**Introducción:**

Durante el desarrollo del proyecto se hizo un método de fuerza bruta que permitía extraer los ítems sets más frecuentes. Sin embargo, se pudo observar que este método era muy ineficiente, además de que no permitía relacionar todos los productos.

Dado que los datos que se manejarán en la aplicación son demasiados, se decidió investigar más sobre métodos alternativos y se encontraron dos más: Apriori y FP-Growth. Estos dos métodos hacen la misma función que el algoritmo de fuerza bruta, con la diferencia de que son capaces de utilizar más productos para el análisis y pueden generar conjuntos de ítems muchos más grandes.

Ahora bien, el Apriori no puede tomar más de 220 datos para generar los conjuntos y esto, en un principio, es una gran desventaja debido a que puede haber conjuntos que se estarían omitiendo. Por otro lado, el FP-Growth es capaz de utilizar todos los ítems frecuentes. A pesar de las grandes diferencias presentadas, el Apriori se podría utilizar en algunos casos. Por ejemplo, en el análisis de conjuntos de ítems para un cliente en un clúster.

En pocas palabras, en los casos particulares en donde se puedan usar el Apriori, si este es lo suficientemente eficiente, se usará.

**Objetivo:** El objetivo de este experimento es determinar bajo diferentes condiciones qué algoritmo es más eficiente. Los resultados de este experimento permitirán determinar cuáles de los algoritmos se debe usar para el análisis de datos del proyecto Allers

**Unidad experimental:** Transacciones disponibles en los datos dados por Allers

**Variables de respuesta:** El tiempo que toma cada algoritmo en generar los itemset frecuentes

**Factores controlables:** Número máximo de ítems por cada itemset frecuente, número de ítems a evaluar, algoritmo a utilizar, tipo de procesador, lenguaje de programación para el desarrollo de los algoritmos, número de transacciones.

**Factores no controlables:** los ítems presentes en cada transacción, ruido en los datos.

**Factores estudiados:** la variante del algoritmo utilizado

**Niveles:** Las transacciones se estudian en 3 niveles (3 algoritmos). Los algoritmos de fuerza bruta y el a priori serán estudiados con 3 condiciones diferentes

**Tratamientos:** En total son 7 tratamientos

|  |  |
| --- | --- |
| **Algoritmo** | **Condiciones** |
| Fp Growth | Ninguna |
| Brute-force | minSup= 0.005,ItemsToEvaluate=20 ,maxItemSetSize=5 |
| Brute-force | minSup= 0.005,ItemsToEvaluate=25 ,maxItemSetSize=10 |
| Brute-force | minSup=0.005 ,ItemsToEvaluate= 30,maxItemSetSize=15 |
| Apriori | minSup= 0.005,ItemsToEvaluate=100 ,maxItemSetSize=5 |
| Apriori | minSup= 0.005,ItemsToEvaluate=150 ,maxItemSetSize=10 |
| Apriori | minSup= 0.005,ItemsToEvaluate=200 ,maxItemSetSize=15 |

**Procedimiento:**

La ejecución del experimento y el análisis de sus resultados lo harán dos personas. La ejecución la deberá hacer una persona con un mismo computador y el análisis de los resultados lo hará la otra persona.

Ejecución del experimento:

El encargado de ejecutar el experimento deberá realizar los 7 tratamientos que se diseñaron en la parte de arriba. Cada tratamiento deberá ejecutarlo 10 veces y plasmar todos los datos en un archivo de Excel, para que entregárselos al responsable del análisis.

En la ejecución del experimento se debe tener en cuenta que mientras el computador esté corriendo la aplicación, ninguna otra aplicación que no sea relevante debe ser abierta. Lo anterior con el fin de que los tiempos de ejecución sean los más precisos posibles.

Análisis del experimento:

Una vez recibido los datos, se procederá a analizarlos. Se utilizará tablas anovas para hacer la comparación entre los tiempos de ejecución. Posteriormente, teniendo en cuenta los resultados, el responsable del análisis deberá hacer las conclusiones indicando qué algoritmo o cuáles algoritmos deben ser usados en la aplicación y bajo qué condiciones.