CAR RESALE VALUE PREDICTION

1. INTRODUCTION

Determining whether the listed price of a used car is a challenging task, due to the many factors that drive a used vehicle's price on the market. The focus of this project is developing machine learning models that can accurately predict the price of a used car based on its features, in order to make informed purchases. We implement and evaluate various learning methods on a dataset consisting of the sale prices of different makes and models .Depending on various parameters we will determine the price of the car. Regression Algorithms are used because they provide us with continuous value as an output and not a categorized value because of which it will be possible to predict the actual price a car rather than the price range of a car. User Interface has also been developed which acquires input from any user and displays the Price of a car according to user's inputs.

1.1 PROJECT OVERVIEW

- Able to understand the problem to classify if it is a regression or a classification kind of problem.
- Able to know how to pre-process/clean the data using different data pre-processing techniques.
- Applying different algorithms according to the dataset.
- Able to know how to evaluate the model.
- Able to build web applications using the Flask framework.

1.2 PURPOSE

This project aims to deliver price prediction models to the public, to help guide the individuals looking to buy or sell cars and to give them a better insight into the automotive sector. Buying a used car from a dealer can be a frustrating and an unsatisfying experience as some dealers are known to deploy deceitful sale tactics to close a deal. Therefore, to help consumers avoid falling victims to such tactics, this study hopes to equip consumers with right tools to guide them in their shopping experience.

2 LITERATURE SURVEY

With the recent arrival of internet portals, buyers and sellers may obtain an appropriate status of the factors that ascertain the market price of a used automobile. Lasso Regression, Multiple Regression, and Regression Trees are examples of machine learning algorithms. We will try to develop a statistical model that can forecast the value of a pre-owned automobile based on prior customer details and different parameters of the vehicle. [2] This paper aims to compare the efficiency of different models' predictions to find the appropriate one.

On the subject of used automobile price prediction, several previous studies have been conducted. To anticipate the value of pre-owned automobiles in Mauritius, Pudaruth employed naive Bayes, k-nearest neighbours, multiple linear regression, and decision trees. However, because there were fewer cars observed, their results were not good for prediction. In his article, Pudaruth concluded that decision trees and naive Bayes are ineffective for continuous-valued variables.[4]

To anticipate the price of a vehicle, Noor and Jan employed multiple linear regression. They used a variable selection methodology to determine the variables that had the highest

influence and then eliminated the remainder. Only a few variables are included in the data, which were utilised to create the linear regression model. With an R-square of 98 percent, the outcome was outstanding. [4]

Peerun et al. conducted study to assess the neural network's performance in predicting used automobile prices. However, especially on higher-priced cars, the estimated value is not very close to the real price. In forecasting the price of a used car, they found that support vector machine regression outperformed neural networks and linear regression by a little margin. [4]

To accurately anticipate the price of a car, many different approaches have been used in the digital world, ranging from machine learning approaches like multiple linear regression, k-nearest neighbor, and naive bayes to random forest and decision tree to the SAS enterprise miner. In [7], [8], [9], [10] and [11] all of these solutions took into account distinct sets of attributes when making predictions based on the historical data used to train the model.

2.1 EXISTING PROBLEM

Using various machine learning algorithms we will predict the price. The algorithms involve Ridge Regression and Lasso Regression. The best model which predicts the most accurate price is selected. After selection of the best model the predicted price is displayed to the user according to user's inputs. User can give input through website to for used car price prediction to machine learning model.

Ridge Regression

A Ridge regressor is basically a regularized version of Linear Regressor. The regularized term has the parameter 'alpha' which controls the regularization of the model i.e helps in reducing the variance of the estimates.

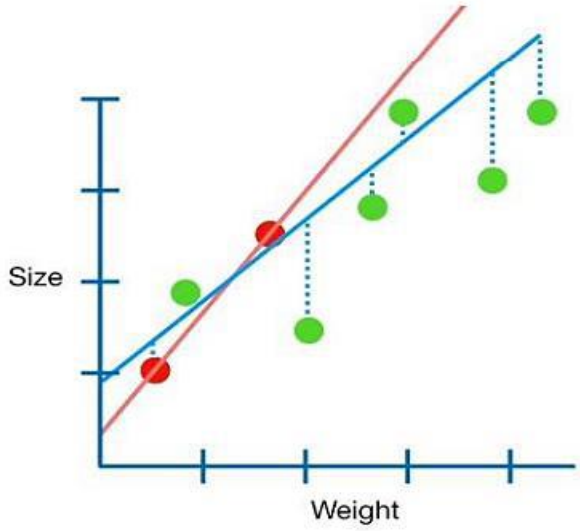


Fig – Ridge Regression

Lasso Regression

The “LASSO” stands for Least Absolute Shrinkage and Selection Operator. Lasso regression is a regularization technique. It is used over regression methods for a more accurate prediction. This model uses shrinkage. Shrinkage is where data values are shrunk towards a central point as the mean. The lasso procedure encourages simple, sparse models (i.e. models with fewer parameters). This particular type of regression is well-suited for models showing high levels of multicollinearity or when you want to automate certain parts of model selection, like variable selection/parameter elimination.

2.2 REFERENCES

[1] Doan Van Thai, “Prediction car prices using quantify qualitative data and knowledge-based system.”
[2] Pattabiraman Venkatasubbu, “Used Cars Price Prediction using Supervised Learning Techniques.”
[3] Nitis Monburinon, “Prediction of Prices for Used Car by Using Regression Models”
[4]https://towardsdatascience.com/used-car-priceprediction-using-machine-learninge3be02d977b2
[5]https://www.semanticscholar.org/paper/vehiclePrice-Prediction-System-using-Machine-NoorJan/fc87ead6754b188b1b8629db77badf361fd24a22
[6] https://www.docsity.com/en/research-projectproposal-online-car-rental-system/5232831/
[7] Comparative Analysis of Used Car Price Evaluation Models, Tongji University, Shanghai 200000, China.
[8] Nitis Monburinon, “Prediction of Prices for Used Car by Using Regression Models”, 5th International Conference on Business and Industrial Research, (ICBIR), Bangkok, Thailand, 2018 [9] Jaideep A Muley, “Prediction of Used Cars’ Prices by Using SAS EM”, Oklahoma State University
[10]Nabarun Pal, “A methodology for predicting used cars prices using Random Forest”, Future of Information and Communications Conference, 2018
[11]Kuiper, Shonda, "Introduction to Multiple Regression: How Much Is Your Car Worth?" - Journal Of Statistics Education, 2008

2.3 PROBLEM STATEMENT AND DEFINITION

The prices of new cars in the industry is fixed by the manufacturer with some additional costs incurred by the Government in the form of taxes. So customers buying a new car can be assured of the money they invest to be worthy. But due to the increased price of new cars and the incapability of customers to buy new cars due to the lack of funds, used cars sales are on a global increase. Predicting the prices of used cars is an interesting and much-needed problem to be addressed. Customers can be widely exploited by fixing unrealistic prices for the used cars and many falls into this trap. Therefore, rises an absolute necessity of a used car price prediction system to effectively determine the worthiness of the car using a variety of features. Due to the adverse pricing of cars and the nomadic nature of people in developed countries, the cars are mostly bought on a lease basis, where there is an agreement between the buyer and seller. These cars upon completion of the agreement are resold. So reselling has become an essential part of today’s world.
The main aim of this project is to predict the price of used cars using the various Machine Learning (ML) models. This can enable the customers to make decisions based on different inputs or factors namely Brand or Type of the car one prefers like Ford, Hyundai, Model of the

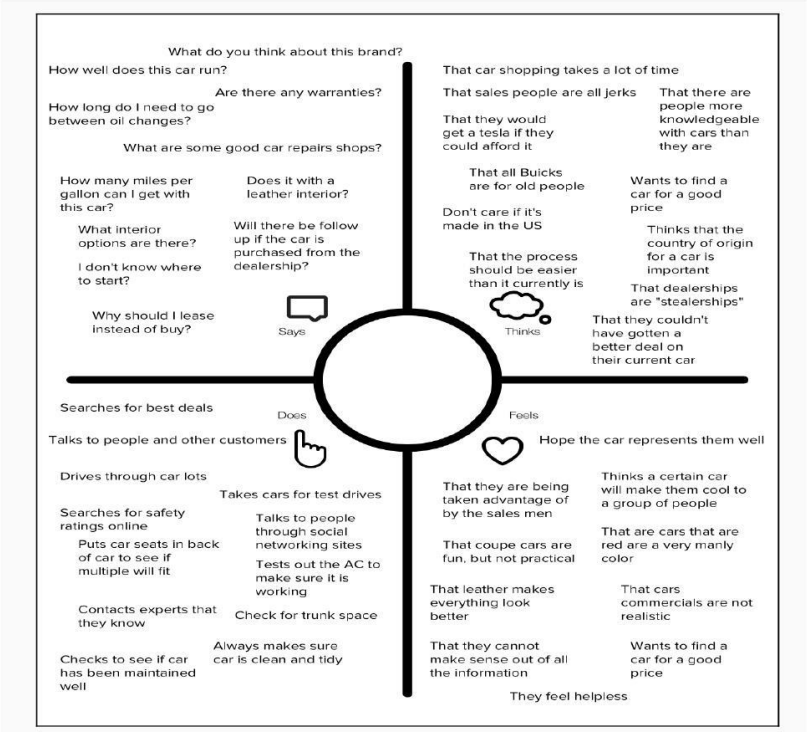
car namely Ford Figo, Hyundai Creta, Year of manufacturing like 2020, 2021, Type of fuel namely Petrol, Diesel, Price range or Budget, Type of transmission which the customer prefers like Automatic or Manual, Mileage to name a few characteristic features required by the customer. This project Car Price Prediction deals with providing the solution to these problems. Different techniques like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions. The predictions are then evaluated and compared in order to find those which provide the best performances.

3 IDEATION AND PROPOSED SOLUTION

3.1 EMPATHY MAP CANVAS

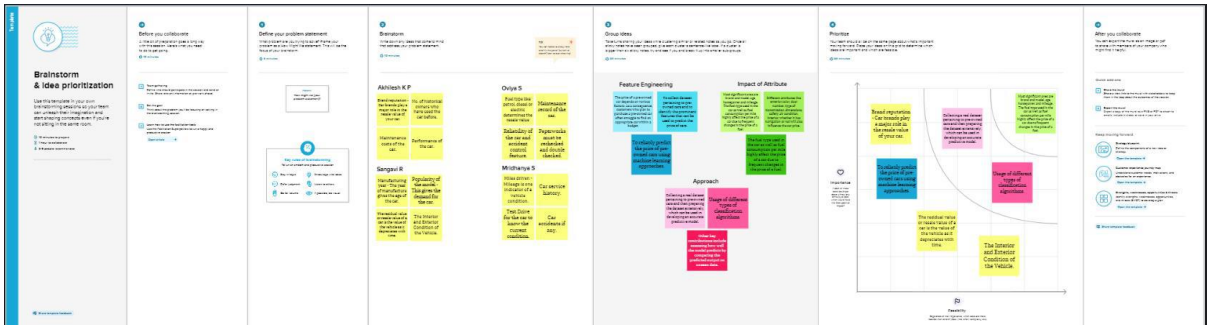
CAR RESALE VALUE PREDICTION

EMPATHY MAP



PAINS	GAINS
Buying a used car means that it can fall apart at any time.	The main attraction of buying a used car is because you do not have to pay for its first depreciation.
Even if the used car is in good working order, you have to deal with repairs sooner than later.	Get a cheaper sales tax and more affordable insurance.
Buying a used car is that it has no warranty.	Can save about a thousand dollars by avoiding the new car registration fees.
You may not know that reason why it was traded in. You need to obtain the vehicle history report first.	You don't have to settle for expensive dealership add-ons when buying a used car.

3.2 IDEATION AND BRAINSTORMING



3.3 PROPOSED SOLUTION

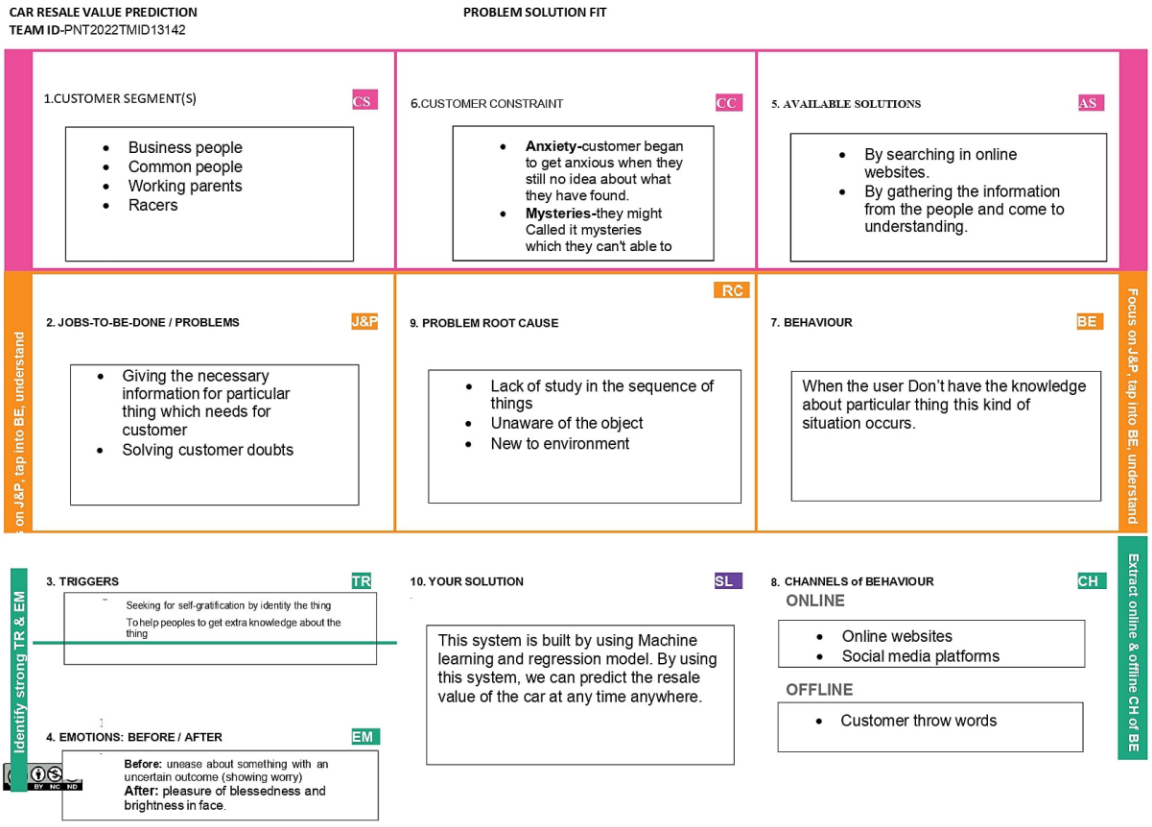
There are two primary phases in the system: 1. Training phase: The system is trained by using the data in the data set and fits a model (line/curve) based on the algorithm chosen accordingly. 2. Testing phase: the system is provided with the inputs and is tested for its working. The accuracy is checked. And therefore, the data that is used to train the model or test it, has to be appropriate. The system is designed to detect and predict price of used car and hence appropriate algorithms must be used to do the two different tasks. Before the algorithms are selected for further use, different algorithms were compared for its accuracy. The well-suited one for the task was chosen.
The main aim of this project is to predict the price of used cars using the various Machine Learning (ML) models. This can enable the customers to make decisions based on different inputs or factors namely Brand or Type of the car one prefers like Ford, Hyundai, Model of the car namely Ford Figo, Hyundai Creta, Year of manufacturing like 2020, 2021, Type of fuel namely Petrol, Diesel, Price range or Budget, Type of transmission which the customer prefers like Automatic or Manual, Mileage to name a few characteristic features required by the customer. This project Car Price Prediction deals with providing the solution to these problems. Different techniques like multiple linear regression analysis, k-nearest neighbours, naïve bayes and decision trees have been used to make the predictions. The predictions are then evaluated and compared in order to find those which provide the best performances. New cars of a particular make, model, and year all have the same retail price, excluding optional features. This price is set by the manufacturer. Used car, however, are subject to supply-and-demand pricing. Further, used cars have additional attributes that factor into the price. These include the condition, milage, and repair history, which sets cars that may have shared a retail price apart.

The purpose of this thesis is to evaluate several different machine learning models for used car price prediction and draw conclusions about how they behave. This will deepen the knowledge of machine learning applied to car valuations and other similar price prediction problems.

This work will focus on answering the research questions. They all entail a comparison of different ML algorithms for price prediction. This will be accomplished by sourcing and

preparing a dataset on which all the algorithms can be trained on and compared fairly. The algorithms selected must therefore be similar enough for the same dataset to be used for all of them. This also means that no large optimization efforts on the dataset will be made to boost the performance, if these changes do not benefit the other models. Maximizing price prediction performance of any one algorithm in ways that do not offer better comparisons is outside the scope of this work.
A revenue model is a blueprint that shows how a startup business will earn revenue or gross income from its standard business operations, and how it will pay for operating costs and expenses.
The optimal parameters were determined in the process of implementing the models, and thus each model was implemented with the parameters that yielded the best performance by trial and error. All of the models approximated geometric appreciation, meaning that a constant percentage of value is lost every year independent of the age of the vehicle. Random Forest Regression had a significantly higher assessed average depreciation at approximately 13.8%, compared to the others with 9.7%. This is closer to the range of 15%-31% assessed by Karl Storchmann in his analysis of international depreciation rates.

3.4 PROBLEM SOLUTION FIT



4 REQUIREMENT ANALYSIS

4.1 FUNCTIONAL REQUIREMENTS

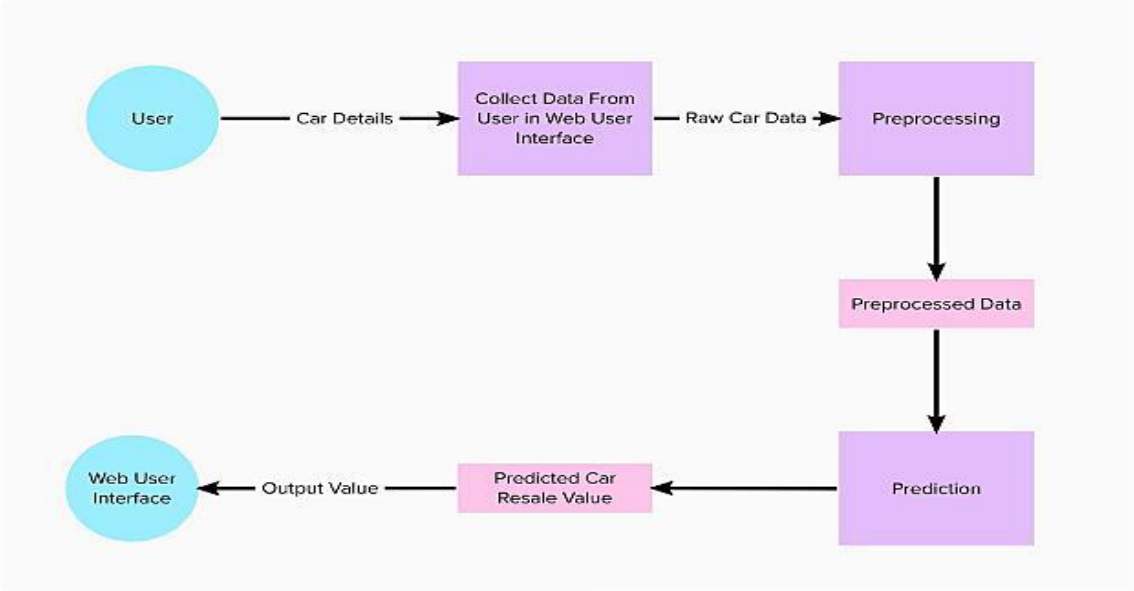
FR No.	Functional Requirement (Epic)	Sub Requirement (Story/ Sub-Task)
FR-1	User Registration	Registration throughWebsite
FR-2	User Confirmation	Confirmation via website
FR-3	Car Registration	Registering the car details
FR-4	Value Prediction	Predicting the car resale value

4.2 NON FUNCTIONAL REQUIREMENTS

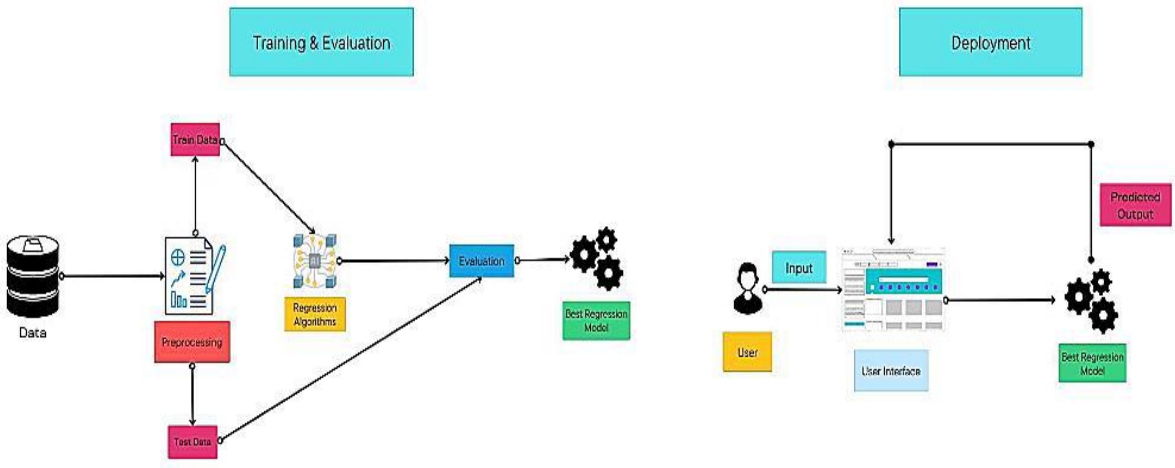
FR No.	Non-Functional Requirement	Description
NFR-1	**Usability**	Predicting the resale value
NFR-2	**Security**	Providing security to the website
NFR-3	**Reliability**	Providing high reliability by predicting values for different types of cars
NFR-4	**Performance**	Providing high performance by using somemachine learning techniques
NFR-5	**Availability**	It is used for all types of cars
NFR-6	**Scalability**	Predicting valuesfor different typesof cars

5 PROJECT DESIGN

5.1 DATA FLOW DIAGRAMS



5.2 SOLUTION AND TECHNICAL ARCHITECTURE



5.3 USER STORIES

User Type	Functional Requirement (Epic)	User StoryNumber	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Data Entry	USN-1	As a user, I can enter the car details in the application.	I can enterthe car details	Medium	Sprint-1
Customer (Mobile user)	Obtain output	USN-2	As a user, I will receive car resale value in the application.	I can receive my car resale value	High	Sprint-1
Customer (Mobile user)	Data Entry	USN-1	As a user, I can enter the car details in the application	I can enter the car details	Medium	Sprint-1
Customer (Mobile user)	Obtain output	USN-2	As a user, I will receive car resale value in the application.	I can recieve my car resale value.	High	Sprint-1

6. PROJECT PLANNING AND SCHEDULING

6.1 SPRINT PLANNING AND ESTIMATION

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Dataset reading and Preprocessing	USN-1	Cleaning the dataset and splitting to dependent and independent variables	2	High	Vignesh
Sprint-2	Building the model	USN-2	Choosingthe appropriate modelfor building and saving the model as pickle file	2	High	Maghibal
Sprint-3	Application building	USN-3	Using flask deploying the ML model	2	High	Sadhan

Sprint -4	Train the modelin IBM	USN-4	Finally train themodel on IBMcloud and deploy the application	2	High	Praveen

6.2 SPRINT DELIVERY SCHEDULE

Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	15	5 Days	24 Oct 2022	29 Oct 2022	15	29 Oct 2022
Sprint-2	15	5 Days	31 Oct 2022	05 Nov 2022	15	05 Nov 2022
Sprint-3	15	5 Days	07 Nov 2022	12 Nov 2022	15	12 Nov 2022
Sprint-4	15	5 Days	14 Nov 2022	19 Nov 2022	15	19 Nov 2022

7. CODING AND SOLUTIONING

7.1 FEATURE 1

index.html

```
<!DOCTYPE html>
<http-methods>
<html lang="en">
<head xmlns="http://www.w3.org/1999/xhtml">
   <meta charset="UTF-8">
   <title>Car Price Predictor</title>
   <link rel="stylesheet" href="static/css/style.css">
   <link rel="stylesheet" type="text/css"
      href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/5.11.2/css/all.css">
   <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.4.1/jquery.min.js"></script>
   <script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.0/dist/umd/popper.min.js"
      integrity="sha384-
Q6E9RHvbIyZFJoft+2mJbHaEWldlvI9IOYy5n3zV9zzTtmI3UksdQRVvoxMfooAo"
      crossorigin="anonymous"></script>

   <!-- Bootstrap CSS -->
   <link rel="stylesheet"
href="https://stackpath.bootstrapcdn.com/bootstrap/4.5.0/css/bootstrap.min.css"
```

```
        integrity="sha384-
9aIt2nRpC12Uk9gS9baDl411NQApFmC26EwAOH8WgZl5MYYxFfc+NcPb1dKGj7Sk"
crossorigin="anonymous">
    <script src="https://cdn.jsdelivr.net/npm/@tensorflow/tfjs@2.0.0/dist/tf.min.js"></script>

</head>
<body class="bg-dark">

<div class="container">
    <div class="row">
        <div class="card mt-50" style="width: 100%; height: 100%">
            <div class="card-header" style="text-align: center">
                <h1>Welcome to Car Resale Value Prediction</h1>
            </div>
            <div class="card-body">
                <div class="col-12" style="text-align: center">
                    <h5>This Web Page predicts the price of the car you want to sell. Try filling the
details below: </h5>
                </div>
                <br>
                <form method="post" accept-charset="utf-8" name="Modelform">
                    <div class="col-md-10 form-group" style="text-align: center">
                        <label><b>Select the company:</b> </label><br>
                        <select class="selectpicker form-control" id="company" name="company"
required="1"
                                onchange="load_car_models(this.id,'car_models')">
                            {% for company in companies %}
                            <option value="{{ company }}">{{ company }}</option>
                            {% endfor %}
                        </select>
                    </div>
                    <div class="col-md-10 form-group" style="text-align: center">
                        <label><b>Select the model:</b> </label><br>
                        <select class="selectpicker form-control" id="car_models" name="car_models"
required="1">
                        </select>
                    </div>
                    <div class="col-md-10 form-group" style="text-align: center">
                        <label><b>Select Year of Purchase:</b> </label><br>
                        <select class="selectpicker form-control" id="year" name="year" required="1">
                            {% for year in years %}
                            <option value="{{ year }}">{{ year }}</option>
                            {% endfor %}
                        </select>
                    </div>
                    <div class="col-md-10 form-group" style="text-align: center">
                        <label><b>Select the Fuel Type:</b> </label><br>
                        <select class="selectpicker form-control" id="fuel_type" name="fuel_type"
required="1">
                            {% for fuel in fuel_types %}
                            <option value="{{ fuel }}">{{ fuel }}</option>
                            {% endfor %}
```

```
                </select>
              </div>
              <div class="col-md-10 form-group" style="text-align: center">
                <label><b>Enter the Number of Kilometres that the car has travelled:</b>
</label><br>
                <input type="text" class="form-control" id="kilo_driven" name="kilo_driven"
                    placeholder="Enter the kilometres driven ">
              </div>
              <div class="col-md-10 form-group" style="text-align: center">
                <button  class="btn btn-primary form-control" onclick="send_data()">Predict
Price</button>
              </div>
            </form>
            <br>
            <div class="row">
              <div class="col-12" style="text-align: center">
                <h4><span id="prediction"></span></h4>
              </div>
            </div>
          </div>
        </div>
      </div>
    </div>

<script>

    function load_car_models(company_id,car_model_id)
    {
        var company=document.getElementById(company_id);
        var car_model= document.getElementById(car_model_id);
        console.log(company.value);
        car_model.value="";
        car_model.innerHTML="";
        {% for company in companies %}
            if( company.value == "{{ company }}")
            {
                {% for model in car_models %}
                    {% if company in model %}

                        var newOption= document.createElement("option");
                        newOption.value="{{ model }}";
                        newOption.innerHTML="{{ model }}";
                        car_model.options.add(newOption);
                    {% endif %}
                {% endfor %}
            }
        {% endfor %}
    }

    function form_handler(event) {
        event.preventDefault(); // Don't submit the form normally
    }
```

```
    function send_data()
    {
        document.querySelector('form').addEventListener("submit",form_handler);

        var fd=new FormData(document.querySelector('form'));

        var xhr= new XMLHttpRequest({mozSystem: true});

        xhr.open('POST','/predict',true);
        document.getElementById('prediction').innerHTML="Wait! Predicting Price.....";
        xhr.onreadystatechange = function(){
            if(xhr.readyState == XMLHttpRequest.DONE){
                document.getElementById('prediction').innerHTML="Prediction:
₹"+xhr.responseText;

            }
        };

        xhr.onload= function(){};

        xhr.send(fd);
    }
</script>


<!-- jQuery first, then Popper.js, then Bootstrap JS -->
<script src="https://code.jquery.com/jquery-3.5.1.slim.min.js"
        integrity="sha384-
DfXdz2htPH0lsSSs5nCTpuj/zy4C+OGpamoFVy38MVBnE+IbbVYUew+OrCXaRkfj"
        crossorigin="anonymous"></script>
<script src="https://cdn.jsdelivr.net/npm/popper.js@1.16.0/dist/umd/popper.min.js"
        integrity="sha384-
Q6E9RHvbIyZFJoft+2mJbHaEWldlvI9IOYy5n3zV9zzTtmI3UksdQRVvoxMfooAo"
        crossorigin="anonymous"></script>
<script src="https://stackpath.bootstrapcdn.com/bootstrap/4.5.0/js/bootstrap.min.js"
        integrity="sha384-
OgVRvuATP1z7JjHLkuOU7Xw704+h835Lr+6QL9UvYjZE3Ipu6Tp75j7Bh/kR0JKI"
        crossorigin="anonymous"></script>
</body>
</html>
```

style.css

```
.{
    margin: 0;
    padding: 0;
    box-sizing: border-box;
}
.bg-dark{

}
```

```
.mt-50{
        margin-top: 50px;
}
#canvas{
   border: 2px solid black;
}
```

7.2 FEATURE 2

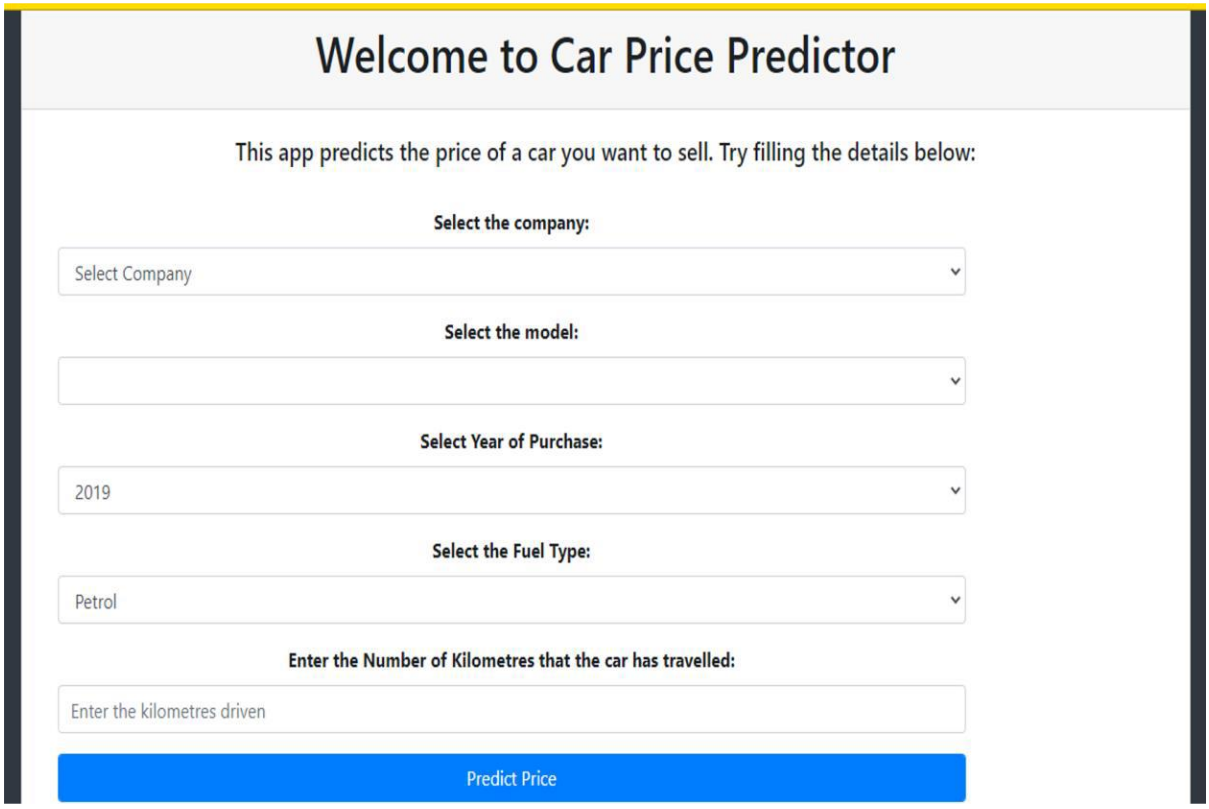
Resale_flask.py

```
from flask import Flask,render_template,request,redirect
from flask_cors import CORS,cross_origin
import pickle
import pandas as pd
import numpy as np

app=Flask(__name__)
cors=CORS(app)
model=pickle.load(open('RandomForest.pkl','rb'))
car=pd.read_csv('Cleaned.csv')

@app.route('/test',methods=['GET','POST'])
def index():
   companies=sorted(car['company'].unique())
   car_models=sorted(car['name'].unique())
   year=sorted(car['year'].unique(),reverse=True)
   fuel_type=car['fuel_type'].unique()

   companies.insert(0,'Select Company')
   return render_template('index.html',companies=companies, car_models=car_models,
years=year,fuel_types=fuel_type)


@app.route('/predict',methods=['POST'])
@cross_origin()
def predict():

   company=request.form.get('company')

   car_model=request.form.get('car_models')
   year=request.form.get('year')
   fuel_type=request.form.get('fuel_type')
   driven=request.form.get('kilo_driven')

   prediction=model.predict(pd.DataFrame(columns=['name', 'company', 'year', 'kms_driven',
'fuel_type'],
                         data=np.array([car_model,company,year,driven,fuel_type]).reshape(1, 5)))
   print(prediction)

   return str(np.round(prediction[0],2))
```

```
if __name__=='__main__':
    app.run()
```

8. RESULTS

8.1 PERFORMANCE METRICS



9. ADVANTAGES AND DISADVANTAGES

9.1 ADVANTAGES

i) Cost is less
Used cars are exponentially more affordable than buying a new vehicle.

ii)Lower Insurance Rates
Another lesser known fact is that car insurance costs less per month for a used car, than a new car. So the savings of buying a used car is evident on a monthly basis in the form of the monthly premium in addition to the sticker price.

iii)Many Used Cars Include a Warranty
One of the most common questions about buying a used car is about warranty. Many used cars are still under factory warranty. Most people who lease a car, turn it back in after three years, and still have 2 years left on its 5 year warranty. Certified Pre-Owned Vehicles and most used cars have gone through a rigorous inspection and will come with an extended warranty by the time you see them on the lot.

iv)High Maintenance
New cars come with basic maintenance packages that allow you to get your oil changed for free at certain checkpoints, such as a 25,000 mileage check-up. Used cars will typically not come with this feature. With a used vehicle, it's your responsibility to get check-ups and oil changes whenever you need to, which may be immediately after you purchase the vehicle.

10. CONCLUSION

The increased prices of new cars and the financial incapability of the customers to buy them, Used Car sales are on a global increase. Therefore, there is an urgent need for a Used Car Price Prediction system which effectively determines the worthiness of the car using a variety of features. The proposed system will help to determine the accurate price of used car price prediction

11. FUTURE WORKS

In future this machine learning model may bind with various website which can 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 learning rates and train on clusters of data rather than the whole dataset.

12. APPENDIX

https://github.com/IBM-EPBL/IBM-Project-714-1658316714