



# CaixaBank Tech Hackathon

RICARDO DURÁN

# EDA (Exploratory Data Analysis)

- ▶ En primer lugar, se realiza un análisis exploratorio de los datos para entenderlos, comprenderlos y encontrar patrones o nuevas variables. De este análisis, se ha determinado:
  - ▶ Se encontraron 133 filas en el dataset con todos sus columnas vacías. Se han eliminado estos datos.
  - ▶ No hay datos de tipo categórico (todos son números reales excepto la variable a predecir, que es binaria)
  - ▶ La columna *Close* y *Adj Close* son prácticamente iguales.
  - ▶ Se pueden crear variables que modelen la tendencia del Ibex, por ejemplo, restando los valores de apertura o cierre.

# Preprocesado de datos

- ▶ Antes de introducir los datos en un modelo, hay que prepararlos:
  - ▶ Se vuelven a eliminar aquellas filas con valores nulos.
  - ▶ Se dividen los datos entre aquellos correspondientes a las características y al target.
  - ▶ Se normalizan los datos, ya que se encuentran escalas distintas entre ellos.
  - ▶ Se dividen los datos entre un conjunto de entrenamiento y uno de validación.

# Modelado

- ▶ Una vez tenemos los datos listos, podemos pasar a proponer distintos modelos.
- ▶ Es un problema de clasificación, por lo que podemos usar los siguientes modelos:
  - ▶ Regresión logística
  - ▶ Árboles de decisión
  - ▶ Random Forest
  - ▶ Ada Boost
  - ▶ Voting (conjunto de modelos)

# Resultados

- ▶ Después de realizar distintas pruebas, el modelo de voting parece ser el que obtiene mejores resultados en el conjunto de validación.
- ▶ Finalmente, se cogen los datos *test\_x*, se preparan de la misma forma que los datos de entrenamiento (se les hace el mismo procesado), y se predice la variable objetivo.
- ▶ El resultado se muestra en los ficheros *predictions.csv* y *predictions.json*