

Proyecto – Allers Group

Integrantes:

* David Erazo
* Daniela Llano
* Manuel Alejandro Quintero
* Juan Esteban Gallo

Método de la Ingeniería

# Paso 1. Identificación del problema

Allers Group, una empresa especializada en la distribución y ventas de productos para la salud busca aumentar sus ventas. La empresa considera que tienen datos históricos relevantes almacenados y creen que existe información importante dentro de estos que permitiría a la empresa alcanzar la meta establecida.

*Se identifican de manera adecuada las necesidades que se presentan en el problema propuesto, así como sus síntomas y condiciones por los que debe ser resuelto.*

*Identificación de sus síntomas y necesidades:*

* Se han recolectado una gran cantidad de datos que hoy en día están en desuso
* Ingresan una cantidad de datos diaria de alrededor 350 a 600 datos
* Se necesita lograr una evolución de beneficio a la empresa con los datos que se tiene en poder.
* Se quiere que el cliente conozca la variedad de productos que ofrece la empresa y cuales podrían ajustarse también a sus necesidades.
* Se quiere que los clientes adquieran una mayor cantidad de los artículos ofrecidos por la empresa.

*Descripción del problema:*

Allers Group requiere una solución a la problemática que tienen con el desuso de datos, permitiendole a la empresa obtener el mayor beneficio mediante análisis de datos para que esta pueda estructurar diferentes áreas, como el área de mercadeo y el área de distribución, mediante los resultados

El sistema debe estar en la capacidad de:

|  |  |
| --- | --- |
| **Identificación del requerimiento:** | RF01 |
| **Nombre del Requerimiento:** | Uso de datos de la empresa |
| **Características:** | La solución estará en la capacidad de utilizar los datos recibidos por la base de datos de la empresa. |
| **Descripción del requerimiento:** | La solución podrá leer los datos obtenidos por la empresa Allers Group y estos se pondrán a disposición de la solución para el proceso que se necesite. |
| **Prioridad del requerimiento:**  Alta | |

|  |  |
| --- | --- |
| **Identificación del requerimiento:** | RF02 |
| **Nombre del Requerimiento:** | Actualizar la información |
| **Características:** | La solución estará en la capacidad de actualizar la información que se recibe por parte de la base de datos si así se necesita. |
| **Descripción del requerimiento:** | La solución podrá leer los datos nuevos que sean obtenidos por la empresa Allers Group y estos se pondrán a disposición de la solución para el proceso que se necesite. |
| **Prioridad del requerimiento:**  Alta | |

|  |  |
| --- | --- |
| **Identificación del requerimiento:** | RF03 |
| **Nombre del Requerimiento:** | Generar predicciones de ventas en base a los artículos |
| **Características:** | La solución estará en la capacidad de realizar una predicción de ventas en base a los datos ya recibidos por la empresa Allers Group generando una salida con los resultados. |
| **Descripción del requerimiento:** | La solución realizará procesos que permitan obtener información acerca de los artículos en las ventas para predecir las posibles promociones que a los clientes les interesaría comprar. Estos resultados deben permitirle al usuario tomar una decisión acerca de cuál o cuáles son los productos que más le benefician y que recomendaciones puede ofrecer a los clientes. |
| **Prioridad del requerimiento:**  Alta | |

|  |  |
| --- | --- |
| **Identificación del requerimiento:** | RF04 |
| **Nombre del Requerimiento:** | Obtener similitudes entre clientes |
| **Características:** | La solución estará en la capacidad de obtener del análisis de datos información que permita generar similitud de ventas entre los clientes |
| **Descripción del requerimiento:** | La solución analizara los datos de una manera más detallada para poder obtener similitudes entre clientes en base a sus ventas y así proveerle información más eficiente al usuario sobre los patrones que cumplen sus clientes en base a otros. |
| **Prioridad del requerimiento:**  Alta | |

|  |  |
| --- | --- |
| **Identificación del requerimiento:** | RF05 |
| **Nombre del Requerimiento:** | Obtener información superficial sobre las ventas. |
| **Características:** | La solución estará en la capacidad de proveerle al usuario información superficial que le permita identificar características sobre sus ventas y que además le permita analizar con mayor claridad la información que la solución le ofrece. |
| **Descripción del requerimiento:** | La solución obtendrá información superficial tales como: porcentaje de aparición de algunos artículos en las ventas y artículos frecuentes por categoría de clientes, para poder darle un mayor sentido a la información suministrada por la solución y permitirle al usuario tener una idea más global sobre sus datos. |
| **Prioridad del requerimiento:**  Media | |

# Paso 2.  Recopilación de Información

Con el propósito de realizar una correcta recopilación de información, se propone abarcar esta fase por medio de 3 factores principales:

## 2.1 Marco Teórico

Se hace una búsqueda de los términos relacionados con el problema planteado. Esta búsqueda se realiza en fuentes reconocidas y confiables para reconocer que conceptos hacen parte del problema.

* Data mining:

*Data Mining* (minería de datos) es el proceso de extracción de información significativa de grandes bases de datos, información que revela inteligencia del negocio, a través de factores ocultos, tendencias y correlaciones para permitir al usuario realizar predicciones que resuelven problemas del negocio proporcionando una ventaja competitiva.

* Big Data:

Big data (en español, grandes datos o grandes volúmenes de datos) es un término evolutivo que describe cualquier cantidad voluminosa de datos estructurados, semiestructurados y no estructurados [que tienen el potencial de ser extraídos para obtener información](https://searchdatacenter.techtarget.com/es/cronica/Big-data-una-espada-de-doble-filo-en-la-ciberestrategia-de-este-CIO).

* Windows Forms

Las características principales de la programación de Windows Forms y cómo puede usar Windows Forms para compilar smart clients que satisfagan las necesidades actuales de las empresas y usuarios.



* C#

es un lenguaje de programación orientado a objetos desarrollado y estandarizado por Microsoft como parte de su plataforma .NET, que después fue aprobado como un estándar por la ECMA (ECMA-334) e ISO (ISO/IEC 23270). C# es uno de los lenguajes de programación diseñados para la infraestructura de lenguaje común

* Visual Studio

Microsoft Visual Studio es un [entorno de desarrollo integrado](https://es.wikipedia.org/wiki/Entorno_de_desarrollo_integrado) (IDE, por sus siglas en inglés) para sistemas operativos [Windows](https://es.wikipedia.org/wiki/Windows). Soporta múltiples lenguajes de programación, tales como [C++](https://es.wikipedia.org/wiki/Visual_C%2B%2B), [C#](https://es.wikipedia.org/wiki/C_sharp), [Visual Basic .NET](https://es.wikipedia.org/wiki/Visual_Basic_.NET), [F#](https://es.wikipedia.org/wiki/F_Sharp), [Java](https://es.wikipedia.org/wiki/Java_(lenguaje_de_programaci%C3%B3n)), [Python](https://es.wikipedia.org/wiki/Python), [Ruby](https://es.wikipedia.org/wiki/Ruby) y [PHP](https://es.wikipedia.org/wiki/PHP), al igual que entornos de desarrollo web, como [ASP.NET](https://es.wikipedia.org/wiki/ASP.NET) MVC, Django, etc., a lo cual hay que sumarle las nuevas capacidades online bajo Windows Azure en forma del editor Monaco.

* Base de datos

Una base de datos de un SI (Sistema de Información) es la representación integrada de los conjuntos de entidades instancia correspondientes a las diferentes entidades tipo del SI y de sus interrelaciones. Esta representación informática (o conjunto estructurado de datos) debe poder ser utilizada de forma compartida por muchos usuarios de distintos tipos.

En otras palabras, una base de datos es un conjunto estructurado de datos que representa entidades y sus interrelaciones. La representación será única e integrada, a pesar de que debe permitir utilizaciones varias y simultáneas.



* Microsoft.NET

Es un [framework](https://es.wikipedia.org/wiki/Framework) de [Microsoft](https://es.wikipedia.org/wiki/Microsoft) que hace un énfasis en la transparencia de [redes](https://es.wikipedia.org/wiki/Red_de_computadoras), con independencia de plataforma de [hardware](https://es.wikipedia.org/wiki/Hardware) y que permite un rápido desarrollo de [aplicaciones](https://es.wikipedia.org/wiki/Aplicaci%C3%B3n_inform%C3%A1tica).

## 2.2 Entrevista con el cliente

Se realizó una visita por parte del personal de la empresa Allers, en donde se logró recolectar información específica del estado en el que opera la empresa actualmente:

* La empresa Allers cuenta con una base de datos en SQL Server para almacenar sus datos generados a medida que transcurre el tiempo.
* La empresa cuenta con equipos de buena capacidad para el manejo de grandes cantidades de datos.
* La solución a plantear va dirigida a los directivos de Allers para la toma de decisiones.
* La empresa cuenta con un inventario activo de aproximadamente 15000 productos activos.
* Se enfocan en la distribución de insumos hospitalarios, teniendo como objetivo cubrir la mayor parte del territorio nacional.
* Realizan sus ventas basándose en promedios históricos.
* Cuenta con personal que realiza análisis según las estadísticas que arrojan los datos que tiene la empresa

## 2.3 Antecedentes

Se realiza una investigación de situaciones y proyectos los cuales sean similares al problema planteado. Esta investigación se realiza teniendo en cuenta proyectos en donde se abarque una solución por medio de Big Data o Data mining.

* Proyecto (Big data para acabar con los atascos en Nueva York):

La ciudad de Nueva York, como muchas ciudades en el mundo, tiene presente una gran problemática en cuanto al tráfico urbano, muchas de las soluciones que se han tratado de abordar son el uso de algoritmos para sincronizar los semáforos, de tal manera que dé prioridad a vías principales. A pesar de esto, solo se ha logrado disminuir un poco los atascos. Según la investigadora del MIT en congestión de redes urbanas, Carolina Osorio, plantea una solución a esta situación para un futuro muy cercano, la cual se trata de utilizar la big data, principalmente extrayendo información de las agendas de las personas, con el propósito de crear algoritmos que permita realizar una organización en las vías, recomendando cuándo realizar una actividad de la vida diaria de la persona. Claramente este proceso se llevaría a cabo gracias a algoritmos capaces de poder tomar decisiones frente a las diferentes prioridades de las agendas de las personas.

* Proyecto (Big data para salvar animales):

El mundo está sufriendo una [sexta ola de extinción en masa](http://wwf.panda.org/about_our_earth/biodiversity/biodiversity/) en la que grandes porcentajes de las especies del planeta están desapareciendo. Un equipo de científicos del Snow Leopard Trust y de la National Conservation Foundation están [usando los datos](http://www.huffingtonpost.com/2015/01/12/india-wild-tigers-big-data_n_6458386.html) para saber cuándo, dónde y cómo los [cazadores furtivos intentarán cazar los tigres salvajes en la India.](http://dataconomy.com/2015/01/how-big-data-can-help-in-the-fight-to-save-indias-tigers/) La fauna de la India alberga la mayoría de la población de tigres salvajes del planeta. Por esta razón están analizando cantidades increíbles de datos (Big data) para intentar interceptar los cazadores furtivos antes de que puedan llegar a su objetivo.

* Proyecto (Big data para detectar y prevenir ciberataques):

El análisis de big data se está empleando de forma intensiva en la mejora de la seguridad informática, como lo está solucionando el sistema de IA creado en el MIT el cual hasta la fecha a logrado predecir el 85% de los ciberataques.

* Proyecto (Big data para la recomendación de libros - Amazon):

La empresa Amazon realiza un análisis de big data sobre los datos que recolectan por sus ventas de libros, gracias a este análisis logran realizar sugerencias de otros libros por medio de un algoritmo. Un análisis en su base de datos permite identificar que un cliente que compra un libro, por lo general su comportamiento está basado en comprar más libros de este mismo tipo, razón por la cual el sistema sugiere otros libros similares basándose en los datos recolectados de miles de clientes.

# Paso 3. Búsqueda de Soluciones Creativas.

Identificación o empatía: Para este caso se decidió llegar a la búsqueda de ideas creativas mediante la técnica de identificación o empatía que consiste en dar la vuelta a una situación y ponerse en la piel de otra persona. Tomando como idea principal ponerse en la situación de nuestro proveedor de información, cuando busca la opción de su manejo de datos de la mejor manera posible. Se llegó a la decisión de usar esta alternativa por consenso del grupo, llegando a que sería la mejor opción para buscar una solución creativa.

**Ideas para estrategias a implementar:**

Contexto: Como se cuenta con una gran cantidad de información acerca de los clientes de la compañía, se realizará un análisis riguroso de sus comportamientos a lo largo del tiempo. Dicho lo anterior, la información de compras y ventas de cada uno serían de gran importancia en el análisis planteado, ya que así se obtendrían las preferencias que tiene el cliente; ya sea de los productos que más frecuenta o los más consultados por clientes similares a él. Esto con el objetivo de sugerir los productos y lograr una conformidad mayor tanto como para el cliente como para la empresa. Por estas razones se sugiere la implementación de un software el cual esté en capacidad de proporcionar una predicción lo más acertada posible, teniendo como alternativas de desarrollo las siguientes opciones:

Alternativa 1: Con el fin de realizar una búsqueda de reglas de asociación, se propone implementar la estrategia “BruteForce” sobre los datos que se tienen en poder. Este algoritmo consiste en enumerar sistemáticamente todos los posibles candidatos, con el fin de revisar si dicho candidato satisface la solución del problema.

Alternativa 2: La implementación de una estrategia “A-Priori” sobre los datos que se tienen en poder~~,~~ conlleva a un buen recorrido de estos, teniendo gran eficacia en el momento de obtener las reglas de asociación, debido a que este método consiste en identificar los ítems frecuentes en la base de datos e ir extendiéndolos a conjuntos de mayor tamaño, siempre y cuando esos conjuntos de datos aparezcan suficientemente seguidos.

Alternativa 3: Una posible implementación para la solución del problema~~,~~ es utilizar la estrategia del algoritmo “F(K-1)XF1”, el cual consiste en combinar los k-1 items frecuentes con los ítems frecuentes que cumplan con la condición del supporCount, y una vez realizado este proceso de combinación se obtendrán los ítems más frecuentes para toda la solución requerida.

**Ideas para el uso de infraestructuras:**

Alternativa 1: Para el manejo de los datos proporcionados, basado en el sistema de manejo de información Hadoop, el cual se tomó en cuenta como una gran alternativa en la minería de datos, el programa estará en capacidad de manejar grandes volúmenes de datos de manera simultánea en varios ordenadores. Dando esto la posibilidad de tener informes necesarios de la manera más eficiente posible. En síntesis, para ayudar en la utilización de los datos para informes de posibles ventas futuras lo más rápido posible del mismo modo que en las consultas propuestas. Cabe resaltar que esta idea se pensó dando como prioridad al tiempo de una consulta y el paralelismo de trabajo en cualquier momento.

Alternativa 2: El cliente, al manejar diversos flujos de datos de pedidos de diferentes partes, requiere que dicho flujo de datos se dé a conocer en tiempo real por un programa de usuario Master, el cual le permita conocer en cualquier momento el manejo de su empresa en cuanto a reportes y reconocimientos por día, dando importancia en cualquier opción de efectividad además de dar comodidad en cuanto la accesibilidad de información se requiera.

Alternativa 3: La accesibilidad web podría unificar toda la información y procesarla para obtener resultados, como, por ejemplo, conocer cuáles son las posibles compras de un cliente, a partir de su comportamiento ya previamente registrado. Ahora bien, esta alternativa se basa más en el objetivo de tener un mejor manejo de datos mediante un programa con accesibilidad web.

Alternativa 4: Realizar una aplicación de escritorio usando los datos proporcionados por el cliente, mediante el análisis de clústeres, de correlación o de regresión, así como con la clasificación de datos, el programa estará en la capacidad de generar informes y reportes. Cabe resaltar, que dichos reportes e informes, se verían principalmente como un análisis a futuro con respecto a los datos recopilados en los últimos seis meses clasificados entre empresas y productos relacionados que pueden ser una posible compra.

Alternativa 5: Realizar una aplicación de escritorio la cual brinde diferentes servicios tales como reportes, gráficos, predicciones y consejos de ofertas. Partiendo del análisis de la base de datos suministrada por la empresa, el programa estaría en la capacidad de analizar los respectivos datos con gran exactitud, para así garantizar que la empresa logre el objetivo de incrementar sus ventas.

# Paso 4. Transición de las Ideas a los Diseños Preliminares

Lo primero que se hará en este paso es descartar la solución o soluciones que se considere que no son factibles. En este caso se descarta la *Alternativa 1* para las ideas asignadas al uso de infraestructuras, debido a que esta plataforma tiene dificultades para adaptarse a .Net Framework y también porque requiere una infraestructura que no se posee.

A partir de una revisión cuidadosa de las otras alternativas, se determina lo siguiente:

*Ideas para estrategias a implementar:*

Alternativa 1:

BruteForce, es una estrategia sencilla de implementar y siempre que exista una solución al problema, esta estrategia la encontrará. Sin embargo, su coste de ejecución es proporcional al número de combinaciones que realice la estrategia, el cual es exponencialmente proporcional al tamaño de los datos, realizando un proceso muy lento en el caso cuando se comienza a trabajar con una gran cantidad de datos.

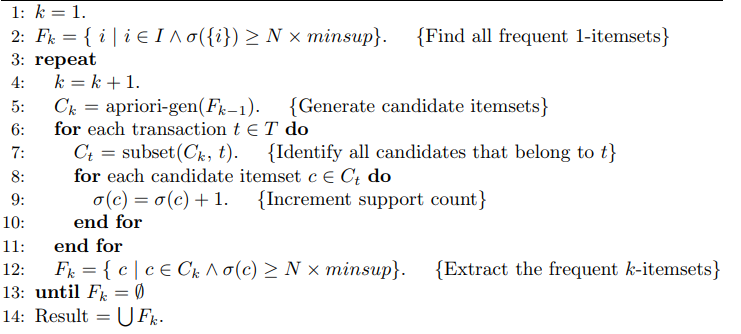
Cuenta con una complejidad temporal de

Alternativa 2:

A-Priori, es una estrategia usualmente usada en la minería de datos, que presenta una gran eficiencia para grandes volúmenes de datos, pero para ciertos datos de entrada esta estrategia presenta unos resultados intermedios que consumen una gran cantidad de recursos, es decir, de memoria.

El número total de iteraciones que requiere este algoritmo es de k(max)+1 donde k(max) es el tamaño máximo de los itemsets más frecuentes.

Pseudocódigo para la implementación del algoritmo A-priori específicamente para la generación del itemset más frecuente:



Alternativa 3:

F(K-1)xF1, es un método que se destaca por su gran velocidad en el conteo del support en sus operaciones, pero tiene el problema de que tiene que realizar 3 combinaciones en su proceso, realizando primero una combinación entre los ítemset más frecuentes en (k-1), luego una combinación para los itemsets más frecuentes que cumplan con la condición del supportCount y por último la combinación entre estos dos conjuntos.

 Cuenta con una complejidad temporal de

Ideas para el uso de infraestructuras:

Alternativa 2:

Pasando a las ideas de implementación que conllevan esta alternativa, se requiere una aplicación capaz de manejar flujos de datos a través de la sincronización de estos mismos mediante la web, esto se debería hacer de manera continua, además que, al tener un usuario master, este será el único que podrá acceder a todos los datos completos y generar los reporte. Por lo cual, si se necesita un reporte seguro y conciso, se ha de tener de la actualización de los últimos datos sincronizados por todos los usuarios.

Alternativa 3:

Esta alternativa consiste en la ayuda de la accesibilidad web para publicar la aplicación por lo cual desde cualquier computador mediante el acceso remoto de la página web se podría obtener los reportes necesarios.

Alternativa 4:

También pasando por los análisis propuestos previamente, en cuanto al análisis de clústeres, llamado en ocasiones Análisis de conglomerados, es una técnica que, como el Análisis de componentes principales, notaría una realidad cuya representación original es multidimensional y es imposible que se pueda ver en su estado puro los elementos de los cuales se quieren filtrar. Para así, luego obtener los reportes solicitados por el usuario. Ahora bien, pasando a la segunda alternativa de análisis de correlación o regresión, este consistiría en analizar la relación entre al menos dos variables. El resultado de dicho reporte mediante esta técnica incluiría la fuerza y el sentido de la relación que estamos analizando.

Alternativa 5:

El hecho de realizar predicciones implica que el programa requiera de un algoritmo el cual sea capaz de realizar diferentes operaciones sobre los datos, además como en este caso se estará trabajando con el concepto de Big Data es necesario la implementación de una correcta y eficiente estrategia para el recorrido sobre los datos.



# D:\David\Desktop\22.jpg

# Paso 5. Evaluación y selección de la mejor solución

Deben definirse los criterios que permitirán evaluar las alternativas de solución y con base en este resultado elegir la solución que mejor satisface las necesidades del problema planteado. Los criterios que escogimos en este caso son los que enumeramos a continuación. Al lado de cada uno se ha establecido un valor numérico con el objetivo de establecer un peso que indique cuáles de los valores posibles de cada criterio tienen más peso, es decir, cuanto más puntaje tenga será el adecuado para realizar la solución.

*Criterios asignados a ideas para estrategias a implementar:*

* *Criterio A.* Eficacia del algoritmo. Se requiere un algoritmo el cual sea capaz de realizar las diferentes actividades en el menor tiempo posible, brindando una gran optimización de tiempo en los diferentes procesos que se van a realizar.

El algoritmo para una cantidad de 10000 datos presenta una eficacia:

- [4] Muy alto (10 min o menos)

- [3] Alto (entre 15 y 20 min)

- [2] Medio (entre 20 y 35 min)

- [1] Bajo (35 min o más)

* *Criterio B.* Facilidad en implementación algorítmica. La alternativa cuenta con un nivel de dificultad que permita al desarrollador generar la solución de la mejor manera con la implementación de menor cantidad de recursos informáticos, para así mejorar el rendimiento del programa. {Entendiéndose recurso informático como plataformas, plug-in, repositorios, bases de datos etc.}
  + [1] Difícil. Se requiere que el desarrollador utilice una gran cantidad de recursos (mayor a 10 min) para el desarrollo de la solución.
  + [2] Neutral. Se requiere que el desarrollador utilice una cantidad suficiente ( entre 3 a 10) de recursos para desarrollar la solución.
  + [3] Fácil. No se requiere que el desarrollador utilice una cantidad mayor de recursos (menos de 3) para desarrollar la solución.

*Evaluación*

Según la evaluación con los criterios anteriores en las alternativas se obtiene la siguiente tabla:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Criterio A  Eficacia | Criterio B  Facilidad de implementación algorítmica | Total |
| Alternativa 1  Método BruteForce | Bajo (35 min o más)  [1] | Neutral  [2] | 3 |
| Alternativa 2  Método A priori | Muy alto (10 min o menos)  [4] | Fácil  [3] | 7 |
| Alternativa 3  Método F(K-1)xF1 | Alto (entre 15 min y 20 min)  [3] | Difícil  [1] | 4 |

*Selección*

En base a los criterios referenciados anteriormente, y la calificación que le dimos a cada una de las alternativas en base a estos criterios, tomamos la decisión de utilizar la alternativa número 2 (Alternativa 2) para la implementación de la estrategia del algoritmo, esto debido a la gran eficacia que tiene, y que cuenta con las mejores características para la resolución de este problema a pesar de que en ocasiones consuma una gran cantidad de recursos.

*Criterios asignados a ideas para el uso de infraestructuras:*

* *Criterio A.* Eficacia. Se requiere una plataforma la cual cuente con una herramienta de análisis lo suficientemente compleja para abarcar los requerimientos necesarios.

Presenta una herramienta con un nivel de eficacia:

- [4] Muy alto (Brinda más de 3 servicios)

- [3] Alto (Brinda entre 2 y 3 servicios)

- [2] Medio (Brinda entre 1 y 2 servicios)

- [1] Bajo (Solo actualiza la información)

* *Criterio B.* Accesibilidad. Se requiere una plataforma la cual cuente con un nivel de acceso adecuado para una empresa que necesita poder tener disponible su información para las personas indicadas.

*-* [3] Restringida (Puede acceder cualquier usuario que este permitido)

- [2] Única (Solo puede acceder el usuario que ingreso los datos)

- [1] Libre (Puede acceder cualquier persona)

* *Criterio C.* Coste de Ejecución. Se prefiere una plataforma que tenga el menor coste de ejecución, es decir, el menor tiempo de ejecución posible cuando se ingresan grandes cantidades de datos.

*-* [1] Grande (Conlleva más de 3 horas en ejecutar)

*-* [2] Intermedio (Conlleva entre 1 y 3 horas en ejecutar)

*-* [3] Bajo (Conlleva menos de una hora en ejecutar)

* *Criterio D.* Seguridad. Esta solución cuenta con un nivel de protección que permite mantener la privacidad de todos los datos de la compañía con la mayor privacidad posible.

- [1] Baja (Cualquier usuario puede acceder y modificar los datos)

- [2] Intermedia (Cualquier usuario permitido puede acceder y modificar los datos)

- [3] Alta (Sólo un usuario puede modificar la información)

*Evaluación*

Según la evaluación con los criterios anteriores en las alternativas se obtiene la siguiente tabla:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Criterio A  Eficacia | Criterio B  Accesibilidad | Criterio C  Coste de ejecución | Criterio D  Seguridad | Total |
| Alternativa 2 | Alto (Brinda entre 2 y 3 servicios)  [3] | Única  [2] | Intermedio (Conlleva entre 1 y 3 horas en ejecutar)  [2] | Alta (Sólo un usuario puede modificar la información)  [3] | 10 |
| Alternativa 3 | Bajo (Solo actualiza la información)  [1] | Libre  [1] | Bajo (Conlleva menos de una hora en ejecutar)  [3] | Baja (Cualquier usuarios puede acceder y modificar los datos)  [1] | 6 |
| Alternativa 4 | Medio (Brinda entre 1 y 2 servicios)  [2] | Única  [2] | Grande (Conlleva más de 3 horas en ejecutar)  [1] | Intermedia (Cualquier usuario permitido puede acceder y modificar los datos)  [2] | 7 |
| Alternativa 5 | Muy alto (Brinda más de 3 servicios)  [4] | Restringida  [3] | Bajo (Conlleva menos de una hora en ejecutar)  [3] | Intermedia (Cualquier usuario permitido puede acceder y modificar los datos)  [2] | 13 |

*Selección*

Con base a los criterios referenciados anteriormente, y la calificación que le dimos a cada una de las alternativas en base a estos criterios, se tomó la decisión de utilizar la alternativa número cinco (Alternativa 5), esto debido a que es la propuesta más completa con un nivel de eficacia muy alto, brindando más de 3 servicios una vez haya sido ejecutada la plataforma, siendo una de las principales razones por las cuales el cliente quede satisfecho.

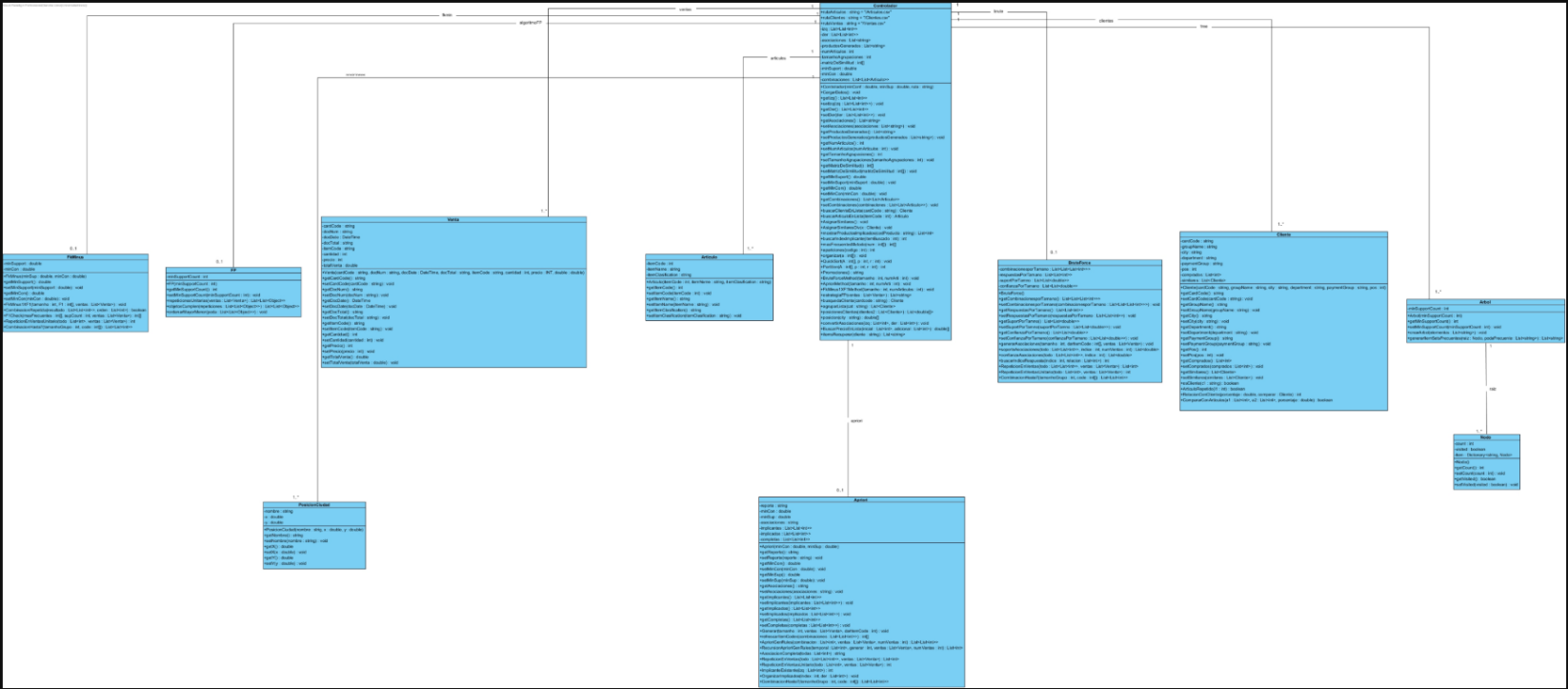
# Paso 6. Preparación de informes y especificaciones

*Especificación del problema (en términos de entrada / salida)*

*Problema:* Se requiere una plataforma que permita realizar diferentes análisis sobre la base de datos brindada por la empresa Allers Group

*Entrada:* Información acerca de clientes, ventas y artículos de la empresa Allers Group

*Salida:* Reportes, predicciones y servicios adicionales dependiendo del correcto análisis de datos.

Diagrama de clases

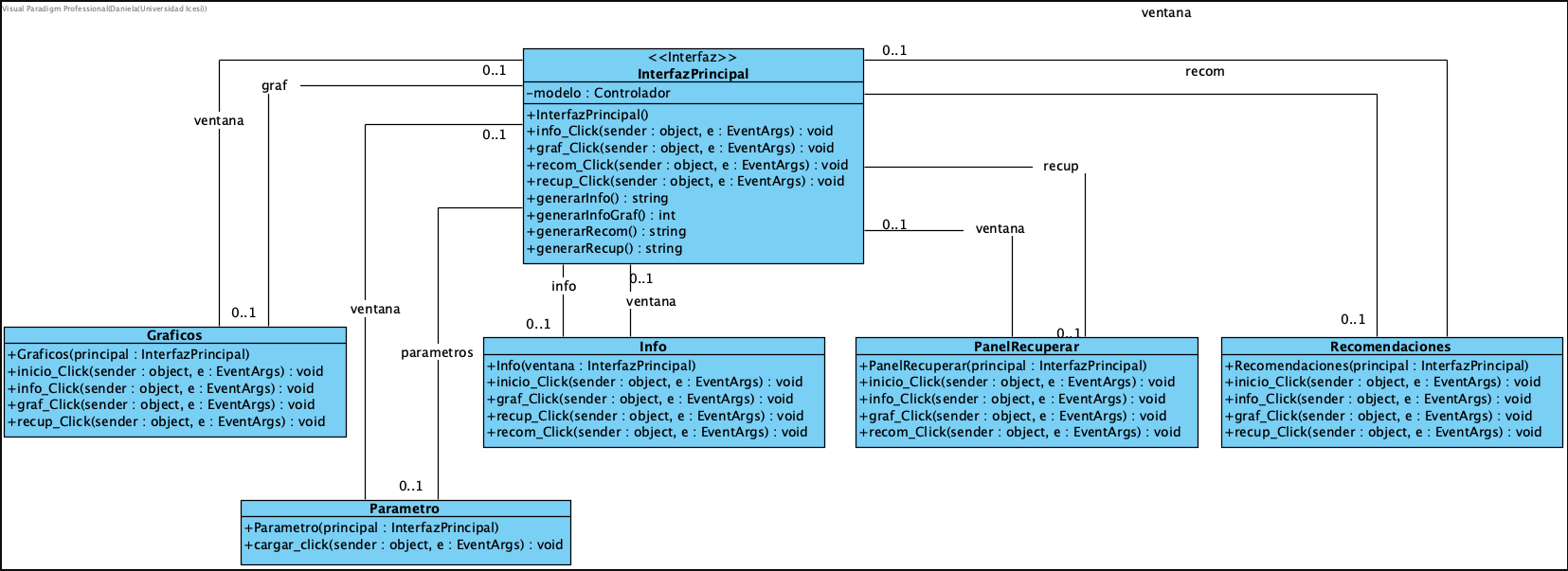
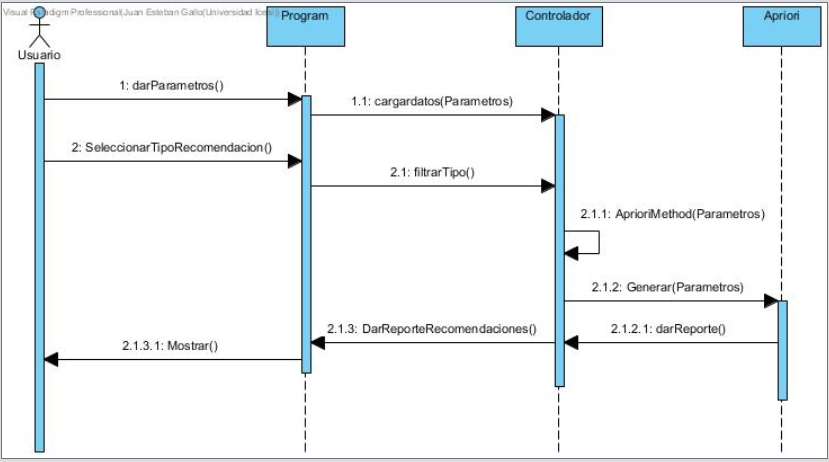
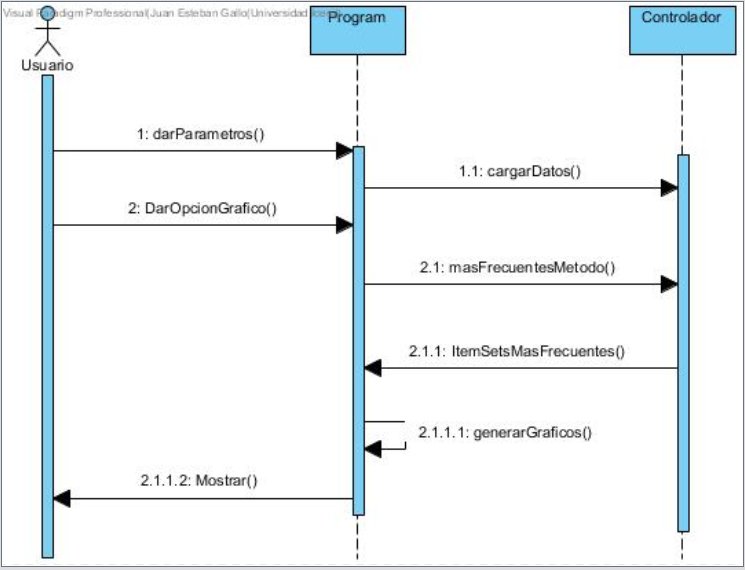
Diagrama de la interfaz gráfica

Diagrama de secuencias

Proceso de recomendaciones



Proceso de generación de gráficas



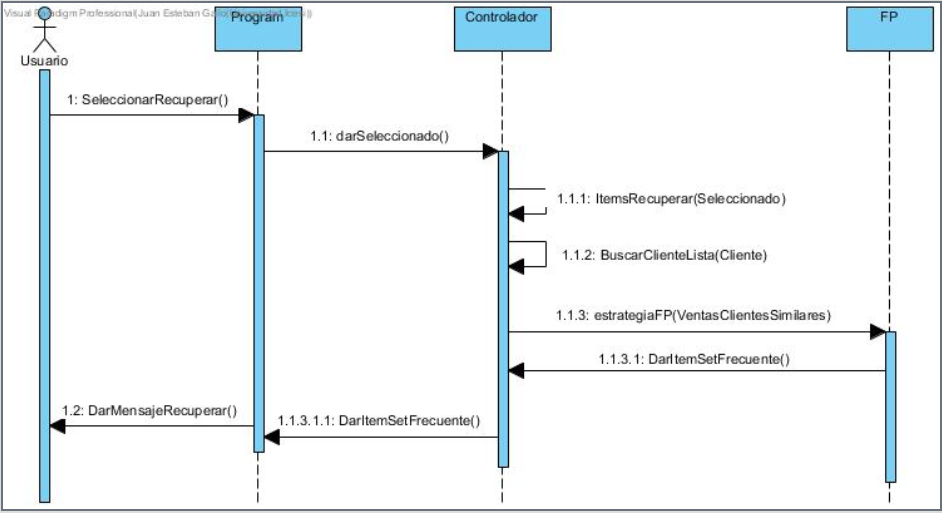
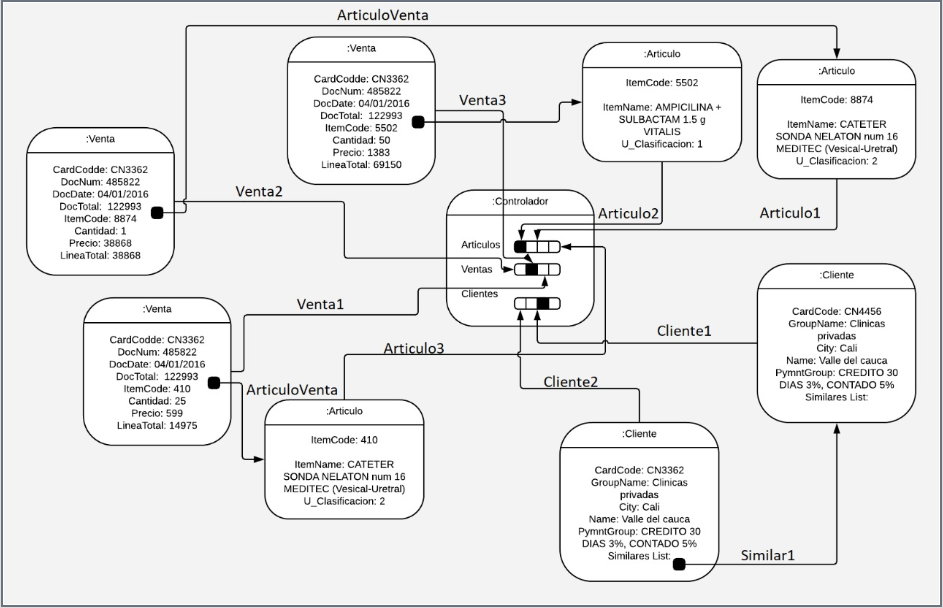
Proceso de Recuperar clientes

Diagrama de objetos



# Paso 7. Implementación del diseño

Implementación en el lenguaje de programación C# bajo la plataforma Visual Studio 2017

Lista de tareas a implementar:

* Implementación del modelo con sus diferentes clases, objetos, métodos, relaciones y atributos que ayuden a cumplir los requerimientos establecidos.
* Implementación de estrategias BruteForce, A-priori y F(k-1)xF1
* Implementación de la interfaz gráfica en cuanto a la visualización de clientes por medio de Gmaps
* Implementación de la interfaz gráfica en cuanto a recomendaciones sobre los productos
* Implementación de la interfaz gráfica en cuanto a similitud y recuperación de clientes.
* Implementación de la interfaz gráfica en cuanto a la visualización de graficas

Construcción en el lenguaje C#

Estrategias

public void BruteForceMethod(int tamanho, int numArti)

{

int[] darItemCode = masFrecuentesMetodo(numArti);

brute.generarAsociaciones(tamanho, darItemCode, Ventas,minSuport,minCon);

reporte += brute.Reporte;

}

public void AprioriMethod(int tamanho, int numArti)

{

TamanhoAgrupaciones = tamanho;

NumArticulos = numArticulos;

int[] darItemCode = masFrecuentesMetodo(numArti);

apriori.Generar(tamanho, ventas, darItemCode);

izq = apriori.Implicantes;

der = apriori.Implicados;

List<List<int>> completas = apriori.Completas;

int reglas = izq.Count;

for(int i = 0; i < reglas; i++)

{

ConvertirAsociaciones(izq.ElementAt(i), der.ElementAt(i));

}

foreach (var n in completas)

{

ConvertirAsociacionesCompletas(n);

}

reporte += apriori.Reporte;

}

public void FkMinus1XF1Method(int tamanho, int numArticulos)

{

int[] F1 = masFrecuentesMetodo(numArticulos);

fkmin.FkMinus1XF1(tamanho, F1, Ventas);

reporte += fkmin.Reporte;

}

public List<string> estrategiaFP(List<Venta> ventas)

{

List<string> retorno = new List<string>();

String mensaje = "";

int supCount = (int)(minSuport \* ventas.GroupBy(a => a.DocNum).Count());

algoritmoFP = new FP(supCount);

List<List<Object>> listaRepeticiones= new List<List<Object>>();

List<List<Object>> poda = new List<List<Object>>();

listaRepeticiones = algoritmoFP.repeticionesUnitarias(ventas);

poda = algoritmoFP.objetosCumplen(listaRepeticiones);

algoritmoFP.ordenarMayorMenor(poda);

Arbol tree = new Arbol((supCount/2));

List<string> listaEncontrada = new List<string>();

var x = ventas.GroupBy(n => n.DocNum);

foreach (var m in x)

{

String[] trans = new String[m.Count()];

int contador = 0;

foreach (var s in m)

{

trans[contador] = s.ItemCode;

contador++;

}

for (int i = 0; i < poda.Count; i++)

{

if (trans.Contains(poda[i][0].ToString()))

{

listaEncontrada.Add(poda[i][0].ToString());

}

}

tree.crearArbol(listaEncontrada);

listaEncontrada.Clear();

}

List<string> itemsFrecuentes = new List<string>();

itemsFrecuentes = tree.generarItemsSetsFrecuentes(tree.raiz, itemsFrecuentes);

mensaje += "El mayor item set es \n";

foreach (string item in itemsFrecuentes)

{

// string nombre = busquedaNombreItem(item);

retorno.Add(item);

string nombre = item;

mensaje +=nombre+ "\n";

}

return retorno;

}

Fuentes:

* Rafael Camps Paré, Luis Alberto Casillas Santillán, Dolors Costal Costa, Marc Gibert Ginestà, Carme Martín Escofet, Oscar Pérez Mora.”Bases de datos”.Fundación Universitaria Oberta de Catalunya.
* Windowa Forms Recuperado de: https://docs.microsoft.com/es-es/dotnet/framework/winforms/windows-forms-overview
* Big Data Recuperado de: <https://searchdatacenter.techtarget.com/es/definicion/Big-data>

* Data Mining Recuperado de: <https://www.gestiopolis.com/que-es-data-mining/>
* Conceptos c#, Visual Studio y Microsoft.NET Recuperado de: https://msdn.microsoft.com/es-co/library/dd492171.aspx
* Vodafone, [El futuro es aspirante].(2017,09,17).Big data para acabar con los atascos en Nueva York. Recuperado de https://www.youtube.com/watch?v=Vljm5JRxHKU&t=90s
* Big data para detectar y predecir ciberataques. Recuperado de <https://www.baoss.es/10-ejemplos-usos-reales-big-data/>
* Big data para salvar animales en extinción. Recuperado de <https://bismart.com/es>
* (s.f) Association Analysis: Basic Concepts and Algorithms Capitulo 6.