Car Price Prediction Using Regression Techniques of Machine Learning

Ms. P.Jasmine, Ms. Sk.Shefa Kouser, Mr. P.Bharath Kumar, Mr. Sk.Shareef

Under The Esteemed Guidance Of Mrs. Sk. Shammi M.Tech Assistant Professor

Department Of Computer Science and Engineering

Bachelor Of Technology

Tirumala Engineering College

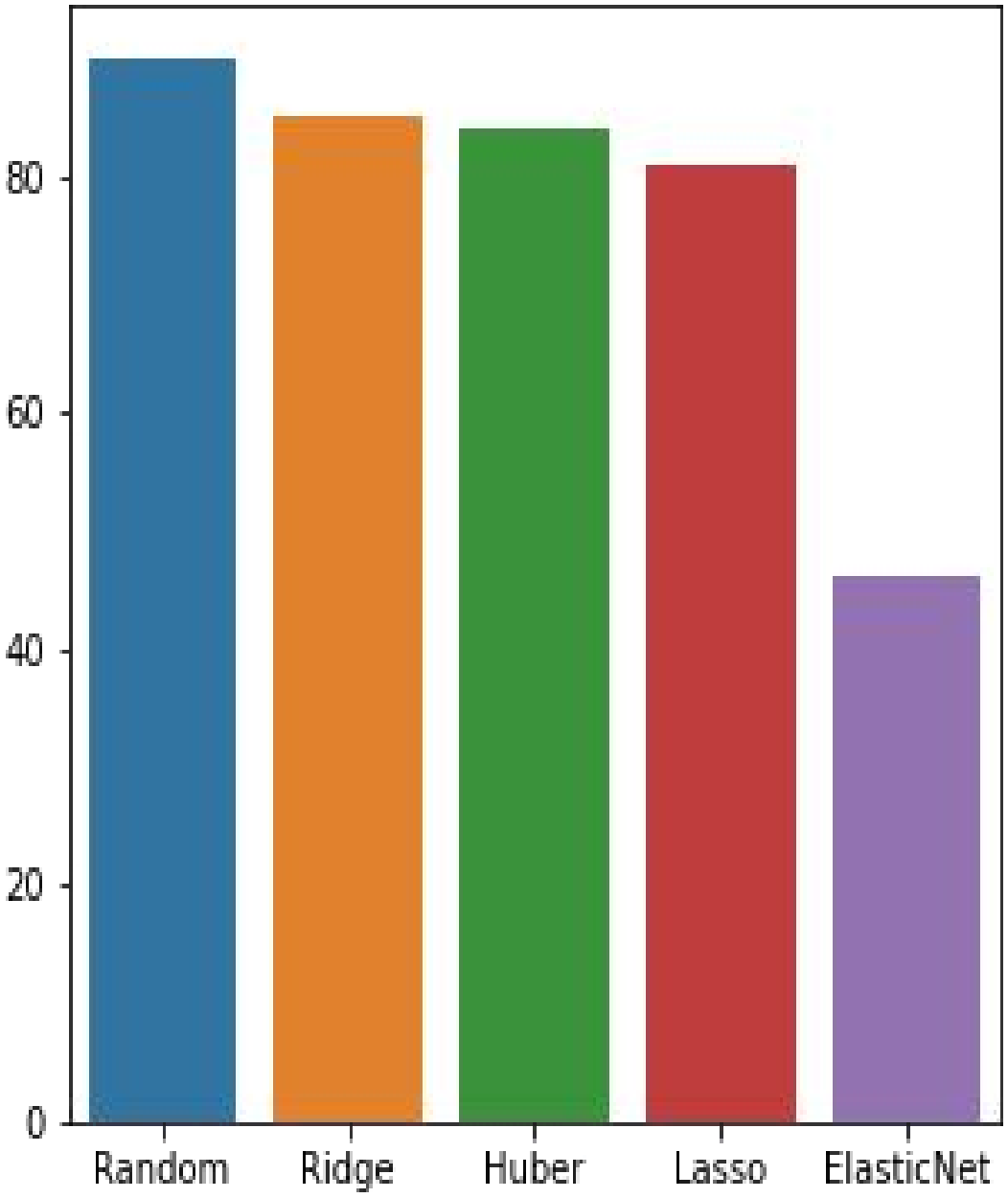
Jonnalagadda, Narasaraopet, Guntur (Dt), A.P.

ABSTRACT

A car price prediction has been a high-interest research area, as it requires noticeable effort and knowledge of the field expert. Considerable number of distinct attributes are examined for the reliable and accurate prediction. To build a model for predicting the price of used cars in Bosnia and Herzegovina, we applied three machine learning techniques (Artificial Neural Network, Support Vector Machine and Random Forest). However, the mentioned techniques were applied to work as an ensemble. The data used for the prediction was collected from the web portal autopijaca.ba using web scraper that was written in PHP programming language. Respective performances of different algorithms were then compared to find one that best suits the available data set. The final prediction model was integrated into Java application. Furthermore, the model was evaluated using test data and the accuracy of 91% was obtained.

1.1 Introduction to Project

Machine learning is one of the applications of artificial intelligence (AI) that provides computers, the ability to learn automatically and improve from experience instead of explicitly programmed. It focuses on developing computer programs that can access data and use it to learn from themselves. The main aim is to allow computers to learn automatically without human intervention and also adjust actions accordingly. Accurate car price prediction involves expert knowledge, because price usually depends on many distinctive features and factors. Typically, most significant ones are brand and model, age, current price and fuel type. Car prices had a great deal of attention in automobile research. The prediction of car price is a challenging task, which can offer automated prediction about the car.

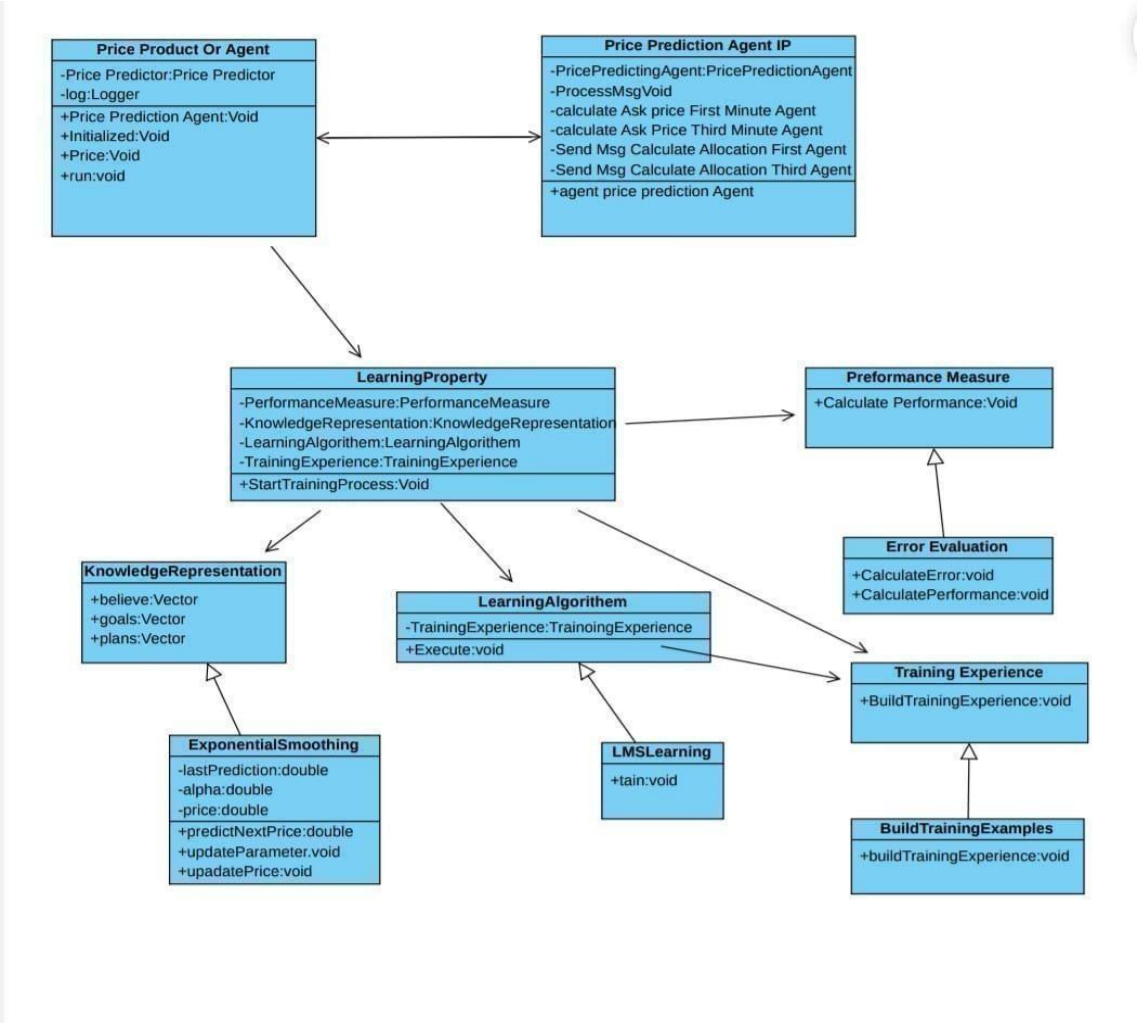


Comparison of accuracy of algorithm

UML DIAGRAMS

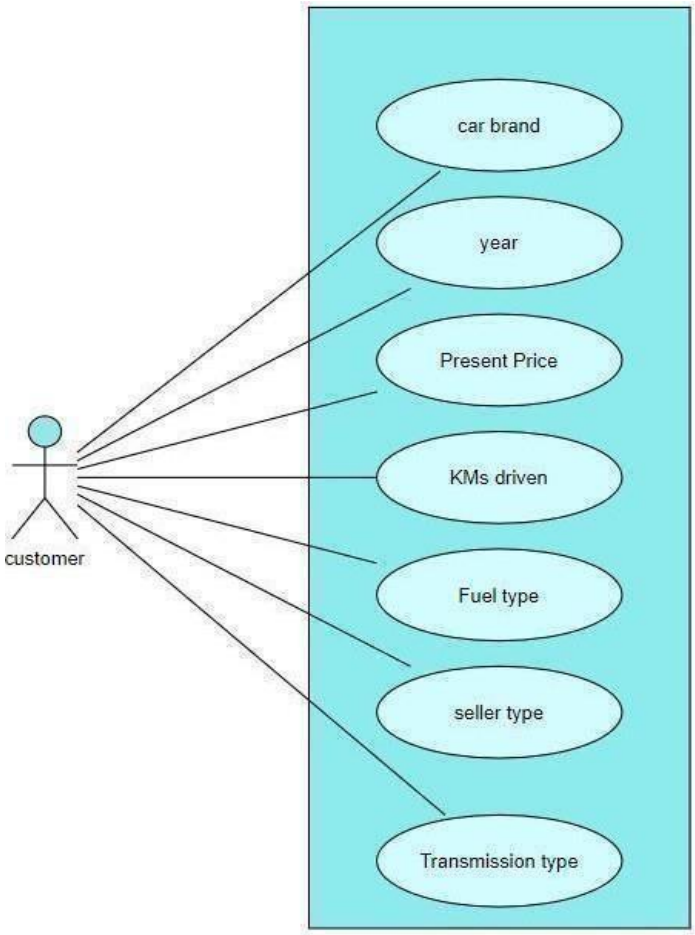
3.1.1 Class Diagram:

In software engineering, a class diagram in the Unified Modeling Language (UML) is a type of static structure diagram that describes the structure of a system by showing the system's classes, their attributes, operations (or methods), and the relationships among the classes. It explains which class contains information.



3.1.2 Use Case Diagram :

In UML, use-case diagrams model the behavior of a system and help to capture the requirements of the system. Use-case diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in use-case diagrams describe what the system does and how the actors use it, but not how the system operates internally.



3.1.3 Sequence Diagram :

A sequence diagram in Unified Modeling Language (UML) is a kind of interaction diagram that shows how processes operate with one another and in what order. It is a construct of a Message Sequence Chart. Sequence diagrams are sometimes called event diagrams, event scenarios, and timing diagrams.

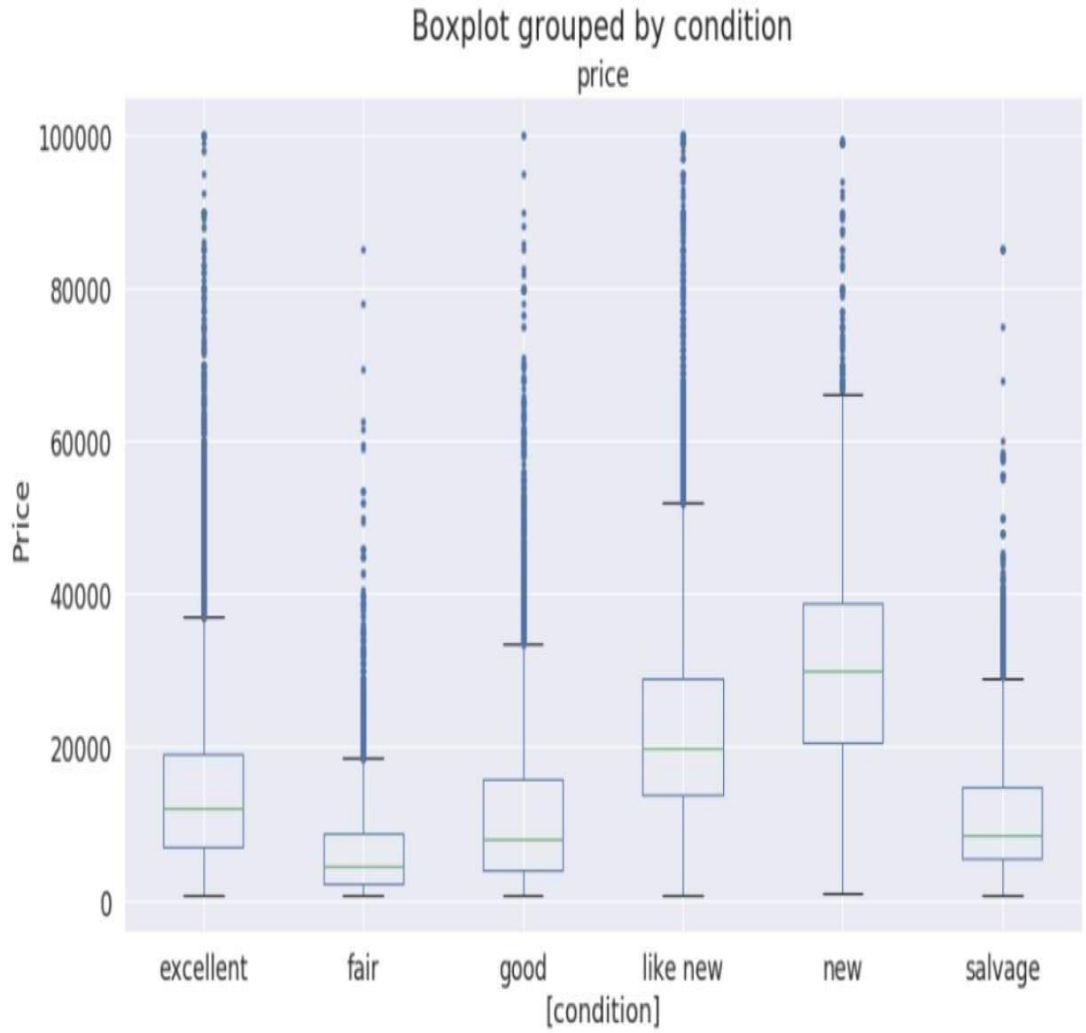
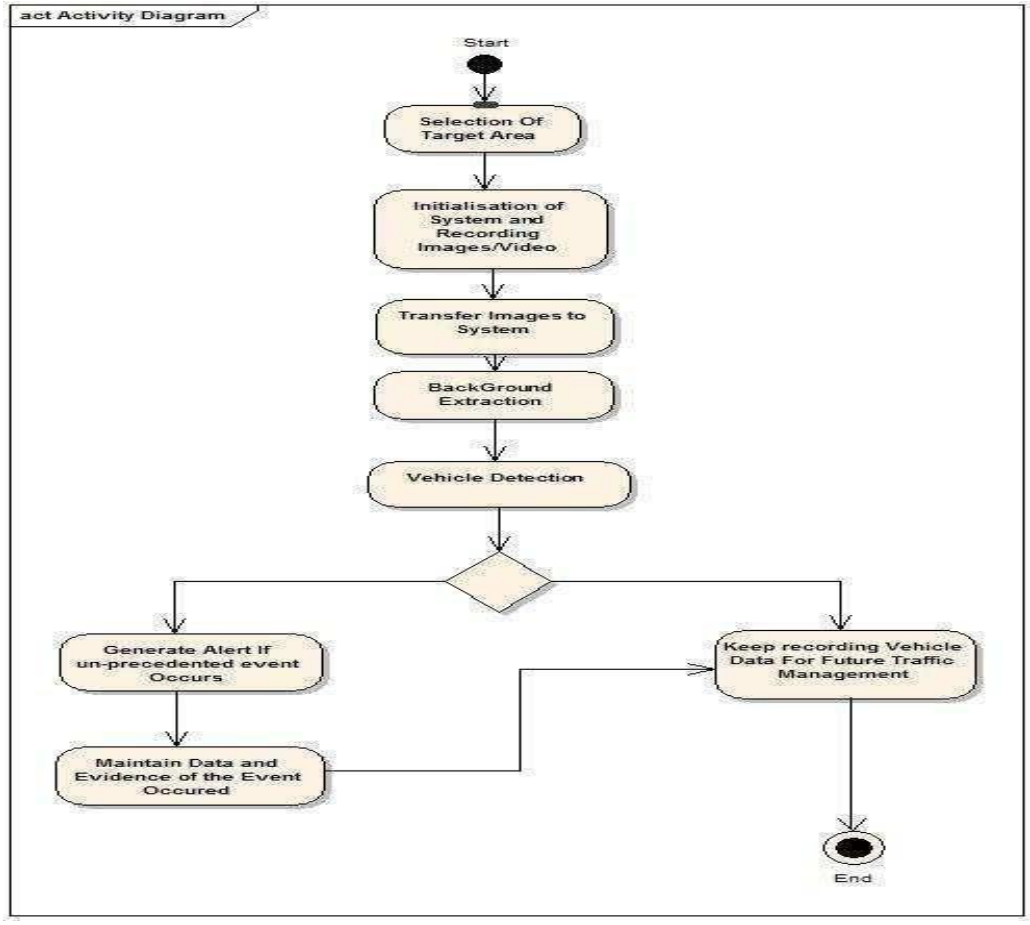


Figure 10: Boxplot of Condition by Price

3.1.4 ACTIVITY DIAGRAM:

Activity diagrams are graphical representations of workflows of step wise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step- by-step workflows of components in a system. An activity diagram shows the overall flow of control.



4.1 Sample Code

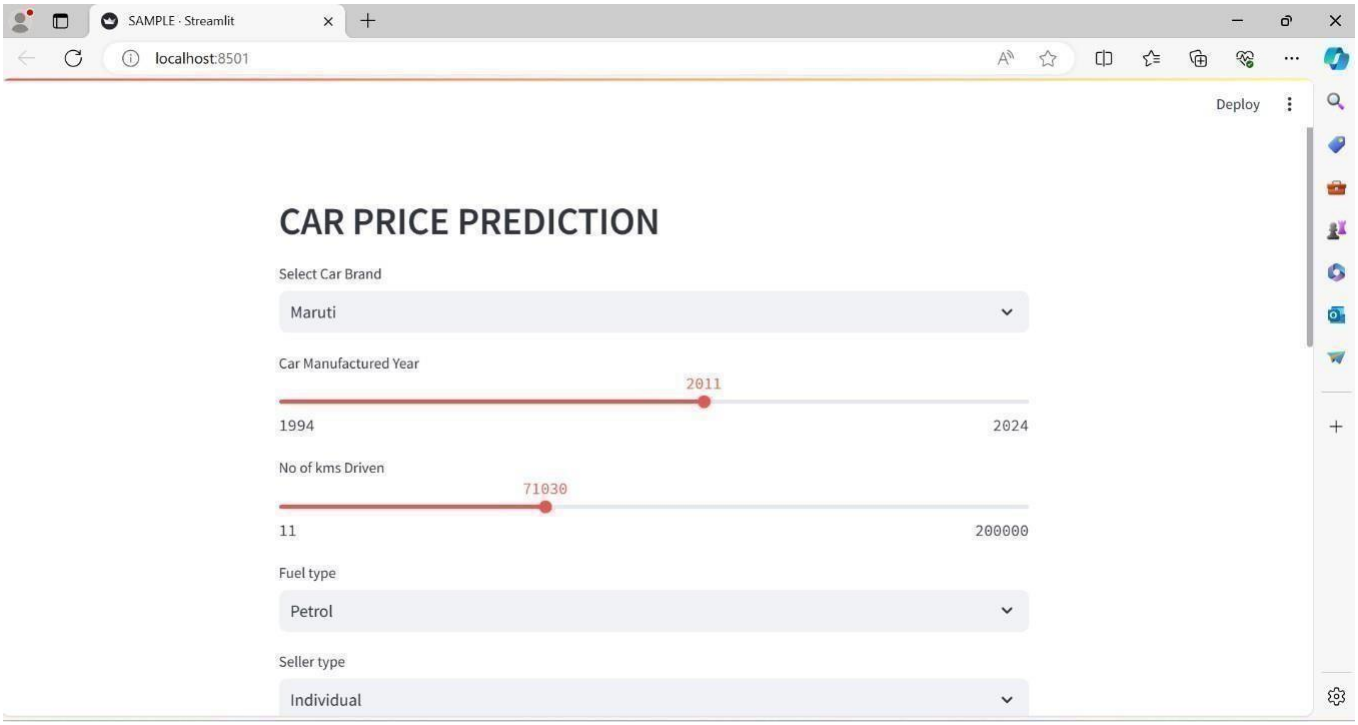
.py code

```
import pandas as pd import numpy as np import pickle as pk import streamlit as st
model = pk.load(open('model (1).pkl','rb')) st.header('CAR PRICE PREDICTION')
cars_data = pd.read_csv('DATA_SET (1).csv') def get_brand_name(car_name):
car_name = car_name.split(' ')[0] return car_name.strip()
cars_data['name'] = cars_data['name'].apply(get_brand_name) name = st.selectbox('Select Car Brand', cars_data['name'].unique()) year = st.slider('Car Manufactured Year', 1994,2024)
km_driven = st.slider('No of kms Driven', 11,200000) fuel = st.selectbox('Fuel type', cars_data['fuel'].unique())
seller_type = st.selectbox('Seller type', cars_data['seller_type'].unique()) transmission = st.selectbox('Transmission type', cars_data['transmission'].unique()) owner = st.selectbox('Seller type', cars_data['owner'].unique())
mileage = st.slider('Car Mileage', 10,40) engine = st.slider('Engine CC', 700,5000) max_power = st.slider('Max Power', 0,200)
seats = st.slider('No of Seats', 5,10
if st.button("Predict"): input_data_model = pd.DataFrame(
[[name,year,km_driven,fuel,seller_type,transmission,owner,mileage,engine,max_power,seats]],
columns=['name','year','km_driven','fuel','seller_type','transmission','owner','mileage','engine','max_p ower','seats'])
input_data_model['owner'].replace(['First Owner', 'Second Owner', 'Third Owner', 'Fourth & Above Owner', 'Test Drive Car'],[1,2,3,4,5], inplace=True)
input_data_model['fuel'].replace(['Diesel', 'Petrol', 'LPG', 'CNG'],[1,2,3,4], inplace=True)
input_data_model['seller_type'].replace(['Individual', 'Dealer', 'Trustmark Dealer'],[1,2,3], inplace=True)
input_data_model['transmission'].replace(['Manual', 'Automatic'],[1,2], inplace=True)
input_data_model['name'].replace(['Maruti', 'Skoda', 'Honda', 'Hyundai', 'Toyota', 'Ford', 'Renault',
'Mahindra', 'Tata', 'Chevrolet', 'Datsun', 'Jeep', 'Mercedes-Benz',
'Mitsubishi', 'Audi', 'Volkswagen', 'BMW', 'Nissan', 'Lexus',
'Jaguar', 'Land', 'MG', 'Volvo', 'Daewoo', 'Kia', 'Fiat', 'Force', 'Ambassador', 'Ashok', 'Isuzu', 'Opel'],
[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31]
,inplace=True)
car_price = model.predict(input_data_model) st.markdown('Car Price is going to be '+ str(car_price[0])
```

.ipynb code

```
Import pandas as pd import numpy asnp
from sklearn.model_selection import train_test_split from sklearn.linear_model import LinearRegression
cars_data =
```

Output Screens :



Home Page

SYSTEM TESTING

4.2 TESTING OBJECTIVE

Software testing is a process used to help identify the correctness, completeness and quality of developed computer software.

Software testing is the process used to measure the quality of developed software. Testing is the process of executing a program with the intent of finding errors. Software testing is often referred to as verification & validation.

STLC (Software Testing Life Cycle):

Testing itself has many phases i.e., is called as STLC. STLC is part of SDLC

- Test Plan
- Test Development
- Test Execution
- Analyze Result
- Defect Tracking

TEST PLAN

It is a document which describes the testing environment, purpose, scope, objectives, test strategy, schedules, mile stones, testing tool, roles and responsibilities, risks, training, staffing and who is going to test the application, whattype of tests should be performed and how it will track the defects.

TESTDEVELOPMENT

Preparing test cases, test data, Preparing test procedure, Preparing test scenario,Writingtest script.

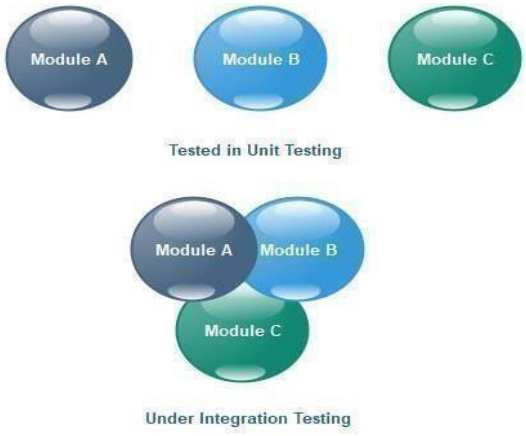
TESTING ACTIVITIES

- **Unit testing**

Initialization testing is the first level of dynamic testing and is the first responsibility of developers and then that of the test engineers. Unit testing is performed after the expected test results are met or differences are explainable/acceptable.

Unit testing helps tester and developers to understand the base of code that makes them able change defect causing code quickly. Unit testing helps in the documentation. Unit testing fixes defects very early in the development phase that's why there is a possibility to occur a smaller number of defects in upcoming testing levels. It helps with code reusability by migrating code and test cases.

Integration testing



All module which make application are tested. Integration testing is to make sure that the interaction of two or more components produces results that satisfy functional requirement

6. CONCLUSION

We have used 5 different algorithms like random forest regressor, Ridge regressor, Huber Regressor, Lasso regressor, ElasticNet CV Regressor for the prediction of the price of the cars. out of these algorithms the accuracy value of random forest regressor is 91%, the accuracy value of Ridge regressor is 89%, the accuracy value of Huber Regressor is 88%, the accuracy value of Lasso regressor is 80% and the accuracy value of Elastic Net CV is 46%. Hence Random forest Regressor exhibits the higher accuracy value out of any of the other values. Hence random forest regression algorithm is used for the prediction of the car prices.

7. FUTURE ENHANCEMENTS

In future this machine learning model may bind with various website whichcan provide real time data for price prediction. Also, we may add large historical data of car price which can help to improve accuracy of the machine learning model. We can build an android app as user interface for interacting with user. For better performance, we plan to judiciously design deep learning network structures, use adaptive.

BIBLOGRAPHY

Theobald, O. (2017). Machine learning for absolute beginners

S. Abhishek.:Ridge regression vs Lasso, How these two popular ML Regression techniques work. Analytics India magazine,2018.

Minaie, N. (2019). A Beginner's Guide to Selecting Machine Learning Predictive Models in Python, https://towardsdatascience.com/the-beginners-guide-to- selecting- machine-learning-predictive-models- in-python-

https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho

https://www.greeksforgreeks.org/

https://www.w3schools.com/

https://www.javatpoint.com/regression-analysis-in-machine-learning

https://www.youtube.com/results?search_query=siddhardhan

https://www.google.com/search?q=siddhardhan&rlz=1C1CHBD_enIN972IN972&oq=siddardhan&aqs=chrome.1.69i57j0i10i512l5.9269j0j4&sourceid=chrome&ie=UTF-8

https://www.google.com/search?q=car+price+prediction+data+set&rlz=1C1CHBD_enIN972IN972&sxsrf=APwXEdeG-R1yPC51p-d4jtTD1ejsRAwMEw%3A1681138890161&ei=yiQ0ZK-1CaHv4-EPjs6c6Ac&oq=car+price+prediction+data+&gs_lcp=Cgxnd3Mtd2l6LXNlcnAQARgAMgUIABCABDIGCAAQFhAeMggIABCKBRCGAzIICAAQigUQhgMyCAgAEIoFEIYDOgoIABBHENYEELADOhAILhCKBRDHARDRAxCwAxBDOgoIABCKBRCwAxBDOgcIABCKBRBDSgQIQRgAUIQEWMEQYIYfaAFwAXgAgAHNAYgBmQiSAQUwLjUuMZgBAKABAcgBCsABAQ&sclient=gws-wiz-serp