

# Competición de Kaggle

Jaime Manrique Sabatel

# Índice

1

Introducción  
de la  
competición

2

Limpieza y  
Métodos

3

Imágenes del  
proceso

4

Análisis final

# Introducción

Para esta competición amistosa de clase se nos pedía emplear modelos de ML (machine learning) para predecir el salario de los trabajos de data.

Entre las normas y condiciones estaban:

- Preparar los datos para los diversos modelos (proceso empírico)
- Entrenar y Testear modelos de Machine Learning
- Subir los resultados con el mejor modelo entrenado de Machine Learning

# Métodos y Limpieza

La limpieza la lleve a cabo en pandas, nos daban dos dataframes. Lo primero que hice fue concatenarlos para realizar la misma limpieza en ambos. Después de limpiarlos los dividi para establecer mi X e Y para introducir en los modelos. Y finalmente subir mis resultados.

**1** Linear Regression: 51915 &  
62146

**2** Lasso: 52515

**3** Random Forest Regressor:  
54277

**4** Logistic Regression: 74249 &  
61087

# Procesos realizados

```
In [8]: data_num = pd.DataFrame(StandardScaler().fit_transform(total._get_numeric_data()), # standardize numeric columns
                                columns = total._get_numeric_data().columns).reset_index(drop=True)

data_obj = total.select_dtypes(include='object').reset_index(drop=True) # get categoric columns

data = pd.concat([data_num, data_obj], axis = 1)
```

```
In [10]: experience_level_ = {'MI':0, 'SE':1, 'EN':2, 'EX':3}
total['experience_level'] = total['experience_level'].map(experience_level_)
```

```
In [21]: total=pd.get_dummies(total, columns=['job_title', 'company_location'], drop_first = True)
total.shape
```

```
In [25]: plt.figure(figsize=(20, 8))

sns.heatmap(testeo_m.isna(),
            yticklabels=False,
            cmap='viridis',
            cbar=False)

plt.show()
```

# Análisis Final

Mi desarrollo a lo largo de la competición fue descendiente, mi primer modelo empleado fue mi mejor puntuación, después todas me empezaron a dar resultados menos precisos. Pese a varios intentos de mejorar la limpieza y utilizar los distintos metodos aprendidos en clase no obtuve mi objetivo de mejorar mi resultado.

# Muchas Gracias



Jaime Manrique Sabatel

Competición de Kaggle