

Juan Fernando Jaramillo

Santiago del Campo

Juan David Carvajal

German Carvajal

**Diseño de Experimento**

* Etapa de planeación.

1. **Entender y delimitar el problema u objeto de estudio:**

Con base en la previa investigación y con la información suministrada por parte de la empresa Allers group, se han identificado los siguientes síntomas y necesidades, que ayudan a entender el problema que se está tratando para posteriormente poder delimitarlo, dejando claro el alcance que va a tener.

**Síntomas:**

**-**Gran cantidad de datos sin aprovechar, pues la información existe, pero se encuentra almacenada en una base de datos sin usos actual.

-Datos de propósito general, que contienen información que puede o no ser útil para la empresa.

-Expansión de la empresa, lo que genera que se produzcan aún más datos.

-Falta de mecanismos para análisis de datos.

**Necesidades**:

**-**Mecanismos de análisis de datos, para tratar toda la información nueva y vieja que posee la empresa, y darle un fin.

-Depuración de los datos a analizar, pues analizar datos no relevantes es una pérdida de recursos computacionales y tiempo.

-Capacidad de predecir comportamiento de compra de los clientes, necesidad establecida por el encargado del manejo de los datos.

**Definición del problema y su alcance:**

La información actual e histórica que posee la empresa Allers Group sobre las transacciones de sus clientes, productos y proveedores no está siendo utilizada, imposibilitando así la predicción de futuras ventas y transacciones.

* Solo se va a abordar el problema con los datos suministrados por la empresa, que fueron limitados por seguridad de la misma.
* Los datos no van a ser actualizados para este análisis.
* No será necesario la compra de servidores o servicios en la nube para alojar datos en grandes cantidades.
* Solo se utilizaran algunas técnicas de minería de datos para el análisis respectivo.
* Estos datos se trataran como la unidad experimental.

1. **Elegir la(s) variable(s) de respuesta que será medida en cada punto del diseño y verificar que se mide de manera confiable:**

Las variables de respuesta en esta problemática son:

* Tiempo de ejecución en los métodos de análisis.
* Precisión en las predicciones.

1. **Determinar cuáles factores deben estudiarse o investigarse, de acuerdo a la supuesta influencia que tienen sobre la respuesta:**

Factores controlables:

* Métodos a usar.
* Cantidad de datos.
* Datos estudiados.

Factores no controlables:

* Ánimo del equipo de trabajo.
* Desbordamientos o errores por las características de la máquina.
* Lentitud de visual studio.

Factores estudiados:

* Cantidad de datos analizados por cada método.

1. **Seleccionar los niveles de cada factor, así como el diseño experimental adecuado a los factores que se tienen y al objetivo del experimento:**

* Pruebas de los métodos con pocos datos (de 5 a 20 datos), o con todos los datos (más de 1000).

1. **Planear y organizar el trabajo experimental:**

Este trabajo será realizado por las personas nombradas en la página 1. Cada integrante tendrá una tarea diferente dentro del desarrollo del proyecto, como lo son: desarrollo de documentos, interfaces gráficas, análisis de métodos a usar, pruebas o repeticiones con los algoritmos, etc. Como primeras medidas, con algunos métodos de análisis se planeó usar un experimento piloto, el cual utiliza un número pequeño de observaciones. El objetivo de su ejecución es ayudar a completar y chequear la lista de acciones a realizar. Las ventajas que proporciona la realización de un pequeño experimento piloto son las siguientes:

— permite practicar la técnica experimental elegida e identificar problemas no esperados en el proceso de recogida de datos,

— si el experimento piloto tiene un tamaño suficientemente grande puede ayudar a seleccionar un modelo adecuado al experimento principal,

— los errores experimentales observados en el experimento piloto pueden ayudar a calcular el número de observaciones que se precisan en el experimento principal.

Posteriormente, se empiezan a implementar otros métodos de análisis que puedan abarcar más datos, y se realizan comparaciones con los anteriores para verificar cual se adapta mejor al problema.

Como últimas medidas se le agregaran funcionalidades a la interfaz gráfica y mejoras a los algoritmos para que se generen reportes, visualizaciones de datos, tableros de control, entre otros.