Orientación del Semi-proyecto

Descripción del proyecto

Cada equipo debe elaborar una **descripción clara y concisa de su proyecto**, respondiendo a las siguientes preguntas:

- **¿Qué se pretende lograr?** → Explicar el objetivo central del proyecto en una o dos frases (ej. analizar, predecir, detectar, clasificar, recomendar…).
- **Dataset seleccionado:** indicar cuál utilizarán (nombre, fuente y formato). **(Electivo)**
- **Justificación del dataset:** por qué es adecuado para este proyecto, destacando: **(Electivo)**
  - **Volumen:** cantidad de datos disponible, qué tan grande es y si permite simular "Grandes Volúmenes de Datos".
  - **Características:** qué contiene (ej. atributos, temporalidad, etiquetas, ruido).
  - **Pertinencia:** cómo se relaciona con el objetivo del proyecto.

Arquitectura propuesta

Deben presentar un **diagrama sencillo del pipeline de trabajo**, con flechas que conecten los principales componentes.

- Especificar si su enfoque será **batch**, **streaming** o una combinación.
- Cada bloque debe ir acompañado de una breve explicación en **una o dos frases** sobre su función en el proyecto.

Avances

Cada equipo debe presentar un **los avances que ya deben tener** (aunque sean prototipos). Los mínimos esperados son:

- ✅ Instalación y configuración de la infraestructura (HDFS, YARN, MapReduce/Spark).
- ✅ Carga de un subset inicial del dataset en HDFS.
- ✅ Ejecución de primeros procesos básicos de limpieza y transformación.
- ✅ Pruebas iniciales de consultas.

Si hay más avances que quisieran revisar se pueden presentar.

Esto require, informe escrito y exposición.