*

Rosado y Rico (2010) definen la inteligencia de negocios como la habilidad corporativa para tomar decisiones a partir del conocimiento obtenido una vez aplicada la recolección, depuración y transformación de datos; identificando así los problemas del negocio que pueden ser corregidos y las oportunidades que pueden ser aprovechadas. De esta manera, el principal objetivo de la inteligencia de negocios es apoyar de forma sostenible y continuada a la organización para mejorar su competitividad y facilitar la información necesaria para la toma de decisiones. Su origen está relacionado a proveer acceso directo a la información para soportar la toma de decisiones.

Según Cano (2007), otros beneficios que la BI brinda a las organizaciones son:

- 1) Beneficios tangibles como la reducción de costos, generación y/o aumento de ingresos y reducción de tiempos para las distintas actividades del negocio.
- 2) Beneficios intangibles como el hecho de tener disponible información para la toma de decisiones.

- 3) Beneficios estratégicos a través de la obtención del conocimiento que indique a qué clientes, a qué mercados o con qué productos dirigirse.

### **B. Minería de datos**

Según Virseda y Román (s.f), la minería de datos se encarga de agrupar las ventajas presentes en las áreas de la estadística, la inteligencia artificial, la computación gráfica, el procesamiento masivo de datos y las bases de datos, utilizando estas últimas como materia prima para obtener como resultado el descubrimiento automático de patrones o de modelos interesantes y no obvios escondidos en dichas bases de datos, los cuales tienen un gran potencial para contribuir en los aspectos principales del negocio; de manera que se apoye la definición de estrategias y tácticas encaminadas a cumplir con los objetivos de la organización.

Paralelamente, según Castañeda y Rodríguez (2003), la minería de datos se puede ver como un sistema de extracción de relaciones que agrupa métodos basados en la computadora (redes neuronales artificiales, árboles de decisión, inducción de reglas, lógica difusa y algoritmos genéticos) y que requiere poca participación por parte del analista en la obtención de información relevante.

### **C. Algoritmos de recomendación**

Según Marina (2014), partiendo de la información que se tenga almacenada en un sistema, los algoritmos de recomendación se encargan de establecer relaciones entre diferentes ítems con base en tendencias y patrones en los datos. Existe gran cantidad de algoritmos de recomendación, la mayoría de ellos con uno o varios parámetros y componentes configurables; por lo que suelen clasificarse principalmente en dos categorías: Filtrado

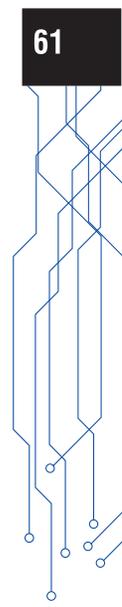
Colaborativo, que utiliza la información de los usuarios para hacer las recomendaciones a ellos mismos y a otros usuarios, y Basados en Contenido, que utilizan la descripción de cada ítem para realizar las recomendaciones.

Sin embargo, Betarte et al. (2006) señalan que un problema común de los algoritmos de recomendación, se presenta cuando un usuario nuevo usa el sistema y no es posible generar recomendaciones hasta que su perfil no se encuentre lo suficientemente determinado, como para encontrar su grupo de vecinos cercanos o elementos que puedan ser de su interés. A este problema se le conoce como Cold Start y se puede presentar también con la llegada de un nuevo elemento al sistema, debido a que no cuenta aún con calificaciones por parte de los usuarios. En la mayoría de los casos, se aborda este problema solicitando a los usuarios nuevas calificaciones sobre una serie de elementos con el propósito de determinar un perfil inicial para realizar las primeras recomendaciones.

### **D. Algoritmo Fp-growth**

Según Pinho (2010), el algoritmo FP-Growth se apoya en una estructura de datos llamada FP-Tree (árbol de patrones frecuentes). En esta estructura de datos se busca almacenar información acerca de conjuntos de ítems frecuentes de forma compacta, para tener mayor eficiencia al momento de realizar las consultas sobre los datos; es decir, se reduce el costo de computación del proceso de obtención de reglas de asociación ya que no es necesario generar los conjuntos de ítems candidatos a frecuentes y tampoco verificar si estos superan el umbral determinado.

Del mismo modo, Pinho (2010) también plantea que debido a que las transacciones de conjuntos de datos suelen compartir ítems frecuentes, el



tamaño del árbol FP-Tree suele ser mucho menor que el tamaño del conjunto de datos original.

La estructura del árbol FP-Tree consta de un nodo raíz que siempre es nulo y un conjunto de subárboles (correspondientes a las transacciones) hijos del nodo raíz (nodo nulo). Para la construcción del árbol, se recorrió el conjunto de transacciones (registros de ventas a clientes del restaurante) en busca de los ítems frecuentes (ingredientes y productos) y se contabilizan sus frecuencias para ser ordenados por esas frecuencias de manera decreciente. De esta forma, los ítems más frecuentes ocupan los niveles de menor nivel (los más cercanos a la raíz) del árbol. Posteriormente, se recorre el conjunto de transacciones solamente con los ítems frecuentes para obtener las ramas del árbol.

### E. Algoritmo *k*-Nearest Neighborhood (KNN)

García y Gómez (2015) definen el algoritmo KNN (K- vecinos más cercanos o algoritmo de los vecinos cercanos) a partir de las siguientes características:

- 1) Las predicciones se realizan basándose en los elementos más parecidos al que hay que predecir.
- 2) Se determina un conjunto de referencia *R* que es el conjunto de prototipos sobre el que se buscan los vecinos más cercanos.
- 3) No hay un modelo global asociado a los conceptos a aprender.
- 4) El coste del aprendizaje es 0 ya que todo el coste pasa al cálculo de la predicción.
- 5) Se conoce como mecanismo de aprendizaje perezoso (lazy learning).
- 6) Se debe especificar una métrica para poder medir la proximidad. La distancia euclidiana suele utilizarse para tales fines.

Con base en la descripción referente a algoritmos de recomendación, el algoritmo KNN se clasifica dentro de los algoritmos basados en contenido.

El procesamiento del algoritmo requiere de un conjunto de ítems de referencia, para los que va a identificar otros ítems similares, y un conjunto de ítems de muestra. El algoritmo se encarga de calcular la distancia euclidiana de cada ítem de referencia con cada ítem de muestra, seleccionando las *K* muestras más cercanas.

### F. Complejidad Ciclomática

La complejidad ciclomática o también conocida como complejidad condicional, según Cardacci (2016), es una métrica de software que ayuda a obtener una medida cuantitativa de la dificultad de crear pruebas automáticas del código y también es una medición orientativa de la fiabilidad del mismo.

Esta métrica no se centra en los bucles del código de un programa, sino en la cantidad de condiciones que se siguen. Además, permite estimar una aproximación al riesgo, costo y estabilidad de un programa y así determinar si el software es confiable y estable. Existen tres formas de calcular la complejidad condicional:

$$CC = A - N + 2 \quad (1)$$

Donde, CC es la Complejidad Condicional, A es el número de aristas del grafo (Una arista conecta dos vértices si una sentencia puede ser ejecutada inmediatamente después de la primera) y N es el número de nodos del grafo correspondientes a sentencias del programa.

La ecuación (1) indica que primero se realiza un conteo de todas las aristas del grafo de nodos, luego se cuenta la cantidad de nodos totales, finalmente se sustrae la cantidad de nodos obtenida a la cantidad de aristas y se suma 2.

$$CC = P + 1 \text{ (2)}$$

Otra forma es realizar un conteo de los nodos predicados y sumar una unidad, así como se observa en (2). P es el número de nodos predicados (Aquellos nodos cuyo grado de salida es dos. Normalmente representados en un diagrama de flujo por condiciones booleanas).

$$CC = R + 1 \text{ (3)}$$

Finalmente, se identifican las regiones cerradas del grafo y se añade 1 para obtener el resultado. R es el número de regiones cerradas del grafo.

Para el contexto de esta investigación dicha métrica fue utilizada con el propósito de estimar la complejidad y fiabilidad de los algoritmos KNN y FP-GROWTH una vez fueron integrados en el Sistema de Información WEB.

Por último, el SEI (1997) plantea una escala de riesgos para clasificar programas a partir de la complejidad condicional obtenida. La Tabla 1 presenta dicha escala.

**Tabla 1.** Evaluación Complejidad Condicional

Complejidad	Evaluación de Riesgo Condicional
1-10	Programa simple sin mucho riesgo
11-20	Programa complejo con riesgo moderado
21-50	Programa muy complejo de alto riesgo
50 +	Programa no testeable de muy alto riesgo

Fuente: Adaptado de SEI (1997)

## G. Data Mart

Según Castillo y Palomino (2012), es una base de datos que se especializa en el almacenamiento de datos de un área de negocio específica. El Data Mart cuenta con una estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de una unidad de negocio específica. Un Data Mart puede alimentarse desde los datos de un Data Warehouse o integrar por sí mismo un compendio de distintas fuentes de información.

Para el contexto de la presente investigación el Data Mart es quien almacena todos los datos relacionados a las ventas asociadas a los clientes del restaurante, por tanto otorga características de homogeneidad, no volatilidad e integridad.

La Tabla 2, presenta las principales diferencias entre un Data Warehouse y un Data Mart:

**Tabla 2.** Comparación entre Data Warehouse y Data Mart

	Data Warehouse	Data Mart
Alcance	Construido para satisfacer las necesidades de información de toda la organización	Construido para satisfacer las necesidades de información de un área de negocios específica
Objetivo	Diseñado para optimizar la integración y la administración de los datos fuente	Diseñado para optimizar la entrega de información de soporte de decisiones
Característica de los datos	Administra grandes cantidades de datos históricos a nivel atómico	Se concentra en administrar resúmenes y/o datos totalizados.
Pertenencia	Pertenece a toda la organización	Pertenece al área de negocio a la cual está orientado

Administración	Es administrado por la unidad de sistema de la organización	Es administrado por el personal de sistema de la unidad propietaria del Data Mart.
----------------	---	--

Fuente: Adaptado de Castillo y Palomino (2012)

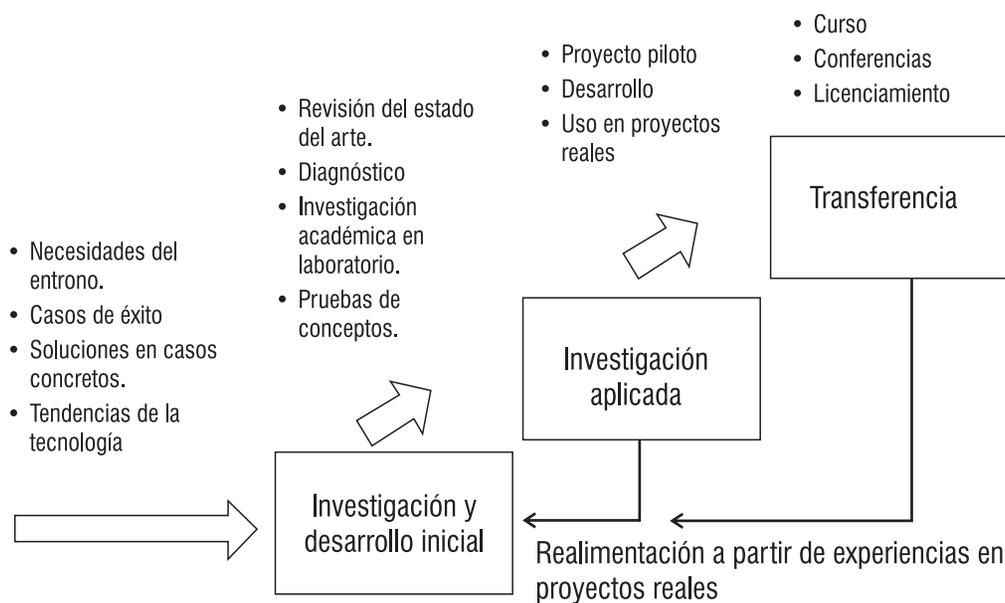
### III. METODOLOGÍA

Marcos (2002) comenta que la investigación en ingeniería difiere significativamente del proceso investigativo tradicional, tanto en objeto de estudio como en método, esto porque las ciencias se ocupan de investigar objetos y fenómenos existentes de diversa naturaleza, mientras que la ingeniería, específicamente la ingeniería de software, basa sus estudios en “cómo hacer” nuevos objetos. A pesar de ello, dicho autor plantea una similitud entre los estudios en el ámbito de la ingeniería de software y las ciencias tradicionales; por tanto, propone inicialmente que hay tres tipos de dominios asociados a la Ingeniería de Software (IS): Las ciencias de la IS, que hacen referencia a objetos de estudio que habitan en la mente de los investigadores y que precisamente con la investigación se busca su materialización o creación. Por otro lado, existe un dominio o Ciencias del Software, que consisten en investigar el software y las técnicas usadas para crearlos. Paralelamente, el autor comenta que alrededor de ambas ciencias mencionadas puede coexistir una tercera categoría, denominada Ciencias de los Sistemas de Información, que se ocupan de estudiar “cómo implantar y usar, tanto el software, como las técnicas y modelos que permiten crearlo” (Marcos, 2002, p.4).

El contexto de la investigación que aquí se plantea, consistió en la elaboración de un producto software alrededor de la gestión de relaciones con los clientes y en consecuencia una investigación en el ámbito de las ciencias de la IS según Marcos (2002). Este último propone que los métodos a

usar son de naturaleza cualitativa y creativa, esto en el caso efectivo de problemas puramente de ingeniería, porque como bien lo menciona, en dichos casos no es posible aplicar métodos empíricos, ya que éstos requieren de técnicas experimentales con un claro conocimiento de la realidad, lo cual implica la incorporación de importantes componentes de creatividad que conduzcan entonces a la producción de una realidad en forma de software. Adicionalmente, comenta que pueden ser cualitativos por la naturaleza misma de los procesos de IS, tales como impacto cultural de la adopción de nuevos paradigmas, incorporación de enfoques de trabajo colaborativo e iniciativas de trabajo en equipo durante los procesos de desarrollo de software.

Por otro lado, Arboleda y Chavarriaga (2004) comentan que la investigación en ingeniería supone una acuciosa reflexión permanente, según ellos porque generalmente actividades relacionadas con la ingeniería de software se confunden fácilmente con “cualquier tipo” de proceso inherente al desarrollo de software. Sin embargo, dichos autores proponen con base en el Software Engineer Institute (SEI) y autores como Martin y McClure la estrategia de investigación sintetizada en la Fig. 2.



**Figura 2:** Estrategia de investigación

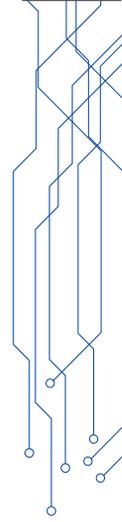
Fuente: Elaborado a partir de Arboleda y Chavarriaga (2004)

El proceso investigativo transita por una primera fase que incluye la revisión documental y revisión del estado del arte, así como la materialización de pruebas en laboratorio o en el contexto del software y la elaboración de prototipos funcionales; la segunda fase hace referencia a la puesta en marcha de proyectos reales que posibiliten realimentar los resultados de la fase inicial. Finalmente se agrega una fase de transferencia que consiste en amplificar los resultados, técnicas y conocimientos a comunidades externas que también puedan realimentar el proceso investigativo.

Con base en lo anterior, la investigación que aquí se expone, combinó la utilización de SCRUM como metodología de desarrollo de software, y enfoques exploratorios y descriptivos transversales a dicha metodología. SCRUM posibilitó la construcción del producto software por medio de iteraciones que se desarrollaron bajo el calendario del trabajo de grado citado

anteriormente<sup>28</sup>, cada iteración tuvo como resultado incrementos en la funcionalidad del sistema WEB, en la documentación asociada y demás artefactos inherentes al proceso de desarrollo de software. Además, los enfoques exploratorios y descriptivos permitieron, en primera medida, indagar por el estado actual de las aplicaciones en el contexto de CRM, revisar el material bibliográfico referente a los conceptos que sirvieron como fundamento de la investigación, y materializar prototipos incrementales en una atmósfera de continua creatividad; En el mismo sentido, el enfoque descriptivo posibilitó la validación de requisitos en la dimensión del software, ya que al describir el software y sus componentes por medio de artefactos, guías, flujos de procesos y documentos, fue posible transformar las ideas de los investigadores en un producto software.

<sup>28</sup> Disponible en/Avaible in/ Disponibilizado em: <http://opac.univalle.edu.co/cgi-olimp/?infile=details.glu&loid=943605&rs=4342312&hitno=6>

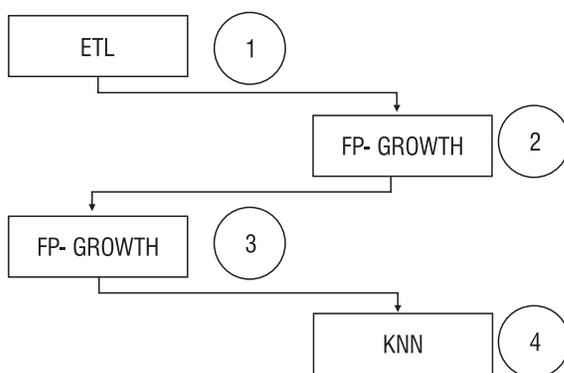


Finalmente, la publicación de un repositorio<sup>29</sup> que contiene la documentación y la integración de los algoritmos KNN y FPGROWTH, así como la presentación de este referente coinciden con lo expuesto en Arboleda y Chavarriaga (2004) en cuanto a la transferencia se refiere.

#### IV. RESULTADOS

La integración de los dos algoritmos fue posible gracias al desarrollo de un sistema Web para un establecimiento de comercio dedicado a la venta de comidas rápidas y a la carta. Para dicho desarrollo se usó el framework Codeigniter 3.0 como marco de trabajo para PHP y Bootstrap 3.3.7 para los componentes del frontend; adicionalmente, el sistema cuenta con dos bases de datos, una transaccional y otra no transaccional (Data Mark), la primera para soportar la dinámica del negocio y la segunda como insumo para realizar análisis de datos después de procesos de extracción, transferencia y carga (ETL) desde la base de datos transaccional.

Por otra parte, la Fig.3 sintetiza el orden de ejecución de los algoritmos FP-Growth y KNN una vez fueron integrados.



**Figura 3:** Integración de los algoritmos  
Fuente: De los autores

1. En primer lugar, se lleva a cabo el proceso de extracción, transformación y carga de datos en el Data Mart a partir de los datos registrados en la base de datos transaccional del sistema. Este proceso se encarga de mantener actualizados los datos del Data Mart ya que sobre éste se ejecutan los algoritmos FP-Growth y KNN.
2. Si se consultan por clientes frecuentes se ejecuta el algoritmo FP-Growth con el fin de obtener dichos clientes a partir de la cantidad de ventas registradas a cada uno mientras cumplan con el umbral definido para ello. Por otra parte, si se busca analizar un cliente específico, el flujo del proceso descarta el paso número 2.
3. Una vez se cuente con el cliente específico objeto de análisis, o los clientes frecuentes resultantes en el paso número 2, se ejecuta el algoritmo FP-Growth para obtener los productos más vendidos a cada uno de ellos, esto por medio de la poda al árbol de ocurrencias resultante de analizar las ventas por cada cliente.
4. Finalmente, se ejecuta el algoritmo KNN para determinar el conjunto de productos recomendados al cliente a partir de la similitud entre la composición de estos productos y la composición de los productos obtenidos en el punto 3. La composición de cada producto está determinada por los ingredientes que contiene.

Con el ánimo de detallar la implementación de los algoritmos se publicó un repositorio<sup>29</sup> que explica dicho proceso, además, se incluyeron en el, todos los artefactos productos de la metodología de desarrollo del sistema de información WEB.

Una vez los algoritmos fueron integrados y teniendo en cuenta el concepto de complejidad

<sup>29</sup> Disponible en: <https://github.com/RoyerUnivalle/KNN-FP-GROWTH>

condicional, se realizaron los respectivos cálculos. Los variables y valores que se usaron se relacionan en la Tabla 3.

**Tabla 3.**

Variabes para el cálculo de la complejidad según Cardacci (2016)

Variable/Algoritmo	FP-GROWTH	KNN
A	40	40
N	31	31
P	16	16
R	1	1

Fuente: De los autores

La Tabla 4 presenta los resultados una vez se reemplazaron los valores de la Tabla 3 en cada una de las ecuaciones presentadas en apartados anteriores.

**Tabla 4.**

Complejidad de cada algoritmo usando tres ecuaciones

Algoritmo/Ecuación	(1)	(2)	(3)
FP-GROWTH	11	17	2
KNN	11	17	2

Fuente: De los autores

Con base en los cálculos realizados, se observa que en la mayoría de los casos la complejidad condicional fue superior a 10 e inferior a 20, lo cual, según la Tabla 1, los algoritmos corresponden a un programa complejo con riesgo moderado, a excepción del cálculo de la complejidad condicional con (3), en la cual, para los dos casos, el resultado fue inferior a 10, lo cual indica que es un programa simple sin mucho riesgo.

Finalmente, se presentan los resultados obtenidos a partir de las pruebas realizadas al sistema para cuatro escenarios específicos sobre

recomendaciones de ventas. Los escenarios se desarrollaron con un total de cincuenta y cien tuplas de clientes, que a su vez, están asociados a diferentes ventas. Se tuvieron en cuenta los siguientes criterios para la recomendación:

1. *Fecha de inicio*: Fecha a partir de la cual se tendrán en cuenta las ventas registradas en el sistema. Este criterio debe ser nulo si el criterio Cliente toma valor.
2. *Fecha final*: Fecha hasta la que se tendrán en cuenta las ventas registradas en el sistema. Este criterio debe ser nulo si el criterio Cliente toma valor.
3. *Frecuencia mínima de ventas*: Número mínimo de ventas registradas a un cliente para considerarlo como cliente frecuente y también número mínimo de ventas de un producto a un cliente para determinarlo como producto frecuente.
4. *Porcentaje de recomendación mínimo*: Es el valor mínimo de similitud entre el producto a recomendar y el producto de referencia (producto comprado frecuentemente por el cliente) para considerarlo como producto recomendado.
5. *Cliente*: Es el código del cliente para el que se desea realizar la recomendación de productos. Este criterio debe ser nulo si los criterios Fecha inicial y Fecha final toman valor.

La Tabla 5 presenta las características de cada uno de los escenarios desarrollados, además relaciona el tiempo de ejecución de cada uno de ellos gracias a la herramienta Profiling de CodeIgniter.

**Tabla 5.** Escenarios de prueba

Escenario	Propósito	Resultado Esperado	Resultado Obtenido	Tiempo Ejecución (s)
1	Obtener productos recomendados para un cliente específico recién registrado en el sistema a partir de los criterios: Frecuencia mínima de ventas, porcentaje de recomendación mínimo e identificación del cliente.	Listado de productos recomendados al cliente.	No se generaron recomendaciones.	0.2437
2	Obtener productos recomendados a vender a los clientes más frecuentes en un lapso determinado a partir de los criterios: Fecha de inicio, fecha final, frecuencia mínima de ventas y porcentaje de recomendación mínimo. Escenarios con cincuenta tuplas.	Listado de clientes frecuentes y productos recomendados por cada cliente.	Listado de clientes frecuentes y productos recomendados por cada cliente.	0.2378
3	Obtener productos recomendados a vender a un cliente específico a partir de los criterios: Frecuencia mínima de ventas, porcentaje de recomendación mínimo y cliente. Escenarios con cincuenta tuplas.	Listado de productos recomendados al cliente.	Listado de productos recomendados al cliente.	0.2426
4	Obtener productos recomendados a vender a un cliente específico habiendo duplicado el volumen de datos de las ventas, a partir de los criterios: Frecuencia mínima de ventas, porcentaje de recomendación mínimo y cliente. Escenarios con cien tuplas.	Listado de productos recomendados al cliente.	Listado de productos recomendados al cliente.	0.2484

Fuente: De los autores

El escenario uno da cuenta del problema de inicio en frío de la integración de los algoritmos, dicho de otra manera, no basta con realizar la integración de los algoritmos sino se cuenta con los datos o tuplas necesarias para el respectivo procesamiento.

En los escenarios dos y tres se recomendaron los productos esperados y también productos adicionales debido a que los algoritmos FP-Growth y KNN determinaron alto grado de

similitud entre la composición de los productos recomendados y la de los productos frecuentes.

Finalmente, el escenario cuatro entrega los resultados esperados y permite inferir que el costo computacional de los algoritmos es de naturaleza polinomial, esto porque depende de los criterios indexados para la búsqueda y además no muestra diferencias mayores entre la ejecución de los algoritmos con cincuenta y cien tuplas respectivamente.

## V. CONCLUSIONES

El uso de algoritmos para la generación de recomendaciones supone problemas de inicios en frío o Cold Start, en consecuencia, es preciso mencionar que a pesar que los algoritmos integrados son eficientes computacionalmente según el cálculo de su complejidad condicional, resta por abordar dicho problema con el fin de posibilitar recomendaciones más precisas, incluso al inicio de contactos con nuevos clientes, lo cual se traduciría entonces en una gestión de relaciones eficiente con clientes aún no fidelizados.

Las técnicas de asociación son determinantes para obtener un análisis preciso basado en grandes volúmenes de datos. Para el contexto del problema abordado en este trabajo, se permitió identificar de manera rápida y fácil los comportamientos presentes en las ventas realizadas a los clientes.

La combinación de una técnica de asociación, como el algoritmo FP-Growth y una técnica de clasificación, como el algoritmo KNN, resultó eficiente en cuestiones de rendimiento y consistente en relación con los resultados generados desde diversos escenarios, estos últimos permitieron observar una relación polinomial en relación al tiempo de ejecución, incluso con diferentes criterios y número de tuplas.

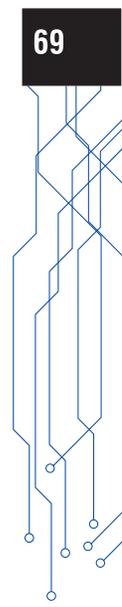
Finalmente, este trabajo se traduce en un referente, en la medida que da cuenta de los procesos y técnicas utilizadas para extender el uso de sistemas informáticos, a sistemas en el ámbito de la gestión de relaciones con los clientes por medio de la aplicación de técnicas a la luz de la minería de datos, lo cual permitirá a compañías que repliquen dichas técnicas el logro de ventajas competitivas por medio del conocimiento de patrones y comportamientos que habitan en sus consumidores.

## VI. AGRADECIMIENTOS

Agradecemos sinceramente a los profesores que enriquecieron este artículo con sus conceptos, críticas y opiniones. Al estudiante de la Maestría en Ingeniería y profesor catedrático, Ing. Carlos Andrés Delgado Saavedra, de igual modo al profesor Joshua David Triana, profesor del curso de Inteligencia Artificial. Ambos de la Universidad del Valle; sin ellos este artículo no se hubiera materializado.

## REFERENCIAS

- Arboleda, H., y Chavarriaga, J. (2004). Modelo de Investigación en Ingeniería del Software: Una propuesta de investigación tecnológica. MIFISIS. Disponible en: <http://web.emn.fr/x-info/harbol07/MIFISIS2004.pdf>
- Betarte, L, Machado, R y Molina, V (2006). PGMUSICA: sistema de recomendación de música basado en filtrado colaborativo. Tesis de grado. Universidad de la República. Facultad de Ingeniería. Instituto de Computación. Uruguay. Disponible en: <https://www.fing.edu.uy/inco/grupos/pln/prygrado/InformePGMusica.pdf>
- Brusco, M. (2011). La influencia del uso de las TICS en la Internacionalización de las PYMES. Universidad Pompeu Fabra.
- Cano, J. (2007). Business Intelligence: competir con información. ESADE Business School, Universidad Ramon Llull, España.
- Cardacci, D. (2016). Refactorización de código y consideraciones sobre la complejidad ciclométrica. Universidad del Cema, Buenos Aires, Argentina. Disponible en: <http://www.ucema>



edu.ar/publicaciones/download/documentos/592.pdf

Castañeda, J. y Rodríguez, M. (2003). La minería de datos como herramienta de marketing: Delimitación y medidas de evaluación del resultado. Gestión científica empresarial. ISBN 84-9745-051-5.

Castillo, J. y Palomino, L. (2012). Implementación de un Data Mart como una solución de Inteligencia de Negocios para el área de logística de T-Impulso. Revista de investigación de sistemas e informática, Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos. Disponible en: <http://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/view/5713/4944>

Chen, I. y Popovich, K. (2003). Understanding customer relationship management (CRM): People, process and technology. Business Process Management Journal. DOI:10.1108/14637150310496758

García, C. y Gómez, I. (2015). Algoritmos de aprendizaje KNN y KMEANS. Universidad Carlos III, Madrid. Disponible en: <http://www.it.uc3m.es/jvillena/irc/practicas/08-09/06.pdf>

Marcos, E. (2002). Investigación en Ingeniería del Software vs Desarrollo Software, MIFISIS. pp. 136-149.

Marina, S. (2014). Suite de algoritmos de recomendación en aplicaciones reales. Tesis de grado. Universidad Autónoma de Madrid. Departamento de ingeniería informática. Madrid. Disponible en: <https://repositorio.uam.es/handle/10486/660903>

Meroño, A. (2004). Tecnologías de información y gestión del conocimiento: integración en un sistema. Revista Economía Industrial. 107-116. Disponible en: <https://dialnet.unirioja.es/revista/445/A/2004>

Pinho, J. (2010). Métodos de clasificación basados en asociación aplicados a sistemas de recomendación. Departamento de informática y automática, Universidad de Salamanca. Disponible en: <http://hdl.handle.net/10366/83342>

Rosado, A. y Rico, D. (2010). Inteligencia de Negocios: Estado del arte. Scientia Et Technica. Vol XVI, núm. 44. Disponible en: <http://www.redalyc.org/pdf/849/84917316060.pdf>

SEI (1997). Software Technology Reference. Página 147. Disponible en: [https://resources.sei.cmu.edu/asset\\_files/Handbook/1997\\_002\\_001\\_16523.pdf](https://resources.sei.cmu.edu/asset_files/Handbook/1997_002_001_16523.pdf)

Virsedá, F. y Román, J. (s.f). Minería de datos y aplicaciones. Universidad Carlos III.