

## Instrucciones Trabajo en Grupo de Predicción

Tu rol es el de un científico de datos que tiene como objetivo resolver un sencillo problema de predicción de si un usuario recibirá un salario superior o inferior a 50.000 u.m. (Unidades Monetarias). Para ello, cuentas con un dataset de unos 50.000 registros y 15 columnas. 14 descriptivas y una predictiva. El problema se sencillo, y tiene como propósito que profundicéis en las técnicas de modelos.

---

A continuación, te desgloso las instrucciones del trabajo:

- El proyecto debe contener un EDA y unas conclusiones analíticas completas, tal y como se hizo en el trabajo individual. Se recomienda que este EDA se haga en un notebook aparte.
- Posterior al EDA, debéis llevar a cabo una Selección y Transformación de variables dedicada. Debéis utilizar, al menos, dos tipos de encoding, y debéis indicar porqué utilizáis ese encoding y a que variables.
- Se deben aplicar **DOS** modelos diferentes al problema. Yo recomiendo utilizar Random-Forest y Lightgbm. Ambos deben tener su tuneo de hiperparámetros mediante randomized-search o grid-search. Se recomienda encarecidamente utilizar cross validation en los modelos. Ambos modelos deben tener una versión baseline (sin tuneado de hiperparámetros - Fine Tuning) y otra más avanzada con Fine tuning.
- Ambos modelos deben contar con todas las métricas vistas en clase y explicadas con unas conclusiones.
- El trabajo es sencillo. Se valorará la realización de experimentos, iteraciones varias de los modelos, variaciones en las técnicas utilizadas... etc.
- **La nota no depende exclusivamente del rendimiento del modelo.** Es una combinación de las técnicas utilizadas, buenas prácticas de código, rendimiento del modelo y suma de iteraciones.

Además, debemos tener en cuenta la siguiente guía de uso de la IA y desarrollo del trabajo:

- Los trabajos son en grupos de tres personas o, alternativamente, individuales. El contenido a entregar será el mismo en ambas tipologías. Los grupos deben comunicarse antes de la fecha estipulada en el anuncio de la Tarea. En caso contrario, los asignará el profesor entre aquellos que no hayan enviado los integrantes.
- La tarea **debe** entregarse en Github, mediante el repositorio. Se recomienda que sea privado con acceso para los integrantes y el profesor. Si la tarea no se entrega por este método, contará como no entregada.
- El código debe poder ejecutarse en ordenadores ajenos sin modificaciones relevantes. Por lo tanto, si el profesor no puede leer el archivo mediante **ruta relativa** se calculará la nota sobre 5 puntos y sólo se revisará la sintaxis. Por lo que, si no es perfecta, no se alcanzará el 4 y **la prueba contará como suspenso**.
- **Todo el código** que tenga origen en **herramientas de IA debe indicarse como tal**, mediante un comentario en la línea de código. Si se abusa de la herramienta, el profesor hará preguntas específicas al alumno del funcionamiento del código y variables instanciadas. En caso de que la consulta no sea respondida correctamente, esa parte del código se omitirá y no contará para la corrección. **El trabajo en grupo será más exigente en este aspecto que el trabajo individual**
- Se recomienda contar con una herramienta de redacción a mano (Haroopad, Word, notas) para comentar los aspectos importantes de las variables, su distribución y tus conclusiones. Todo lo que consideres necesario, redáctalo, es tu forma de explicar el trabajo que estás llevando a cabo. Es mejor explicar algo erróneo, pero con buen criterio que introducir código sin contexto.
- La tarea se realizará en horario fuera de clase y, si el profesor lo considera oportuno, también en ciertas horas de clase.