**Diseño de experimentos: A-priori VS Clúster.**

**Problemática**: ¿Cuál de los dos algoritmos es más rápido? (A-priori VS Clúster).

Para esto trabajaremos sobre la base de datos que nos ha proporcionado la organización Allers. De igual manera, es necesario saber que los datos pasados por la organización han sido previamente analizados y corregidos con el fin de eliminar datos innecesarios. Dentro de estos se encuentran:

* Clientes atípicos (su presencia en las transacciones es muy baja).
* Productos erróneos (productos que se encuentran sin código en la base de datos).

Esto se realiza con el propósito de trabajar sobre datos más precisos, ya que si se decidiera escoger toda la base de datos, habría información que entorpecería el experimento, ocasionando resultados erróneos.

**Fase 1: Unidad experimental.**

Las unidades experimentales del experimento serán los algoritmos de:

* A-priori.
* Clúster.

**Fase 2: Variables de respuesta.**

Para este diseño de experimentos las variables de respuestas serán:

* Tiempo de ejecución de los algoritmos.

Es importante mencionar que la prueba se ejecutará veinte veces por algoritmo, es decir que se tomarán veinte tiempos para A-priori y veinte tiempos para Clúster.

**Fase 3: Factores controlables.**

Los factores de estudio controlables para este experimento son:

- Repeticiones del algoritmo.

- Cantidad de clientes.

- Cantidad de productos.

- Número de clústeres.

**Fase 4: Factores no controlables.**  
Los factores de estudio no controlables para este experimento son:  
  
- Capacidad de procesamiento del equipo donde se realice el experimento.

- Tareas que en ese momento se estén ejecutando en segundo plano en el sistema.

**Fase 5: Factores a estudiar.**

- Tamaño de entrada (MinConfidence y MinSupport).

- Variante técnica a analizar.

**Fase 6: Niveles.**

Para el desarrollo del experimento optamos por trabajar con un MinSupport de 2. Por otra parte, decidimos ir modificando la variable de MinConfidence para darnos cuenta de cómo trabaja y de la misma manera poder medir la velocidad de los algoritmos en cuestión.

Los niveles que se van a estudiar son:

* Tamaño de entrada (MinConfidence).
* 20
* 50
* 80
* Variante técnica a analizar.
* A-priori (técnica de análisis de asociación).
* Clúster (técnica de clusterización).

**Fase 7: Tratamientos.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Factores a estudiar** | |  |
| **# Prueba** | **Variante técnica a analizar** | **Tamaño de entrada** | **Tiempo en s** |
| **1** | A-priori | 20 | 12,85752784 |
| **2** | A-priori | 50 | 13,56504401 |
| **3** | A-priori | 80 | 11,51000736 |
| **4** | Clúster | 20 | 0,014074465 |
| **5** | Clúster | 50 | 0,018576185 |
| **6** | Clúster | 80 | 0,01585146 |