

Modelado y presentación de la solución basada en datos

Ciencia de datos aplicada: Tercer entregable

Objetivo:

En este entregable, se espera que cada equipo avance en la implementación de una solución funcional al problema planteado en las primeras etapas del curso. Dicha solución deberá implementar modelos a partir de las técnicas de ciencia de datos –ya sea a través de modelos predictivos, sistemas basados en recuperación de información (como RAG), análisis descriptivo avanzado o visualizaciones interactivas–, según la naturaleza de cada proyecto. Además, se espera que el modelo sea validado a partir de las métricas propias de cada tipo de solución.

Por último, se propone la presentación de cada proyecto en el encuentro final del curso, generando un ámbito de intercambio y aprendizaje compartido.

Entrenamiento y validación del Modelo

Cada equipo deberá implementar y validar un enfoque que permita abordar el problema planteado. El modelo o sistema puede ser de tipo predictivo, descriptivo, basado en recuperación de información (RAG), análisis exploratorio avanzado u otra técnica justificada.

Se espera que este apartado incluya:

- Descripción del enfoque adoptado y su justificación.
- Implementación del modelo o sistema de base (ej.: algoritmo de ML, arquitectura RAG, lógica de análisis, etc.).
- Evaluación de la solución:
 - Si se trata de modelos predictivos: métricas apropiadas (accuracy, F1, MAE, etc.).
 - Si se trata de sistemas no tradicionales (como RAG): criterios de evaluación acordes (precisión en recuperación, tests de uso, resultados representativos).

- Análisis de los resultados.
- Reflexión crítica sobre el rendimiento y posibles mejoras.

Formato de entrega: Se propone la entrega de una Notebook (.ipynb o link a Google Colab).

Modelo de Entrega: [Ejemplo de Notebook con Entrenamiento y validación del modelo](#)

Prototipo funcional y propuesta de despliegue

Los equipos deberán desarrollar un prototipo mínimo viable que permita interactuar con el modelo o sistema implementado.

El prototipo debe:

- Simular el uso del modelo/sistema con nuevos datos (hacer una clasificación, predicción o responder ante un prompt, según sea el caso).
- Incluir una propuesta (al menos teórica) de despliegue:
 - ¿Cómo podría funcionar en un entorno real?
 - ¿Qué recursos requeriría?
 - ¿Qué alternativas existen para escalarlo?

Formato de entrega: Notebook, video-demostración, app simple (Streamlit, Flask, etc), dashboard, etc.

Modelos de Entrega (opciones):

- Notebook: colab.research.google.com/drive/1bTmAV8yLe9W3GKpj4FSvBW9X2uN0c36
- App Streamlit: <https://iris-svm.streamlit.app/>
- Code Streamlit: github.com/jumafernandez/itba-data-science/tree/main/model-app-streamlit
- Flask (code): github.com/jumafernandez/itba-data-science/tree/main/model-app-flask

Presentación oral del proyecto

En la última clase del curso, cada equipo realizará una exposición de su proyecto de entre 10 y 15 minutos. Esta instancia será compartida con los demás estudiantes y el equipo docente.

Se espera que la presentación incluya:

- Presentación del problema abordado y los objetivos del proyecto.
- Estrategia general y decisiones tomadas a lo largo del proyecto.
- Demostración en vivo o grabada del prototipo funcional.
- Resultados obtenidos y conclusiones.
- Reflexión sobre dificultades, aprendizajes y próximos pasos.