



PREDICCIÓN DE SALARIOS COMPETICIÓN DE KAGGLE

Fecha: 08/03/2023

Realizado por: Pedro Jódar

ÍNDICE

1

Exploración y limpieza

*Primer análisis y
limpieza de los
datos*

2

Entrenamiento y testeo

*Lazy, H2O y
SKLearn*

3

One-hot-encoding y normalización

*get_dummies y
StandardScaler*

4

Búsqueda de hiperparámetros

Grid searching



EXPLORACIÓN Y LIMPIEZA

Continent

AME	135574.089855
AUS	87064.000000
EUR	70662.116667
ASI	33655.366667
LTM	20601.333333

Name: salary_in_usd,

experience_level

EX	196078.850000
SE	136341.794239
MI	90384.610465
EN	59989.569231

Name: salary_in_usd,

employment_type

CT	222750.000000
FT	113914.630165
FL	48000.000000
PT	35053.375000

Name: salary_in_usd,

company_location

AME	141093.854037
AUS	99709.333333
EUR	68349.083969
ASI	43941.514286
AFR	42318.000000
LTM	21160.800000

Name: salary_in_usd,

RETO

**GRAN CANTIDAD DE
COLUMNAS CATEGÓRICAS**

BÚSQUEDA DE JERARQUÍAS

**TRANSFORMACIÓN DE
COLUMNAS CATEGÓRICAS**

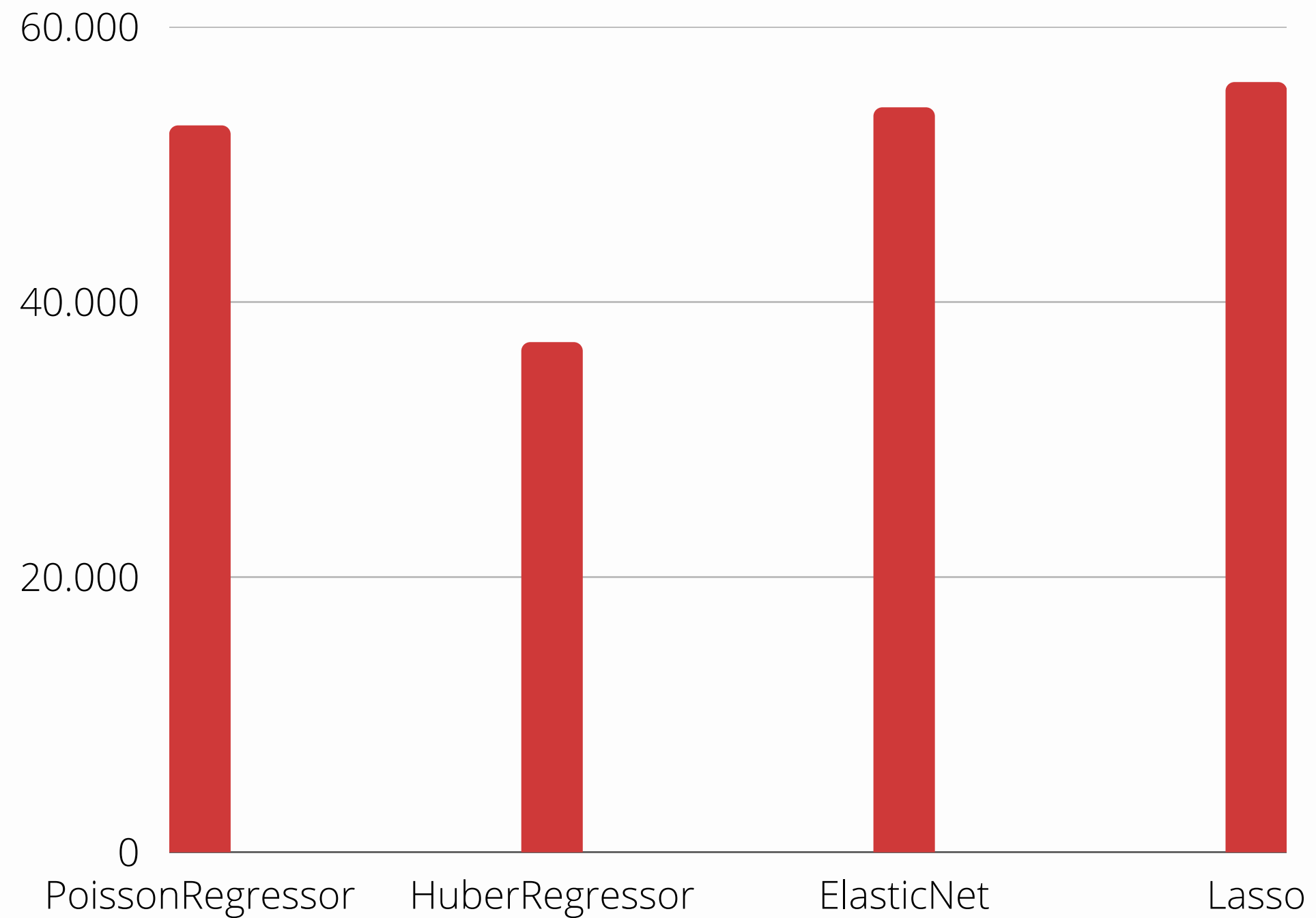


H2O Y LAZY

POCA EFICIENCIA CON H2O

LAZY NOS DA LA **CLAVE**

HUBER REGRESSOR



HIPERPARÁMETROS

HuberRegressor()

```
epsilon = [e + 0.01 for e in range(1, 2)]  
max_iter = [e for e in range(200)]  
alpha = [e + 0.0001 for e in range(200)]  
warm_start = [False, True]  
fit_intercept = [True, False]
```

EPSILON = 1.01

MAX_ITER = 18



Your Best Entry!

Your most recent submission scored 33510.29635, which is an improvement of your previous score of 41646.07344. Great job!

¡MUCHAS GRACIAS!

Autor: Pedro Jódar

Repositorio