



Proyecto ML - Machine Learning

En este proyecto vas a aplicar todo lo aprendido en el bootcamp hasta el momento, para lo que deberás tener en cuenta lo siguiente:

1. **El proyecto es individual**
2. **Datos:** Tienes que conseguir los datos por tu cuenta (puedes recurrir al apartado de datos del Proyecto EDA)
3. **Fecha de Entrega del Tema:** Hasta el viernes 14 de marzo, usando este [formulario](#).
4. **Fecha de Entrega Final de la Presentación:** Hasta el jueves 3 de abril.
5. **El proyecto lo expondréis en un vídeo** (*opciones explicadas más abajo*)
6. **Todas las entregas se realizarán en el foro que habilitaremos en el Campus Virtual:** Mediante enlace a vuestro repositorio personal en GitHub y enlace a la presentación en vídeo.

Objetivo

El objetivo del proyecto de Machine Learning es el de evaluar tus capacidades para construir un modelo de ML que resuelva un problema concreto de principio a fin, sin considerar la parte de productización o industrialización. Para ello, se piden desarrollar un conjunto de actividades que den lugar a entregables y luego realizar una exposición en vídeo del trabajo y los resultados. La aproximación puede ser supervisada o no supervisada, y puede utilizar ML tradicional, DL o probar ambos y quedarse con el más adecuado.

No es objetivo del proyecto resolver un problema específico de la vida real, aunque eso ayuda :-), ni tampoco demostrar que se conoce al dedillo cualquier framework (sklearn, keras) de los vistos, aunque también ayuda :-). Se valorará más la resolución sensata, el seguimiento razonado de los pasos aprendidos y el análisis claro y bien expuesto de los resultados ligándolos con el problema que se quiere resolver.

Podéis escoger el dataset y problema que queráis siempre que defendáis con argumentos que este problema tiene algún tipo de solución o ayuda a su solución empleando Machine Learning. En los casos indicados, será necesario hacer un EDA que apoye nuestro proyecto de Machine Learning.

Entregables

1. La dirección de **un repositorio Github público nuevo y dedicado exclusivamente al proyecto** de Machine Learning con el nombre de `ML_<temática-del-proyecto>`. En este repositorio debe haber un fichero `README.md` en el que se explique el problema que se quiere resolver, se indique el dataset empleado (es decir se describa brevemente, y se indique si es público o privado y en el primer caso se de una forma de acceder a este), se describa brevemente la solución adoptada y se comente la estructura de directorios del repositorio. **Puntuación (5/100)**, Extra: Hacerlo en inglés y en español (5/100)
2. **Una muestra del dataset** (no más de 5MB de tamaño) siempre y cuando sea público a incluir en un directorio `/src/data_sample`.
3. **Un notebook claro y conciso con el paso a paso del proceso**. Debe ser un notebook que al ejecutar todo termine por generar el modelo y lo evalúe (como si fuera una práctica obligatoria de clase). Es necesario destacar cada paso empleando headers y negritas, (Por ejemplo **Paso 1: Entendiendo el problema, Paso n: Evaluación contra test**, etc). No es tu notebook de trabajo, es el notebook que lo resume, sin alternativas ni pruebas, y que debe funcionar completo y sin errores. Debe ir en una carpeta `/src/results_notebook`. **IMPORTANTE**: Este notebook debe tener como paso final la grabación del modelo en formato "pickle", "joblib" o "h5" (tensorflow) que creáis conveniente en la carpeta `src/models`. Se valorará tanto el código, como el orden y la claridad. **Puntuación (50/100)** Extra: (10/100) realizar el notebook en inglés.
4. **El documento o soporte que se emplee como apoyo a la exposición en vídeo de resultados** (ver apartado siguiente). Puede ser en el formato que se desee mientras sea reproducible sin necesidad de software comercial. **Puntuación (5/100)**. Extra: Entregar una versión en inglés (5/100)

Resumiendo lo indicado arriba, el código deberá seguir la siguiente estructura de carpetas:

1. `src/data_sample` : los archivos de datos de muestra utilizados en el proyecto que permitan ejecutar el código. [Recuerda no subir a GitHub archivos demasiado pesados]
2. `src/img` : Imágenes que hayas necesitado para tu proyecto.
3. `src/notebooks` : los notebooks usados para pruebas.
4. `src/results_notebook` : el notebook final resultante del proyecto.
5. `src/models` : los modelos guardados al ejecutar el código del proyecto.
6. `src/utils` : todos los módulos, funciones auxiliares o clases creadas para el desarrollo del proyecto.

Exposición del proyecto

A diferencia del proyecto del EDA en el que la presentación es casi tan importante o más que el desarrollo del análisis, en este caso lo realmente importante es llegar al mejor modelo posible de forma razonada. Pero no por ello, exponer los resultados es algo a descuidar.

Se pide:

1. Realizar **una exposición en vídeo** del problema, proceso, resultados y análisis de los mismos.

2. **La duración máxima del vídeo es de 7 minutos**, a partir de ahí el exceso de tiempo consumido contará negativamente. **Puntuación (20/100)**

Guía de exposición:

A continuación se da una guía para la exposición que debe considerarse como lo que, es una guía, no es obligatorio ceñirse a ella mientras se cubran los aspectos requeridos.

1. Presentación del problema de negocio. *Por ejemplo: Ayudar a reducir la mortalidad por cáncer de mama. Reducir los gastos de mantenimiento de las centrales nucleares.*
2. Planteamiento del problema técnico y cómo resuelve o ayudar a resolver el problema de negocio. *Por ejemplo: Predicción de aparición de cáncer en los próximos dos años, ayuda porque pueden aplicarse técnicas preventivas. Predicción del mejor momento para hacer una revisión completa de la central nuclear, ayuda en la planificación de las revisiones (vaciados de los generadores principales) y a hacerlas coordinadas para que no suba el gasto eléctrico.*
3. Descripción del dataset y EDA/miniEDA (target si lo hubiera, agrupaciones más interesantes en un clustering, features más predictivas, etc.)
4. Arquitectura de la solución, modelos probados, resultados de la comparativa (si la hubiera). Modelo seleccionado final (o modelos si fuera un combinado) y resultado contra test o resultado de las agrupaciones.
5. Conclusiones finales y acciones de mejora.

Plazos

Todas las entregas se realizarán en el foro que habilitaremos en el Campus Virtual, antes de que finalice el jueves 3 de abril. Y deberán incluir:

1. **Enlace a vuestro repositorio personal en GitHub**
2. **Enlace a la vídeo-presentación. IMPORTANTE: Vídeo en formato MP4**

Tutorías

Durante el periodo no lectivo del 24 de marzo al 4 de abril estaremos disponibles para consultas y/o tutorías específicas para el proyecto.

Vídeo

Para la presentación en vídeo aquí tienes una lista de recursos:

- Microsoft Powerpoint:
 - La herramienta de presentaciones por excelencia, también tiene función de grabar.
 - **Evidente necesita licencia, pero si ya lo tenéis instalado -> problema resuelto.**
 - Recordad exportar el vídeo en MP4, muy importante.
- Presentaciones de Google:

- Solo se puede acceder a la función de grabación con una cuenta de pago de tipo Google Workspace.
- Existe la posibilidad de crear una cuenta de prueba gratuita de 14 días, si lo hacéis usad una cuenta de correo secundaria.
- [Loom](#):
 - Loom es un capturador de pantalla que permite grabar toda vuestra pantalla y compartir los vídeos resultantes.
 - Su versión gratuita permite grabar como máximo vídeos de 5 minutos.
 - Existe la posibilidad de crear una cuenta de prueba gratuita de 14 días para grabar vídeos sin límite de duración, pero tened en cuenta que la entrega del proyecto de ML/DL también será en vídeo, así que si lo hacéis usad una cuenta de correo secundaria.
 - Importante, si hacéis dos vídeos en Loom, juntadlos en un único archivo MP4 para la entrega.
- [OBS Studio](#):
 - El capturador de pantalla open source por excelencia.
 - Totalmente gratuito, tiene funcionalidades para captura de pantalla, streaming, etc.
 - Es la opción más potente para hacer el vídeo, pero también la más difícil de configurar y utilizar.
 - Recordad exportar el vídeo en MP4, muy importante.