PRÁCTICA 1 – ML MUCD CUNEF UNIVERSIDAD

Trabajaremos con este dataset: <https://www.kaggle.com/tbsteal/canadian-car-accidents-19942014?select=drivingLegend.pdf>

Preguntas:

- ¿Qué tipos de vehículos (modelos, antigüedad, etc.) y conductores son más propensos a tener accidentes (acción correctiva en prima)?

- ¿Qué tipos de vehículos (modelos, antigüedad, etc.) y conductores son menos propensos a tener accidentes (descuento en prima)?

- ¿Qué es lo que más contribuye a que existan fallecimientos en un accidente?

- Dado un accidente, ¿se puede generar un modelo que prediga si habrá fallecimientos o no? ¿Si se va a necesitar tratamiento médico o no? Las aseguradoras tienen que inmovilizar capital para pagar estas casuísticas.

- Plus: complementar con datos abiertos de clima (aunque Canadá es muy grande) y de otra tipología, ¿hay algún tipo de relación con temperaturas medias, precipitación media del día/mes, nieve...? ¿a más días festivos o de vacaciones, más accidentes? etc.

Libertad para generar análisis de valor y nuevas ideas.

Pasos:

1. EDA. (en notebook aparte -> 01\_EDA)
2. Establecer objetivos a modelar.
3. Para cada variable target
   * Modelo base (imputar el mayoritario)
   * Modelo de regresión lineal simple (¿se puede usar?) + Lasso
   * Selección de variables (justificación).

Para cada modelo extraer sus resultados (accuracy, precisión, ROC, …)

* + Modelo GLM
  + SVM
  + Modelo RandomForest
  + Modelo XGBBoost
  + Modelo LightGBM
  + Optimización de hiperparámetros / pipeline sobre el mejor modelo que consideremos]
  + Interpretabilidad

1. Conclusiones finales

Muy importante sería introducir comentarios sobre explicabilidad (tenéis código de ejemplo en la plantilla que os proporcioné):

[https://christophm.github.io/interpretable-ml-book/ (Enlaces a un sitio externo.)](https://christophm.github.io/interpretable-ml-book/) (muy importante aquí estudiar el apartado de SHAP)

[**https://pybonacci.org/2018/10/14/explicando-la-explicabilidad/** (Enlaces a un sitio externo.)](https://pybonacci.org/2018/10/14/explicando-la-explicabilidad/)

No olvidéis comentar las distintas matrices de confusión, métricas varias (como la f1 score) y resultados varios que salen.

Entregad en la tarea del canvas una carpeta comprimida en formato **ZIP** nombrada con las iniciales de los alumnos que han hecho la práctica separados por un guión bajo (ejemplo:DBS\_ARS.zip), aunque sea el mismo material, debe ser entregado por todos los alumnos que han participado en la práctica. Obligatorio seguir la siguiente estructura mínima (estructura directorios, pueden aparecer otros directorios como ref o src si se precisa):

- Archivo README.md conteniendo dentro los nombres y dirección de correo de los alumnos que han trabajado de forma conjunta en la práctica. Además, aquí debe aparecer la ruta GitHub al proyecto.

- Directorio data: pero NO incluir los datos, porque por el tamaño daría problemas al subirlo. Por tanto, este directorio estará vacío en la subida. Sí estará completo en GitHub.

- Directorios notebooks: con todos los notebooks.

- Directorio html: con la versión html de los notebook (en salvar como... de jupyter).