

Indicaciones para el proyecto final de la asignatura

La asignatura “**Machine Learning para Analítica de Negocios**”, plantea la utilización de algoritmos de Machine Learning para la predicción de eventos en un entorno específico de desarrollo. Para ello, los estudiantes, a lo largo de la asignatura, han descubierto y desarrollado nuevas habilidades técnicas relacionadas a la creación de modelos y ajuste de los mismos para obtener un valor de predicción más fiable en entornos de desarrollo y producción.

El desarrollo del producto final se llevará a cabo en equipos, donde, solo uno de los integrantes subirá los avances. A continuación, se detallan los incisos que serán evaluados como parte del proyecto final de la asignatura.

Entregables:

- Los equipos deberán crear un nuevo repositorio en **Github** (privado) con la siguiente nomenclatura: ml_[nombre_proyecto]. Si la información (datos, Excel u otro tipo) son confidenciales, entonces no subirlos a **Github** aunque sea en un entorno privado, enviarlo a través de correo electrónico al docente.
- Generar una estructura de proyecto en **VSCode** tal y como se plantea al final del documento del docente. (Revisar parte final del documento)
- El archivo readme.md deberá contener información general del proyecto y como un usuario puede acceder al proyecto, descargarlo y ejecutar los comandos para poder ser utilizado
- Generar un nuevo **environment** diferente al que se utiliza en clases para el desarrollo de sus modelos. Por ejemplo, puede crearlo conforme a la siguiente nomenclatura: ml_proyecto_[Nombre_Proyecto]
- Generar un modelo predictivo usando métodos de aprendizaje supervisado (regresión o clasificación)
- Para el desarrollo de los modelos deberá aplicar los siguientes elementos:
 - **EDA** (Exploratory Data Analysis). Aplique los métodos que el equipo considere necesario para eliminar outliers
 - En caso de ser útil para los estudiantes, pueden utilizar algoritmos de aprendizaje no supervisado (clustering) para encontrar patrones dentro de los conjuntos de datos a procesar
 - Métodos estadísticos para **Feature Selection**. Analice correctamente la distribución de los datos para considerar los métodos a aplicar.
 - Aplicar **Feature Engineering** en caso de encontrar la necesidad de crear una variable que no está presente dentro de su conjunto de datos, pero será clave para el desarrollo de su modelo y predicción
 - Generar los modelos utilizando los algoritmos de Machine Learning aprendidos a lo largo del curso
 - Utilizar **GridSearch**, **RandomizedSearch** y **Bayesian** para encontrar los mejores parámetros para ajustar los modelos de clasificación seleccionados
 - Describir a través de visualizaciones, los resultados obtenidos para cada modelo generado. Considere la inclusión del baseline como parte de sus resultados.

- Valorar la inclusión de un componente de NLP para el proyecto (esto puede ser tomado en consideración hasta el final)
- Los equipos deberán generar un archivo en formato **LaTeX** con una estructura en dos columnas mostrando los avances de sus proyectos. En este documento, deberán ir aspectos técnicos del desarrollo de sus modelos.
- Se considerarán los siguientes acápites dentro del documento **LaTeX**: Introducción, planteamiento del problema, objetivos, datos, referencias bibliográficas. Revisar la parte final del documento para mayor detalle sobre cada uno de los puntos a evaluar.

Descripción de los acápites del documento final:

- **Introducción:**
 - Describir el contexto y motivación, porqué este problema es importante
 - Objetivos del proyecto y cómo el modelo generado promueve una solución al problema
- **Datos:**
 - Fuente de los datos: indicar la fuente de datos (personal, recolectado a partir de sistemas OLTP, etc.)
 - Estructura de los datos: formato de los datos, número de ejemplares, features, tipos de datos, explicación de cada variable, etc.
 - Particionamiento de los datos: descripción de las particiones de entrenamiento, testing y validación (en caso de aplicar)
 - Preprocesamiento de datos: detalle de cómo procesó los datos y los transformó para ser utilizados durante el entrenamiento
- **Metodología:**
 - Selección de los algoritmos utilizados y técnicas que se implementaron
 - Feature Engineering (si aplica)
 - Arquitectura del modelo: estructura del modelo, hiperparametros, etc.
- **Resultados:**
 - **Métricas cuantitativas:** precisión, accuracy, recall, F1-Score, matriz de confusión o cualquier otra métrica relevante para demostrar el resultado óptimo de su modelo
 - **Análisis cualitativo:** Ejemplos de clasificación, predicciones y análisis de errores
 - **Tablas y figuras**
 - **Interpretación**
 - **Limitaciones**
 - **Comparaciones entre los modelos generados**
- **Conclusión**
 - Contribución de su modelo
 - Recomendaciones para futuros trabajos
- **Referencias bibliográficas:**
 - Formato APA

Árbol de estructura propuesta:

my_ml_project/

├── data/

│ ├── raw/

│ │ └── ...

│ ├── processed/

│ │ └── ...

├── models/

│ └── ...

├── notebooks/

│ └── ...

├── src/

│ ├── __init__.py

│ └── ...

├── tests/

│ └── ...

├── requirements.yml

├── README.md

└── .gitignore