# Car Performance Prediction Using IBM Watson Machine Learning

Team Members:

Arikeshi Srivastava(19BCE10038,VIT Bhopal)  
 Adarsh Bhasker Srivastava(19BCE10060,VIT Bhopal)  
 Saarth Srivastava(19BCE10074,VIT Bhopal)  
 Mohit Soni (19BCE10133, VIT Bhopal)

**INTRODUCTION**

IBM Watson Machine Learning helps data scientists and developers accelerate AI and machine learning deployment on IBM Cloud Pak for Data. Deploy AI models at scale across any cloud on an open, extensible architecture.

With IBM you can:

1. Deploy any models including machine learning and deep learning models and decision optimization models.
2. Dynamically retrain models with continuous learning.
3. Automatically generate APIs to build AI-powered applications through DevOps.
4. Manage and monitor models for model drift, bias and risk.
5. Bring any projects into production, including open source, third-party and IBM tools

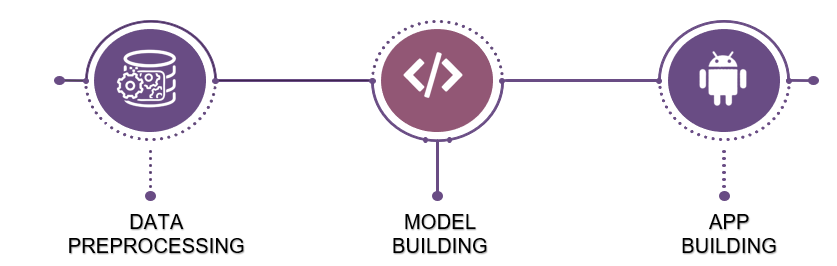
**PROBLEM STATEMENT**

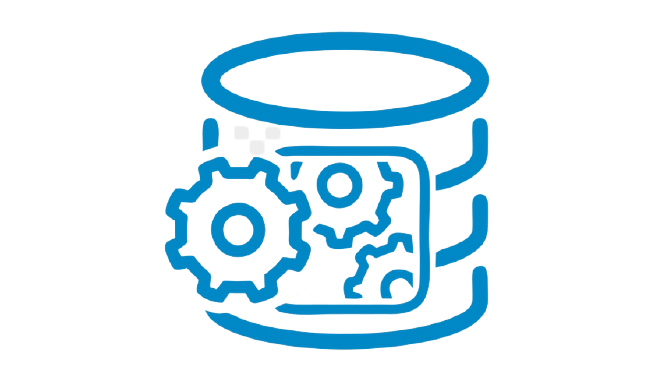
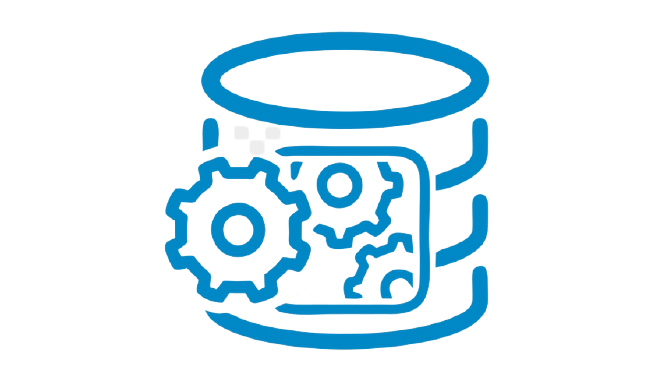
Predicting the performance level of cars is an important and interesting problem. The main goal is to predict the performance of the car to improve the certain behaviour of the vehicle. This can significantly help to improve the system's fuel consumption and increase efficiency.

The performance analysis of the car is based on the engine type, no of engine cylinders, fuel type, and horsepower, etc. These are the factors on which the health of the car can be predicted. It is an on-going process of obtaining, researching, analysing, and recording the health based on the above three factors. The performance objectives like mileage, dependability, flexibility, and cost can be grouped together to play a vital role in the prediction engine and engine management system. This approach is a very important step towards understanding the vehicle's performance.

**SOLUTION**

By using the given dataset to train machine learning model, model data from pkl file and h5 file to load data into python and flask to launch for development in app.



Dataset consists of parameters like displacement, acceleration, horsepower, cylinders, weight, origin, model year which helps determine performance and health of a car . So, the dataset is pre-processed and split into train and test data and trained using random forest classifier. The model predicts the mileage of car based on these given parameters. This can help to get better understanding of engine and engine management system and helps analyse the vehicle's performance in a better way.

Since the dataset involves independent input variables that is displacement, acceleration, horsepower, cylinders etc and one dependent variable that is mileage. Regression Machine Learning is made to calculate the car performance. The mileage of a car is not depended on the mileage of the other cars hence, Recurrent Neural Network cannot be used. The app will take in the input values and predict the mileage of the car.

**LITERATURE SURVEY**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| S.No. | Title | Author | Publication | Summary |
| 1. | Car Price Prediction using Machine Learning Techniques | Enis Gegic, Becir Isakovic, Dino Kečo, Zerina Mašetić, Jasmin Kevrić | UIKTEN - Association for Information Communication Technology Education and Science | To build a model for predicting the price of used cars three machine learning techniques were applied (ANN, SVM and Random Forest classifier).The data set was made and taken from the websites using web scraper that works based on PHP. The model was further integrated into Java application and gave an accuracy of 87.38% |
| 2. | Crash prediction based on traffic platoon characteristics using floating car trajectory data and the machine learning approach | Junhua Wang, Tianyang Luo,  Ting Fu | Accident Analysis & Prevention | A data preparation method, involving crash data filtering, floating car data filtering and data matching on the road network, is introduced for the safety analysis purpose. Results suggest that the traffic platoon information collected from floating cars accompanied works reasonably in predicting crashes on expressways. The support vector machine, with an overall accuracy of 85%, outperformed the binary logistic model which had an overall accuracy of 60%. |
| 3. | Machine learning or discrete choice models for car ownership demand estimation and prediction? | Miguel Paredes,Erik Hemberg,Una-May O'Reilly,Chris Zegras | IEEE | Discrete choice models are widely used to explain transportation behaviors, including a household's decision to own a car. They show how some distinct choice of human behavior or preference influences a decision. in contrast, machine learning models are derived to maximize prediction accuracy through mechanisms such as out-of-sample validation, non-linear structure, and automated covariate selection, albeit at the expense of interpretability and sound behavioral theory. |

**EXPERIMENTAL INVESTIGATIONS**

The experimental investigations include the search of the datasets with ample number of data to give a model whose r2\_score is high enough. When a model is made with the help of RandomForestsRegressor the r2\_score is given as 87%. Hence The required neural Network must also have a similar accuracy. The features of the model made are given below:

1) 7 neurons in the input layer because the input matrix has 7 columns.

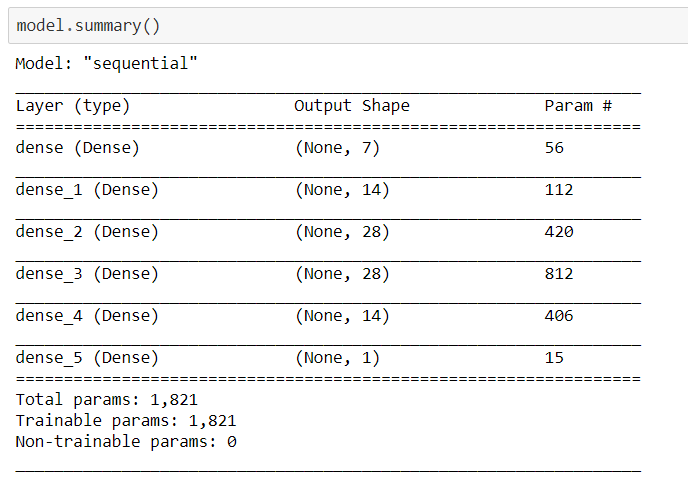
2) 4 hidden layer with 14,28,28,14 neurons repectively.

3) 1 output layer because the y matrix has 1 column.

4) All the layer have 'random\_ uniform' as initilizer and 'relu' activation function.

5) Optimizer: adam and loss: Mean\_Absolute\_Error loss function

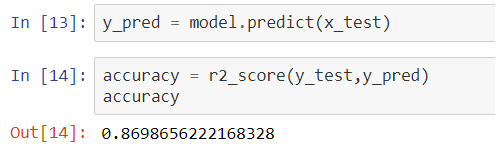
6) batch\_size: 24 and no. of epochs: 10000

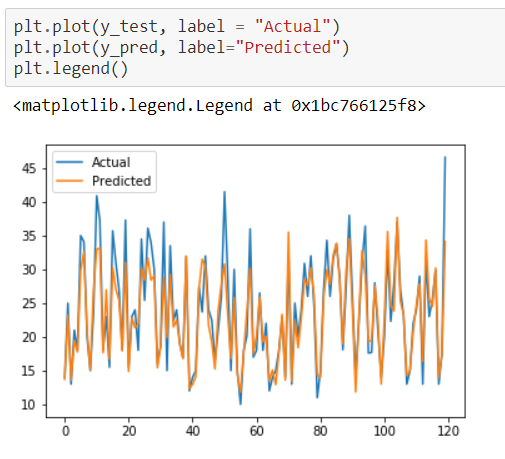


After Training the model the train vs test per epochs accuracy is given by



Accuracy of the predicted vales with the test data output





**SOFTWARE SPECIFICATIONS**

Anaconda (based on Python 3.7): Spyder, Jupyter notebook

App: Python Flask App Development, HTML and CSS Web page Development

**FLOWCHART**

1. Data Collection

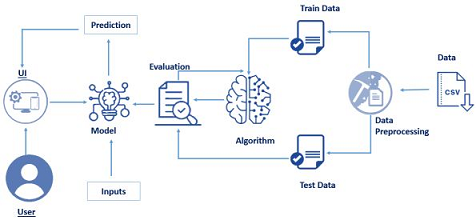
* Dataset is collected
* Data Pre-processing
* Libraries are imported.
* Dataset is imported.
* Checking for Null Values is done.
* Splitting Data into Train and Test.
* Feature Scaling.

1. Model Building

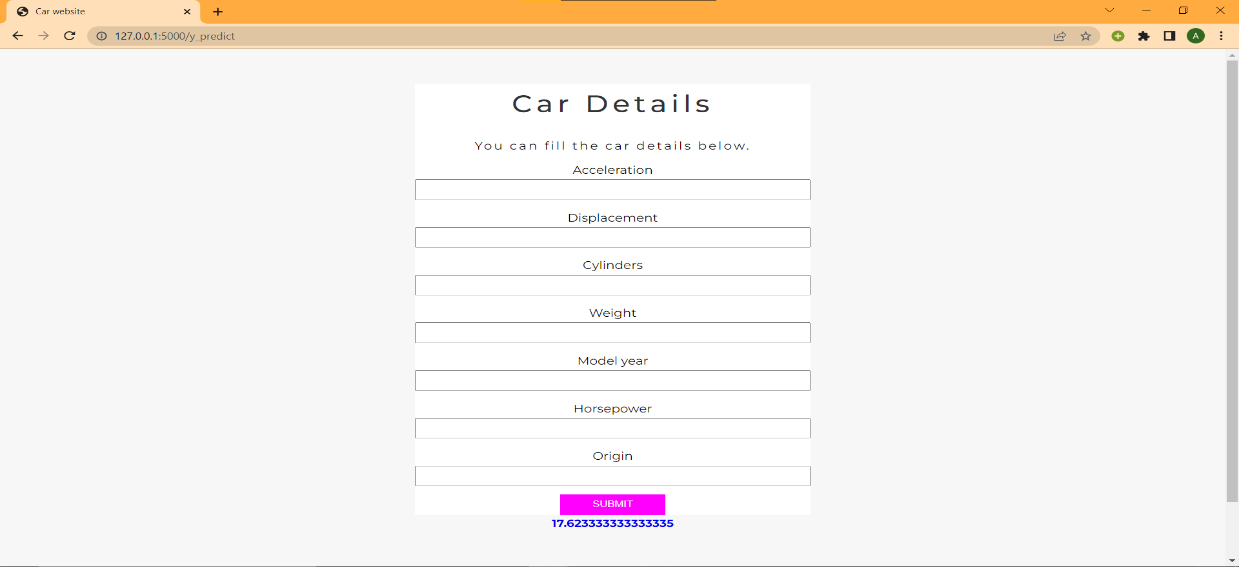
* Training and testing the model
* Evaluation of Model

1. Application Building

* Create an HTML file
* Build a Python Code



**APP SCREENSHOTS**



**CONCLUSION**

Regression model compiles the data and gives a clear relation between dependent and independent variables. Since the mileage of the car depends on the factors like weight, displacement, horsepower etc which are independent of each other, regression model acts as a better ML algorithm. Recurrent Neural Network cannot be used because the mileage of one car doesnt depends on the mileage of the other car.

This is a regression kind of problem since mileage has been predicted based on other numerical parameters. This model has been trained using random forest classifier and when data was tested for this model it gave an accuracy of around 87%. The prediction of mileage gives a better understanding of performance and usage of the car. The app helps by providing a user-friendly interface for the users to provide the input data and obtain the required mileage.

**FUTURESCOPE**

Although, this system has achieved astonishing performance in car mileage prediction problem our aim for the future research is to test this system to work successfully with various data sets with high number of input values to gain more appropriate results. Our future scope will be to explore more dense and complex neural network models to find a model which gives a very high accuracy and exact predictions of mileage for determining car performance.

**BIBLIOGRAPHY**

1. <http://www.sthda.com/english/wiki/regression-analysis-essentials-for-machine-learning>
2. <https://www.sciencedirect.com/science/article/abs/pii/S0001457519307468>
3. <https://ieeexplore.ieee.org/abstract/document/8005618>
4. <https://smartinternz.com/>