CAR RESALE VALUE PREDICTION USING MACHINE LEARNING ALGORITHMS

MADDUKURI NIVAS (19BCE1010),
KOVVURI UDAY SURYA DEVESWAR REDDY (19BCE1253)
DR. SALEENA B | SCSE

MOTIVATION AND INTRODUCTION:

The objective of the project is to create an algorithm that predict the value of the used cars located in different places of India. We will implement this using two different algorithms they are Random forest Regression, support vector Regression.

SCOPE OF THE PROJECT:

The idea evolved from the fact that the resale value of the car differ from area to area in the same state, this is because of the company. So, by this predictor the seller can check the resale value by entering all the details correctly. And the seller can sell at the correct price. So, that there will be no misleading.

METHODOLOGY:

RANDOM FOREST REGRESSION

The random forest model is a type of additive model that makes predictions by combining decisions from a sequence of base models. More formally we can write this class of models as:

$g(x) = f0(x) + f1(x) + f2(x) +...$
Where the final model g is the sum of simple base models fi. Here, each base classifier is a simple decision tree. This broad technique of using multiple models to obtain better predictive performance is called model ensembling. In random forests, all the base models are constructed independently using a different subsample of the data.

SUPPORT VECTOR REGRESSION

Support Vector Regression (SVR) is quite different than other Regression models. It uses the Support Vector Machine (SVM, a classification algorithm) algorithm to predict a continuous variable. While other linear regression models try to minimize the error between the predicted and the actual value, Support Vector Regression tries to fit the best line within a predefined or threshold error value.

Assuming that the equation of the hyperplane is as follows: y = nx + b

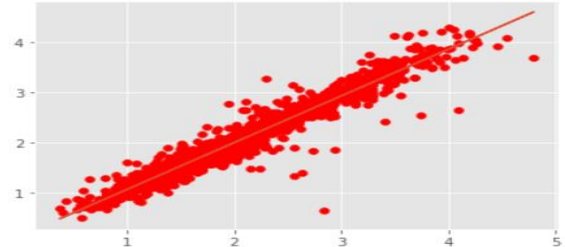The lines will be nx + b=+a, nx + b = -a

Thus, any hyperplane that satisfies our SVR should satisfy: -a < y – nx + b < +a

RESULTS:

predicting test set values using random forest

```
y_pred = regr.predict(X_test)
print(y_pred);

[2.21700654 1.47785114 1.62545259 ... 2.50062231 1.90864515 1.7904186 ]
```
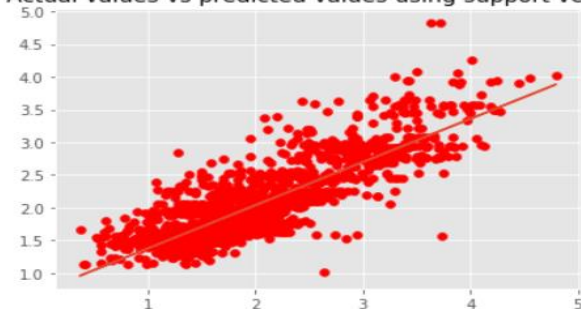
Actual values vs predicted values using random forest

Predicting test set values using support vector

```
y_pred = regressor.predict(X_test)
print(y_pred);

[2.48783871 1.58652453 1.96866888 ... 2.2411862  1.66746063 1.67121376]
```

Actual values vs predicted values using support vector

Accuracy Scores
Random forest Regression:

```
R-squared score (training): 0.986
R-squared score (testing): 0.932
```

Support vector regression:

```
regressor.score(X_test,y_test)

0.683128821493263
```

CONCLUSION:

We implemented the car resale value model using random forest regression and support vector regression algorithm, and we found that random forest regression model gave the more accuracy score of 93.56%. So, random forest regression is best fit for the dataset we taken.

ACKNOWLEDGEMENT: