



KAGGLE COMPETITION

# SALARY PREDICTION



# CONTENIDOS

- 01 Introducción
- 02 Exploración de los datos
- 03 Limpieza y transformación
- 04 Machine Learning
- 05 Resultados

# INTRODUCCIÓN

Se trata de buscar un modelo de Machine Learning que mejor se ajuste para predecir el salario en USD, que depende de otras variables como el puesto de trabajo, la localización de la empresa o el tipo de contrato, entre otras.

# EXPLORACIÓN DE LOS DATOS

Se estudia la consistencia de todas las variables y en qué medida influyen sobre la columna a predecir.

# LIMPIEZA Y TRANSFORMACIÓN

- Eliminación de outliers.
- Eliminación de algunas columnas.
- Get dummies de las columnas categóricas.
- Normalización.

# MODELOS LINEALES

- LinearRegression
- Lasso
- Ridge
- ElasticNet

# NO LINEALES

- Support Vector Regressor (SVR)
- Random Forest Regressor(RFR)
- Extra Tree Regressor (ETR)
- Gradient Boosting Regressor (GBR)
- XGB Regressor (XGBR)
- Cat Boost Regressor (CTR)
- LGBM Regressor (LGBMR)

# RESULTADOS

MODELO

MODELO LINEAL  
RIDGE  
ENTRENANDO CON  
EL 100 % DE LOS  
DATOS DE  
SALARIES

RMSE

35698.751

R2

0.63