**Fase 4: Transición de la formulación de ideas a los diseños preliminares**

**Ideas descartadas:**

Para generar Frequent Itemsets:

Como en el problema se extraen relativamente pocos itemsets comparados con todos los ítems dados, entonces el recorrido deja de ser eficaz ya que se espera que haya una gran densidad de transacciones con respecto a la barrera de los frecuentes.

Para el análisis de las transacciones de Allers:

El método de Support Vector Machine utiliza la técnica del aprendizaje supervisado, lo cual se limita para desarrolladores estudiantes debido al poco tiempo que se tiene, requeriría hacer etiquetas a conjuntos de datos y hacer un proceso mucho más elaborado, además, existen otros métodos ligados a la predicción y con menores recursos.

Presentación de la información en la GUI:

La idea de que el programa muestre la información por medio de gráficos de barras, dispersión y tortas, apoyadas por datos numéricos representando dispersión y medidas de tendencia central no es atractiva para el cliente, dejo explicito el hecho de que ya poseía software que contenía tales características, por lo tanto, no se tomara en cuenta esta idea.

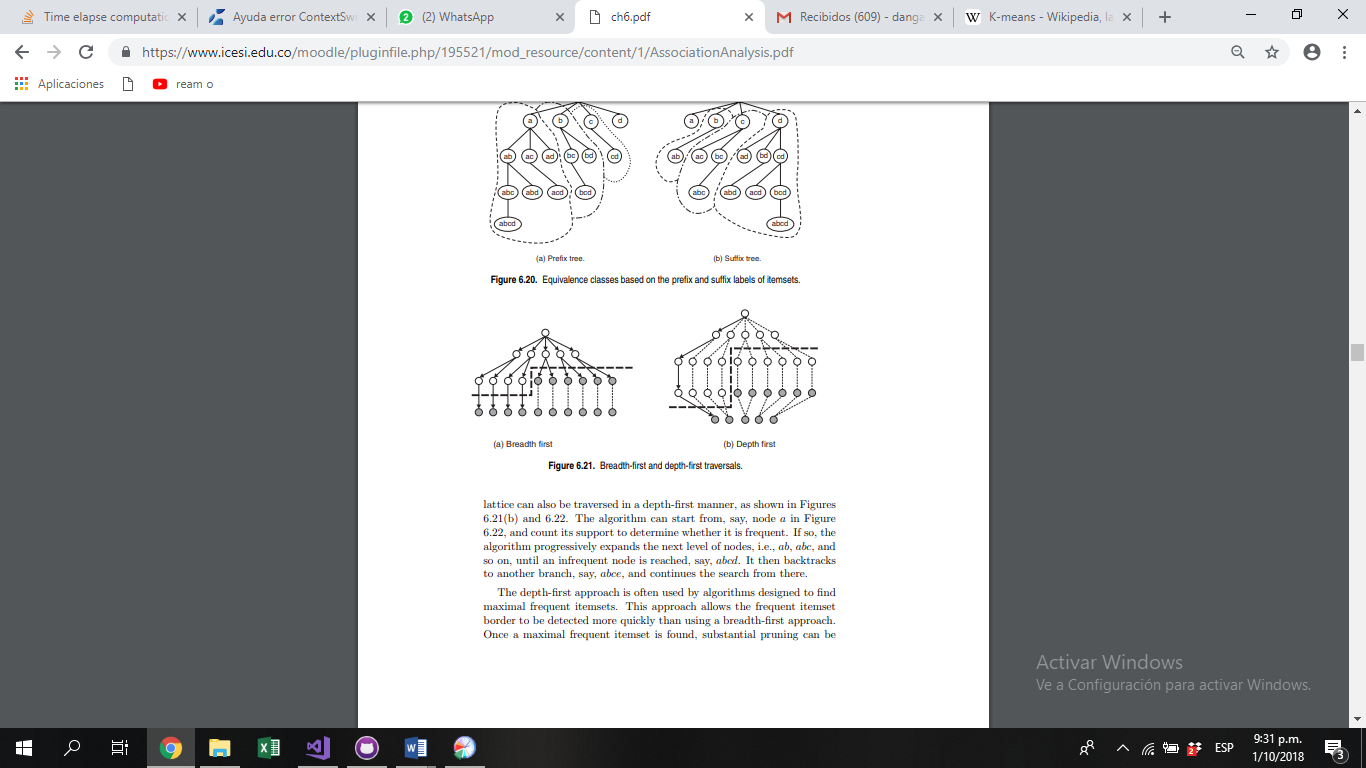
Análisis de los clientes:

Por razone similares a la idea descartada en análisis de transacciones, el SVM resulta ser una alternativa poco probable tanto para transacciones como para clientes, en general, la técnica del aprendizaje supervisado requiere de más detalle para su elaboración.

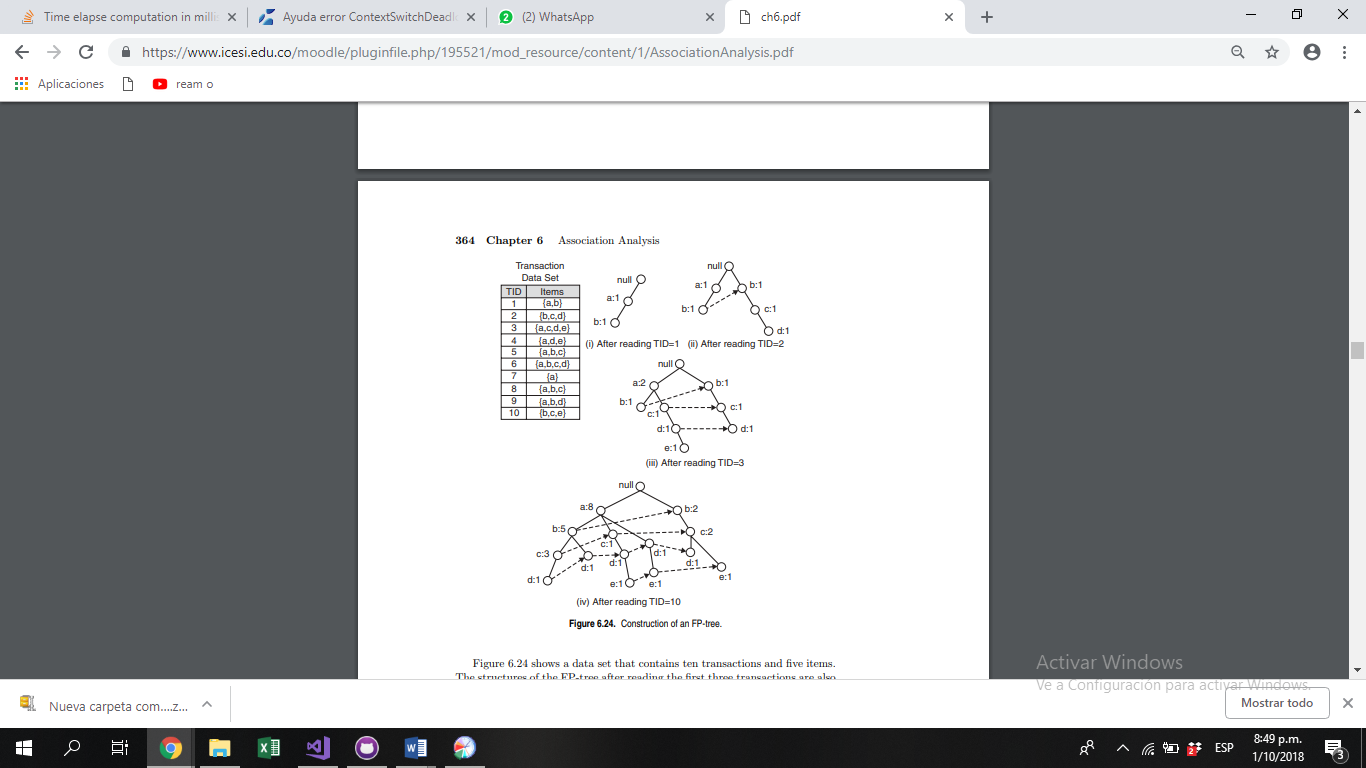
**Diseños preliminares:**

Para generar Frequent ItemSets:

1. El algoritmo va recorriendo los itemsets candidatos en profundidad, en vez de por anchura (Análogamente a los recorridos DFS y BFS de un grafo). Así pues, si bien el Apiori recorre primero los itemsets A, B, C, D ,E, después AB, AC, AD, etc.., recorrerlos por profundidad sería de la forma A, AB, ABC, ABD etc… Así, al encontrar un itemset no frecuente, el algoritmo se devuelve al itemset anterior y pasa a buscar el siguiente (Si para en ABD, se devuelve y pasa a ABCE). Su principal uso es el de encontrar Maximal Frequent Itemsets.



1. Representar las transacciones a través de un FP-Tree, con el objetivo de elaborar un FP Algorithm. Dicha representación es útil para representar las transacciones de manera condensada como un árbol. Los Frequent ítem sets se van encontrando a partir de los “sufijos” de los itemsets (Con qué elemento termina el itemset) siguiendo los caminos que contienen dichos sufijos. Emplea una técnica de “Divide y vencerás”. El siguiente es un ejemplo de FP – tree:



Análisis de las transacciones de Allers:

1. La aplicación contiene métodos relacionados con análisis de asociación. Esto es, la aplicación ejecuta procedimientos que aplican el análisis de asociación teniendo en cuenta los métodos predefinidos, las diferentes técnicas y variabilidad en las maneras de recolectar y analizar los datos.
2. Un prototipo del aplicativo que contenga métodos como medidas de tendencia central, anovas, estudio de distribuciones, aplicar probabilidades y concluir con lo que se obtiene usando estadística descriptiva.
3. La aplicación se implementa con matrices multidimensionales para asociar a cada ítem con la transacción y a su vez con los clientes, discretizando algunos atributos para hacer respectivas aproximaciones, en esta aplicación se dejan de utilizar filas y/o columnas para agrupar los objetos del problema.

Presentación de la información en la GUI:

1. La idea de presentar la información por clientes es relevante ya que uno de los objetivos con la aplicación es separar grupos de clientes a los cuales se les podría hacer descuentos u otro tipo de estrategias de marketing mediante el análisis de sus compras históricas, en la GUI, se observaría posibles predicciones de compra para cada cliente y su similitud con otros.
2. Se separa la información de transacciones por mes, esto para facilitar las predicciones y notar más fácilmente el crecimiento o decrecimiento en las ventas de ciertos productos. Esta vez se muestran las reglas extraídas por mes y soportes.

Análisis de los clientes:

1. A partir de los datos de las transacciones, se determinan clientes que son similares entre sí según unas métricas, para así recomendar a un cliente productos que él no está comprando, pero que los clientes similares si adquieren.
2. La aplicación contiene un método de agrupamiento denominado k-means, que ayuda en gran medida a la división de los clientes basados en sus transacciones y al trato de los mismos para realizar análisis más específicos. Si bien se trata de un algoritmo NP – Hard, existen heurísticas empleadas y diferentes variantes que hacen que el problema se solucione con un óptimo local.

