**Idea de proyecto Fernando Salazar – Aplicación de IA en análisis de datos operativos**

**Problemática a resolver**

En mi entorno laboral, una gran parte del tiempo del equipo de análisis de datos se invierte en la lectura manual y categorización de comentarios y observaciones escritas por técnicos de mantenimiento en reportes operativos. Estos reportes, usualmente generados en lenguaje natural, contienen descripciones de fallas, condiciones de operación, acciones correctivas y observaciones contextuales que no siguen un formato estándar.

Este proceso manual de revisión y etiquetado implica alto consumo de tiempo, subjetividad y una gran dificultad para detectar patrones que permitan tomar decisiones predictivas o correctivas a gran escala.

**Tipos de datos involucrados**

* Registros históricos de mantenimiento en formato de texto libre.
* Tiempos de respuesta y resolución asociados.
* Códigos de equipo, fechas y tipo de mantenimiento.
* Comentarios de técnicos escritos en español.

**Origen de los datos**

* Sistema de gestión de mantenimiento (CMMS).
* Base de datos interna (PostgreSQL).
* Formularios digitales utilizados en campo.

**Áreas o departamentos involucrados**

* Departamento de Mantenimiento y Operaciones.
* Departamento de Tecnología / Data Engineering.
* Gerencia de Planeación Operativa.

**Modelo de IA propuesto**

Se utilizaría un modelo LLM especializado en lenguaje español como **LLaMA 3 (Meta)** o **Gemini 1.5**. Se buscaría usar capacidades de *embedding* para agrupar por similitud semántica, y de clasificación para etiquetar automáticamente los textos en categorías predefinidas como: "falla recurrente", "revisión programada", "reparación crítica", etc.

El modelo también generaría sugerencias de acciones a partir del análisis de comentarios.

**Entregable del proyecto**

* Sistema de clasificación automática de reportes.
* Dashboards con analítica visual de fallas más frecuentes, zonas críticas, etc.
* Recomendador de acciones basado en texto libre.
* Dataset etiquetado para uso interno y mejora continua.

**Tecnologías y herramientas requeridas**

* Python (pandas, langchain, llama-cpp-python).
* PostgreSQL y SQL para extracción de datos.
* FastAPI para crear endpoints de clasificación.
* Streamlit o Power BI para dashboards.
* Infraestructura cloud (Google Cloud / AWS).

**Tiempos estimados**

| **Fase** | **Duración estimada** |
| --- | --- |
| Recolección y preprocesamiento | 2 semanas |
| Entrenamiento/pruebas del modelo | 3 semanas |
| Desarrollo de API y dashboard | 2 semanas |
| Validación con usuarios | 1 semana |
| Ajustes finales y documentación | 1 semana |
| **Total estimado** | **9 semanas** |

**Costos aproximados**

* Infraestructura cloud: $1,500 USD anuales.
* Desarrollo e implementación: $4,000 USD.
* Licencias de software y APIs: $800 USD.
* Capacitación y documentación: $700 USD.

**Costo total estimado inicial: $7,000 USD**