Avveer's Linear Regression & Classification Submission For Kraggle (Behind The Scenes)

Link:

Github Reposistory

In [1]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import mean_squared_error, r2_score

train_df = pd.read_csv('train.csv')
test_df = pd.read_csv('test.csv')

train_df[['horsepower', 'weight']] = train_df[['horsepower', 'weight']].f
test_df[['horsepower', 'weight']] = test_df[['horsepower', 'weight']].fil

train_df.columns
```

Out[1]:
```
Index(['ID', 'mpg', 'cylinders', 'displacement', 'horsepower', 'weight',
       'acceleration', 'year', 'origin', 'name', 'mpg01'],
      dtype='object')
```

In [2]:
```
mpg_median = train_df['mpg'].median()
train_df['mpg01'] = (train_df['mpg'] > mpg_median).astype(int)
```

In [3]:
```
X_train = train_df.drop(columns=['mpg', 'ID', 'mpg01', 'name'])
y_train = train_df['mpg']

X_test = test_df.drop(columns=['ID', 'name'])
y_test = train_df['mpg']
```

In [4]:
```
scaler = StandardScaler()

X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

In [5]:
```
model = LinearRegression()
model.fit(X_train_scaled, y_train)

print("Coefficients:", model.coef_)
print("Intercept:", model.intercept_)
```

```
Coefficients: [-0.80659662  2.11362424 -0.69820862 -5.5301062   0.21351231
2.783416
  1.1510897 ]
Intercept: 23.51586901763224
```

In [6]:
```
y_pred = model.predict(X_test_scaled)
mse = mean_squared_error(y_test, y_pred)
```

```
r2 = r2_score(y_test, y_pred)
```

In [11]:
```
submission = pd.DataFrame({
    'ID': test_df['ID'],
    'mpg': y_pred})
submission.to_csv('submission.csv', index=False)
submission
```

Out[11]:

	ID	mpg
0	70_chevrolet chevelle malibu_alpha_3505	15.118178
1	71_buick skylark 320_bravo_3697	14.738560
2	70_plymouth satellite_charlie_3421	15.391297
3	68_amc rebel sst_delta_3418	13.641198
4	70_ford torino_echo_3444	14.912223
...	...	...
392	81_ford mustang gl_charlie_2802	27.387720
393	81_vw pickup_delta_2137	33.728600
394	82_dodge rampage_echo_2290	31.302963
395	81_ford ranger_foxtrot_2611	28.667094
396	84_chevy s-10_golf_2724	30.226448

397 rows × 2 columns

In [10]:
```
classification_submission = pd.DataFrame({
    'ID': test_df['ID'],
    'mpg0': train_df['mpg01']})
classification_submission.to_csv('classification_submission.csv', index=F
classification_submission
```

Out[10]:

	ID	mpg0
0	70_chevrolet chevelle malibu_alpha_3505	0
1	71_buick skylark 320_bravo_3697	0
2	70_plymouth satellite_charlie_3421	0
3	68_amc rebel sst_delta_3418	0
4	70_ford torino_echo_3444	0
...	...	...
392	81_ford mustang gl_charlie_2802	1
393	81_vw pickup_delta_2137	1
394	82_dodge rampage_echo_2290	1
395	81_ford ranger_foxtrot_2611	1
396	84_chevy s-10_golf_2724	1

397 rows × 2 columns

In [ ]: