RAMAIAH
Institute of Technology

Department of Information Science and Engineering

Deep Learning
ISE41

Assignment 2-Report

Predict Fuel Efficiency of a vehicle Using Deep Learning

Submitted by:
Pavana H A : 1MS19IS086
Prajwal Moger : 1MS19IS090

Submitted to:
Dr. Vijaya Kumar B P
Dept. of ISE, RIT

Predict Fuel Efficiency of a vehicle Using Deep Learning

Introduction:

The automotive industry is extremely competitive. With increasing fuel prices and picky consumers, automobile makers are constantly optimizing their processes to increase fuel efficiency.But, what if you could have a reliable estimator for a car's mpg given some known specifications about the vehicle?

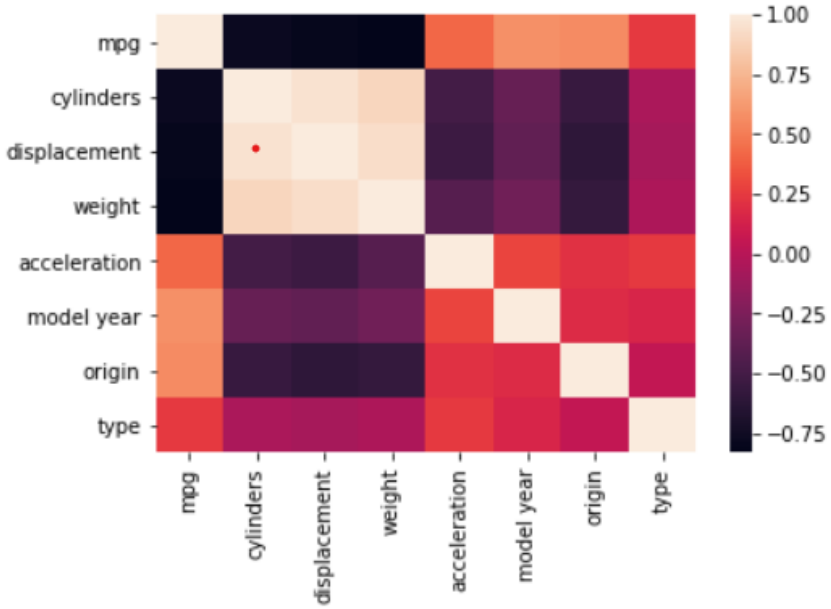
This notebook uses a public dataset from Kaggle.It contains the following columns: mpg, cylinders, horsepower, weight, acceleration, etc.

Data preprocessing:

- Checking Null Values: Dataset has no null values.
- We are removing the Car_name column since it doesn't add much value.

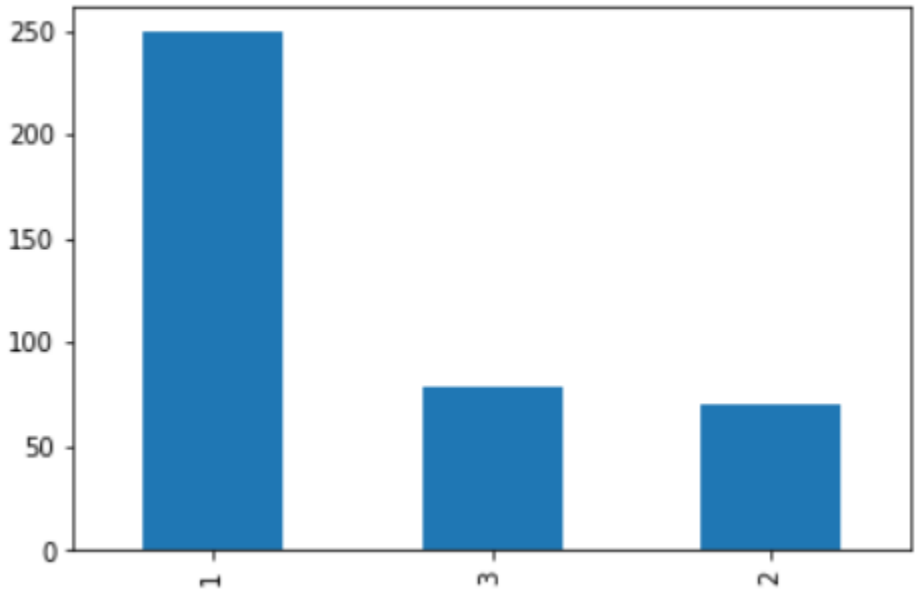
Exploratory data analysis:

Correlation map for this dataset:



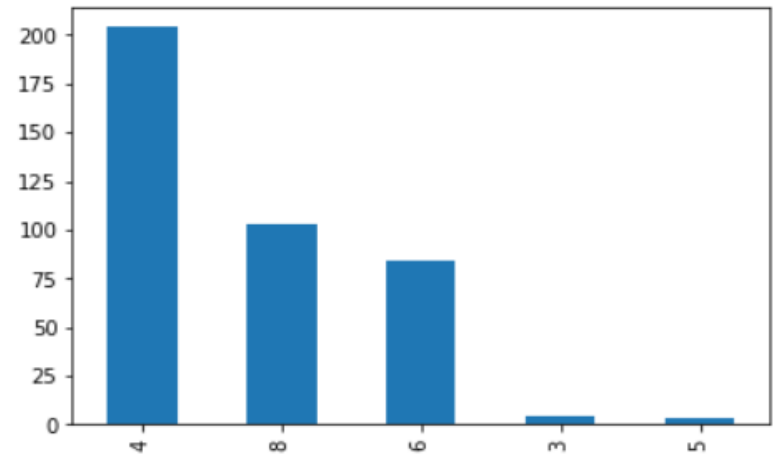
There are some strong correlations between each column. For cylinders, displacement, horse-power, and weight, it makes sense that the mpg would be negatively correlated with rising trends in any of the named features.

Number of cars based on their origin (US = 1, Asia = 2, Europe = 3):



- With 1 corresponding to American cars, we can see that the US accounts for the majority of the cars produced here.
- This may be a problem for our model, but if our accuracy is too low, then we can always normalize the presence of each area in the dataset to get predictions that aren't skewed towards American car mpg.

Distributions of different cylinder counts among our dataset:



- Notice how many cars have 4 cylinders versus 8/6 and 3/5. The V3 is an older engine design that was rarely used in production cars, while the V5 counts can be attributed to Volkswagen's VR5 engine design.
- Since the dataset uses pre-2000s cars, it makes sense how 4 cylinder cars are extremely popular. As time went on, the popularity of the SUV led to more cars having 6–8 cylinders in their engines.

Model training:we will be creating and training our model for predicting what a car's mpg will be. Since there are multiple algorithms we can use to build our model, we will compare the accuracy scores after testing and pick the most accurate algorithm.

We are using four different algorithms to train our model.

- XGBRegressor
- DecisionTreeRegressor
- RandomForestRegressor
- KNeighborsRegressor

we are using XGBoost, DecisionTree, RandomForest, and KNeighborsRegressor to perform our predictions. We will then see which algorithm produces the highest accuracy and select it as our algorithm of choice for future use.

Testing the model:we are doing regression testing, we use sklearn's built-in mean_squared_error() method and then use math.sqrt() to get the RMSE (root mean squared error).

From the RMSE values, we will select the *lowest* number since that algorithm has predicted the closest to the actual value.

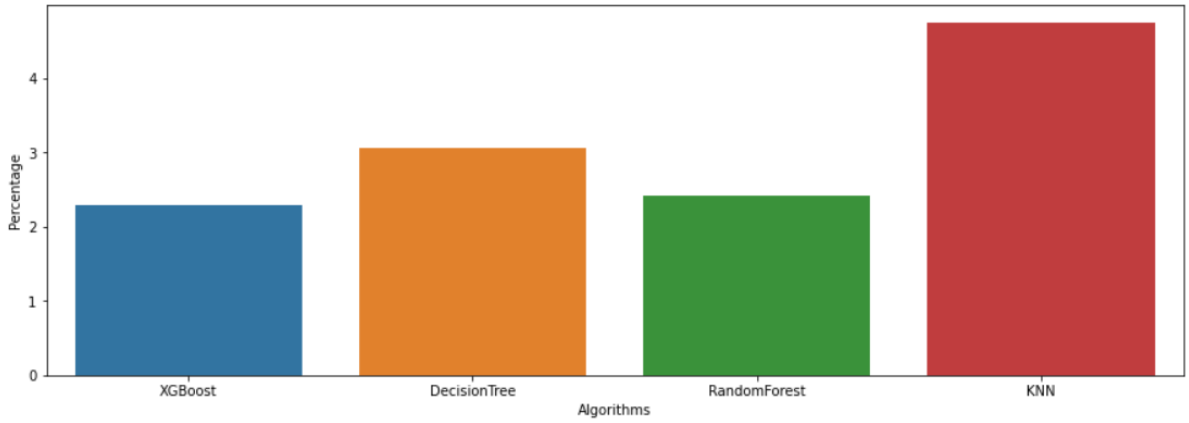
```
xgb_predict = xgb.predict(valid)
dtc_predict = dtc.predict(valid)
rfc_predict = rfc.predict(valid)
knn_predict = knn.predict(valid)

accuracy['XGBoost'] = math.sqrt(mean_squared_error(valid_label, xgb_predict))
accuracy['DecisionTree'] = math.sqrt(mean_squared_error(valid_label,dtc_predict))
accuracy['RandomForest'] = math.sqrt(mean_squared_error(valid_label,rfc_predict))
accuracy['KNN'] = math.sqrt(mean_squared_error(valid_label,knn_predict))
print(accuracy)
```

```
{'XGBoost': 2.2980936072558675, 'DecisionTree': 3.0668348815148474, 'RandomForest': 2.423530747837603, 'KNN': 4.754733091613742}
```

	label mpg	prediction
35	17.0	19.012188
357	32.9	26.992594
324	40.8	38.208786
132	25.0	21.423006
224	15.0	15.306767

	Algorithms	Percentage
0	XGBoost	2.298094
1	DecisionTree	3.066835
2	RandomForest	2.423531
3	KNN	4.754733



Conclusion:we built a model that could reliably predict a car's mpg given some information about the car within 2.3 mpg of the actual value.

This model could be trained with newer car data and be used to predict competitor's future mpg ratings for upcoming cars, allowing companies to potentially resources currently used on R&D today on making more efficient, more popular vehicles that outshine competitors.

While our model may be inaccurate in some cases, we talked about how our dataset can contain inaccurate values for the mpg, and oftentimes, our predictions are more accurate than the values in the dataset.