Project Report – Car Resale Value Prediction

1. INTRODUCTION

1.1 Project Overview

With difficult economic conditions, it is likely that sales of second-hand imported (reconditioned) cars and used cars will increase. In many developed countries, it is common to lease a car rather than buying it outright. After the lease period is over, the buyer has the possibility to buy the car at its residual value, i.e. its expected resale value. Thus, it is of commercial interest to sellers/financers to be able to predict the salvage value (residual value) of cars with accuracy.

1.2 Purpose

In order to predict the resale value of the car, we proposed an intelligent, flexible, and effective system that is based on using regression algorithms. Considering the main factors which would affect the resale value of a vehicle a regression model is to be built that would give the nearest resale value of the vehicle. We will be using various regression algorithms and algorithm with the best accuracy will be taken as a solution, then it will be integrated to the web-based application where the user is notified with the status of his product.

2. LITERATURE SURVEY

2.1 Existing problem

It is expected that sales of old cars and second-hand imported (reconditioned) autos will rise in tough economic times. Leasing a car rather than purchasing one entirely is typical in many affluent nations. After the lease term is up, the buyer will have the option of purchasing the vehicle for its residual value, or anticipated resale value. But It was difficult to accurately anticipate the salvage value (residual value) of cars for sellers and financiers from a business standpoint considering the various configurations of the car.

2.2 References

S.NO	PAPER	AUTHOR	YEAR	METHOD AND ALGORITHM
1.	Car resale price forecasting: The impact of regression method, private information, and heterogeneity on forecast accuracy	Stefan Lessmann, Stefan Vob	2017	Resale price forecasting is first done with Random Forest Regression. Then the same price forecastign is done with externally generated residual value estimates and finally the two results are compared to determine the best approach.
2.	Prediction of Resale Value of the Car Using Linear Regression Algorithm	Kiran S	2020	A correlation with each attribute to that of target attribute is found and linear regression curve with the target attribute is drawn. As a final step the total error and accuracy is measured.
3.	Car Price Prediction in the USA by using Liner Regression	Huseyn Mammadov	2021	They proposed a model using linear regression since the dependent variable price is linearly related to many independent variables and they have eliminated the irrelevant features by using the recursive feature elimination to reduce the dimensionality. Then R-square and root mean squared error is used to reduce the errors produced.

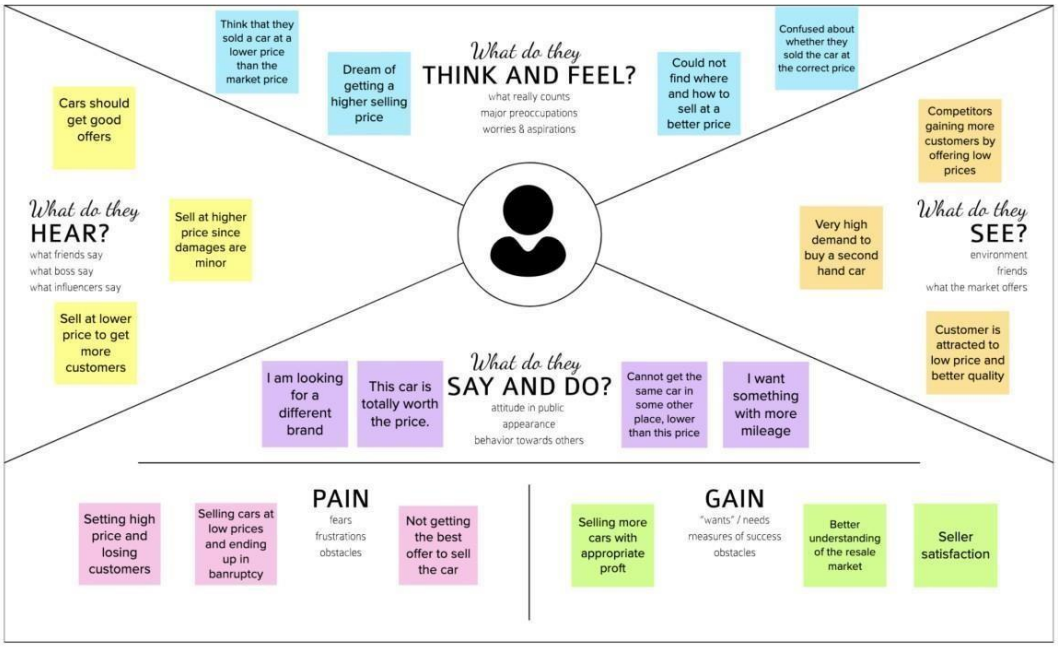
S.NO	PAPER	**AUTHOR**	YEAR	**METHOD AND ALGORITHM**
4.	Predicting the Price of Used Cars using Machine Learning Techniques	Sameerchand Pudanith	2013	Different techniques like multiple linear regression analysis, k-nearest neighbor, naive baye5 and decision trees have been used to make the predictions. The predictions are then evaluated and compared in order to find those which provide the best performances.
5.	Used Cars Price Prediction using Supervised Learning Techniques	Pattabiraman Venkatasubbu, Mukke5h Ganesh	2019	They proposed a model using mulhple and lasso regression. Using Lasso regression on the training data set, we first select the subset of attributes that lead to less error while predicting the price. It makes use of 10-fold cross-validation and Ll regularization. A general linear model, which models price lothe Set of selected attributes from lasso regression is used for multiple regression training.

2.3 Problem Statement Definition

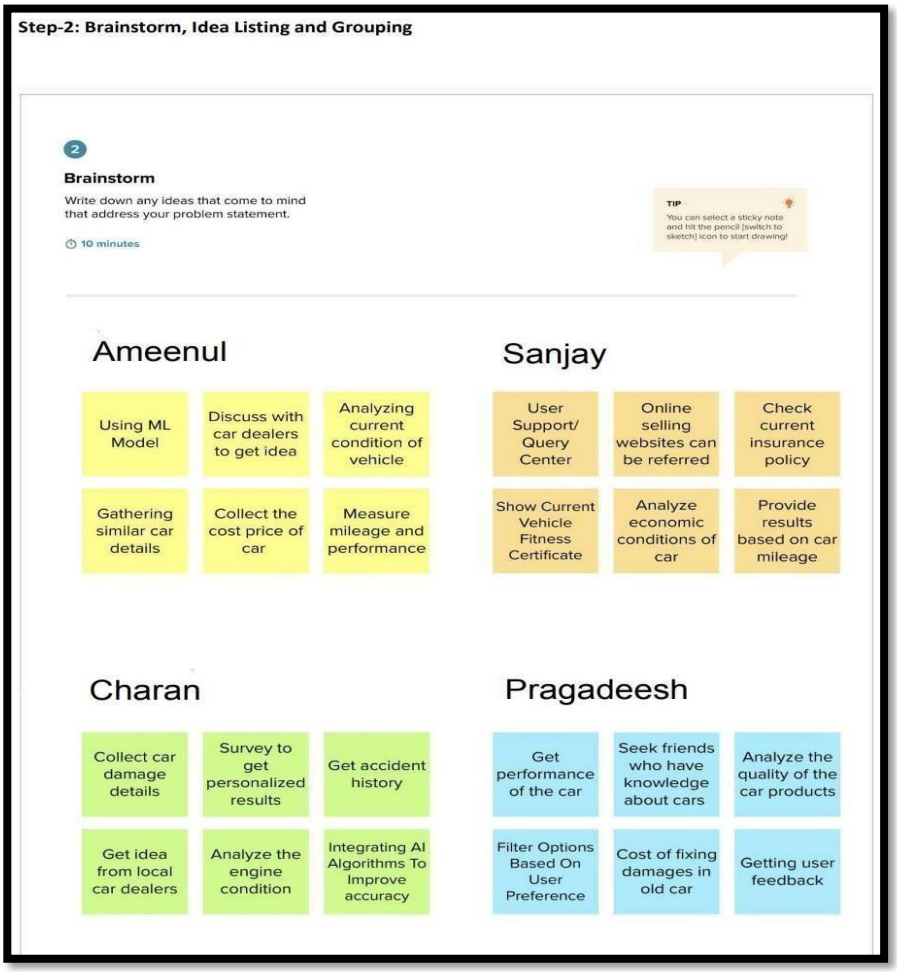
Problem Statement (PS)	I am (Customer)	I'm trying to	But	Because	Which makes me feel
PS-1	Car dealer	sell used cars	Increasing price	rising interest rates, tariffs, and energy concerns, car dealerships are expected to have fewer sales, especially with newer vehicles	Sad and Worried
PS-2	Common People	Buy 2nd hand cars	Can't decide on cars	Too many car models & prices. Common man can't decide correct car.	Fear of making wrong choice
PS-3	Budget Oriented People	Buy cheap cars	Price is not justified	Price is increased and cannot justify price	Paying more for cars
PS-4	Seller	To sell my car at reasonable price	Deciding on the price is hard	Too many complications in calculating the correct price for selling cars	Unhappy for not selling car at correct price.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



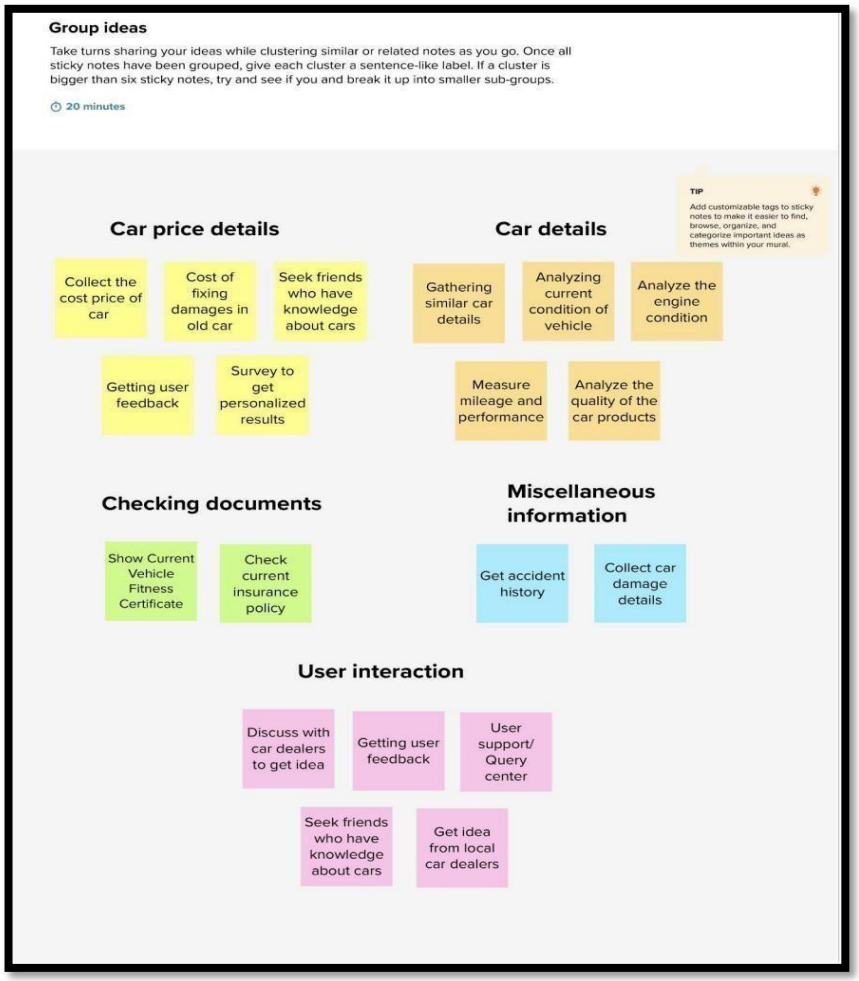
3.2 Ideation & Brainstorming

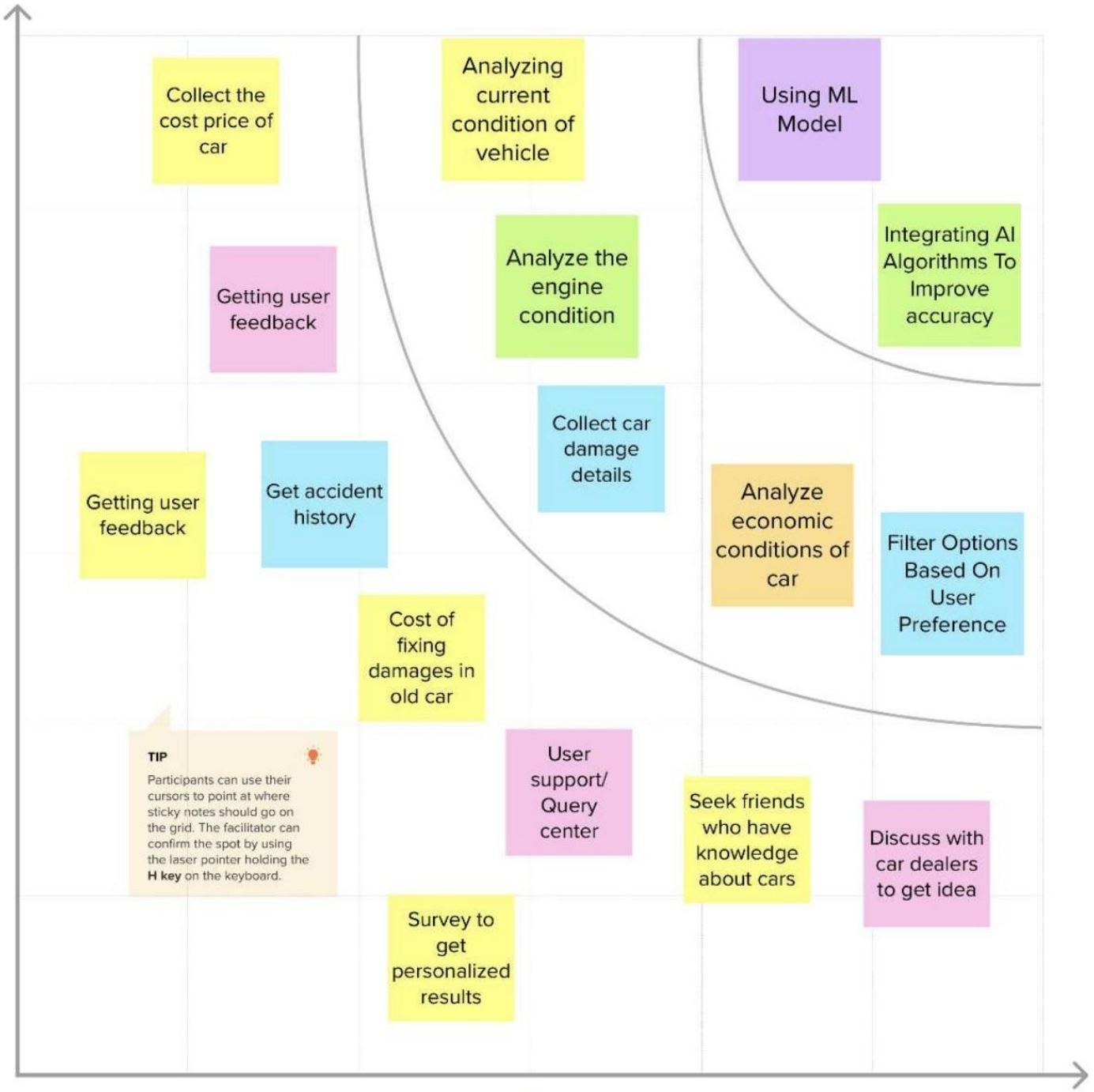


Group ideas

Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes





Importance

If each of these tasks could get done without any difficulty or cost, which would have the most positive impact?

Feasibility

Regardless of their importance, which tasks are more feasible than others? (Cost, time, effort, complexity, etc.)

3.3 Proposed Solution

S.No.	Parameter	Description
1.	Problem Statement (Problem to be solved)	• The main aim of this project is to predict the resale value of a used car using regression algorithms. • This could help the customers to find the best price of the used car that is going to be sold.
2.	Idea / Solution description	• The resale value of a car depends on factors such as price, vehicle type, gearbox, model, kilometres run, fuel type, etc. • The data is then pre-processed to handle missing values and outliers, to normalize the data and split it into dependent and independent variables. • After that the model is developed using regression algorithms to predict the resale price of the car.
3.	Novelty / Uniqueness	• This is a real-time problem which can benefit both customer and seller. • The novelty of this proposal is to predict the resale value as near as possible to the actual value.
4.	Social Impact / Customer Satisfaction	• Provided the current economic times, it is more likely that the usage of second-hand cars will increase. • This is a mutual commercial interest to both the customers and the sellers. • It predicts the resale values of the car based on all its features and prevents over-pricing or under-pricing. • This sets an understanding or trust between the seller and the customer.
5.	Business Model (Revenue Model)	• The proposed model could be sold to resellers so that they could use it to find the perfect price for bidding. • It could be developed into an application and get revenue from it if more no of users started to using it to find the best value of a second-hand car.
6.	Scalability of the Solution	• The primary model is targeted only for a lower number of audiences. • However, as the customer base increases for the model it can be extended to the cloud for effective services.

3.4 Problem Solution fit

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

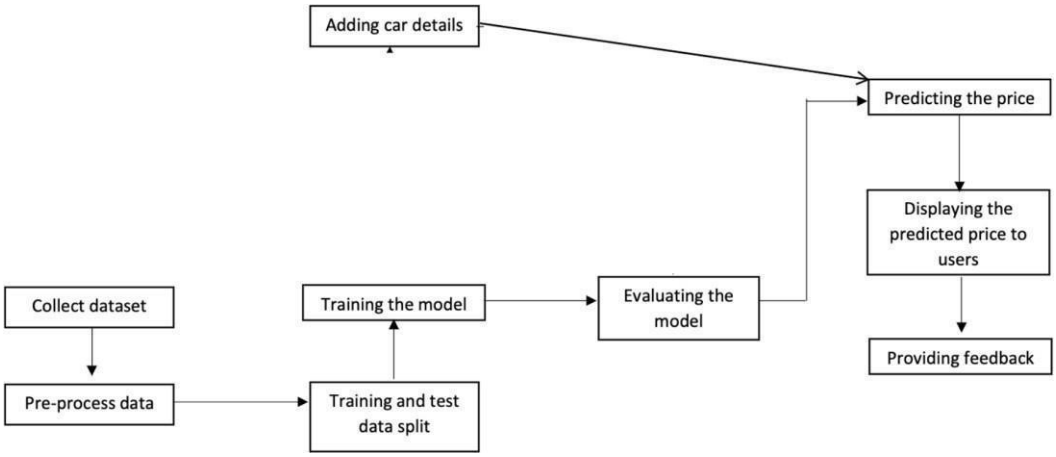
FR NO	Functional Requirement (Epic)	Sub Requirement (Story / Sub-Task)
FR-1	Car details	• Adding car details • Parsing the details using label encoder • Posting the data to IBM cloud model
FR-2	Value Prediction	• Predicting the resale value using the trained model • Returning the result
FR-3	Result	• Parsing the result from returned JSON • Displaying it on the website

4.2 Non-Functional requirements

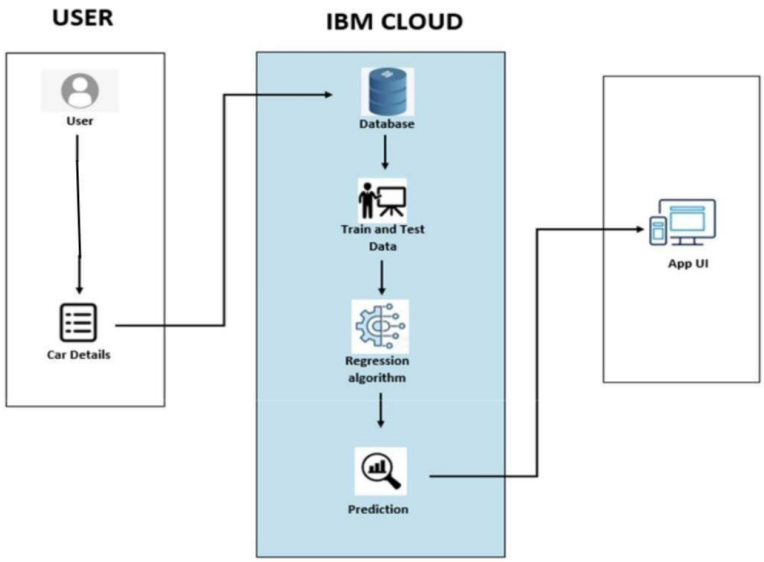
NFR NO	Non-Functional Requirement	Description
NFR-1	Usability	• User friendly UI • Clear Instructions • Easy process flow
NFR-2	Performance	• Quick prediction result • ML algorithm with better accuracy and less time complexity • Fast website loading
NFR-3	Availability	• Application can be accessed from both mobile and desktop • Uninterrupted user service

5. PROJECT DESIGN

5.1 Data Flow Diagrams



5.2 Solution & Technical Architecture



5.3 User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Admin	Dataset	USN-1	Collect the required data for the Car resale prediction.	Enough data collected for training model	High	Sprint-1
	Data preprocessing	USN-2	Perform data cleaning to optimize the dataset	Clean Dataset enough to make correct predictions	High	Sprint-1
	Training & Building Model	USN-3	Build the model using regression algorithms to classify the data	Model should be predicting prices with acceptable accuracy	High	Sprint-2
	Deploy the model	USN-4	Deployment of ML model using IBM Cloud	Model should be working fine from the cloud	High	Sprint-2
	Integrate the web app with the IBM model	USN-5	Use flask for the integration purpose.	Model should be easy to use & working fine from the web app.	High	Sprint-3
Customer	Homepage	USN-6	Details about the application and the car resale process	I can get an idea about the app	Medium	Sprint-3
	Car Details	USN-7	As a user, I should give the car details like car model, engine and fuel type, etc...	Car details should be accepted & taken for further processing	High	Sprint-4
	Car Price	USN-8	As a user, I can view the current rate of the used car price	Predicted price should be shown	High	Sprint-4

6. PROJECT PLANNING & SCHEDULING

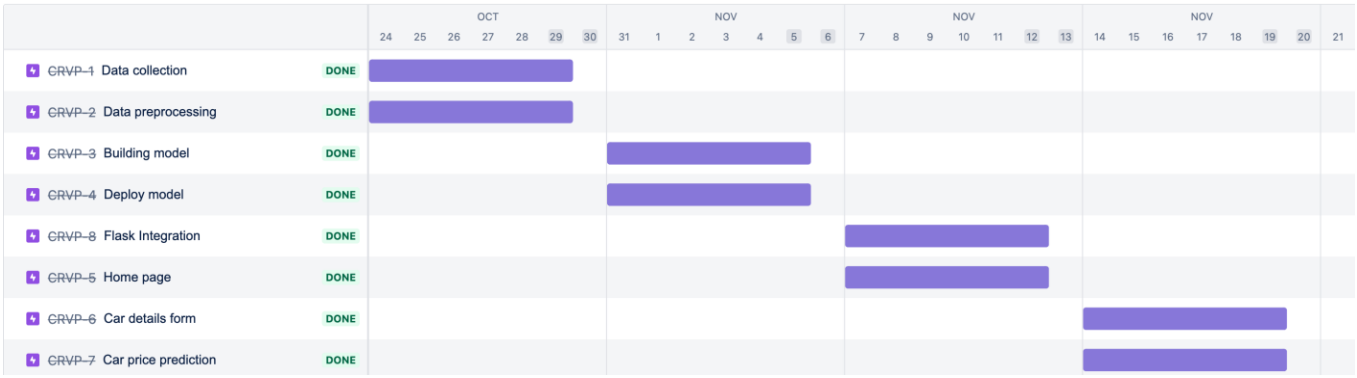
6.1 Sprint Planning & Estimation

User	Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Admin	Sprint 1	Dataset collection	USN-1	Collect the required data for the Car resale prediction	4	High	Ameenul, Sanjay, Charan, Pragadeesh
	Sprint 1	Data pre-processing	USN-2	Perform data cleaning to optimize the dataset	6	Medium	Sanjay, Ameenul
	Sprint 2	Training & Building Model	USN-3	Build the model using regression algorithms to classify the data	6	High	Ameenul, Sanjay, Charan, Pragadeesh
	Sprint 2	Deploy the model	USN-4	Deployment of ML model using IBM Cloud	4	High	Charan, Pragadeesh
	Sprint 3	Integration	USN-5	Integrate the web app developed using flask with IBM model	5	High	Sanjay, Ameenul
Customer	Sprint 3	Homepage	USN-6	Details about the application and the car resale process	5	Low	Charan, Pragadeesh
User	**Sprint**	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Story Points**	**Priority**	**Team Members**
	Sprint 4	Car Details	USN-7	As a user, I should give the car details like car model, engine and fuel type, etc…	5	Medium	Ameenul, Sanjay
	Sprint 4	Car Price	USN-8	As a user, I can view the current rate of the used car price	5	High	Ameenul, Sanjay

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	10	6 Days	24 Oct 2022	29 Oct 2022	10	29 Oct 2022
Sprint-2	10	6 Days	31 Oct 2022	05 Nov 2022	10	05 Nov 2022
Sprint-3	10	6 Days	07 Nov 2022	12 Nov 2022	10	12 Nov 2022
Sprint-4	10	6 Days	14 Nov 2022	19 Nov 2022	10	19 Nov 2022

6.3 Reports from JIRA



7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Random Forest Regressor with optimal depth

```
#Model building and Fitting
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import r2_score
regressor = RandomForestRegressor (n_estimators=1000, max_depth=15, random_state=34)
```

The maximum depth is 32, If it is higher the model takes too long to train and run and sometimes the memory gets filled up and the system crashes. If it is too low then the accuracy is very poor, as a result we took an optimal depth of 15.

7.2 Deployed the model in IBM cloud

```
model_details = client.repository.store_model(model=regressor,meta_props={
client.repository.ModelMetaNames.NAME:"CRVP-deployment",
client.repository.ModelMetaNames.TYPE:"scikit-learn_1.0",
client.repository.ModelMetaNames.SOFTWARE_SPEC_UID:software_spec_uid}
)
model_id = client.repository.get_model_id(model_details)
```

Stored the model in the IBM cloud using the above code, after that It is deployed using IBM watson through GUI.

7.3 Using Deployed model to predict the result.

```
#method to predict from deployed model
def predictFromDeploymentModel(userInput):
    API_KEY = "YOUR-APP-KEY"
    token_response = requests.post('https://iam.cloud.ibm.com/identity/token', data={"apikey":
    API_KEY, "grant_type": 'urn:ibm:params:oauth:grant-type:apikey'})
    mltoken = token_response.json()["access_token"]
    header = {'Content-Type': 'application/json', 'Authorization': 'Bearer ' + mltoken}
    payload_scoring = {"input_data": [
        {"fields": [
            'yearOfRegistration'
            ,'powerPS'
            ,'kilometer'
            ,'monthOfRegistration'
            ,'gearbox_labels'
            ,'notRepairedDamage_labels'
            ,'model_labels'
            ,'brand_labels'
            ,'fuelType_labels'
            ,'vehicleType_labels'], "values": [userInput]}]}
    response_scoring = requests.post(
        'https://us-south.ml.cloud.ibm.com/ml/v4/deployments/9528ebfb-f57a-4b7b-9684-d745eea3da24/predictions?version=2022-11-08', json=payload_scoring,
    headers={'Authorization': 'Bearer ' + mltoken})
    predictions = response_scoring.json()
    return (predictions['predictions'][0]['values'][0][0])
```

Used IBM watson to deploy the model, instead of storing the large model.sav file in the local which is not feasible for practical use. This code sends the API request to the deployed model along with the data that the user had entered for which we want to predict the result. After successful prediction the result comes in the form of json which is later parsed and the resale value is obtained.

7.4 Dynamic prediction page using flask

```
    <div class="result-div">
      <p>{{ data.result }}</p>
</div>
```

The Result element is shown only when the result is available, it is dynamically rendered using flask data, which can be passed while rendering the page.

```
result = predictFromDeploymentModel(list(X[0]))
data = {"result" : "Predicted price - $ "+str(round(result,2))}
return render_template('prediction.html' , data = data)
```

If the result is available the result text is passed

```
@app.route('/predict')
def predict():
    return render_template('prediction.html',data = {'result' : ""})
```

If there is no result, it will be empty in the screen

7.4 Beautiful UI using jQuery and bootstrap

```
<head>
    <meta charset="UTF-8">
    <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/css/bootstrap.min.css">
    <link rel="stylesheet" href="{{ url_for('static',filename='css/prediction.css') }}">
    <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.6.0/jquery.min.js"></script>
    <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.4.1/js/bootstrap.min.js"></script>
</head>
<body>
    <form class="form-horizontal" action="{{ url_for('y_predict') }}" method="post">
      <fieldset>
      <h2 class="text-center">Car Resale Value Prediction</h2>
      <br>
      <div class="form-group">
        <label class="col-md-4 control-label" for="regyear">Registration Year</label>
        <div class="col-md-4">
        <input name="regyear" type="text" class="form-control input-md" required>
        </div>
      </div>
      <div class="form-group">
        <label class="col-md-4 control-label" for="selectMonth">Registration Month</label>
        <div class="col-md-4">
          <select name="regmonth" class="form-control" id="selectMonth">
            <option value="january">january</option>
          </select>
        </div>
      </div>
      <div class="form-group">
        <label class="col-md-4 control-label" for="powerps">Power of car in PS</label>
        <div class="col-md-4">
        <input name="powerps" type="text" class="form-control input-md" required>
        </div>
      </div>
```

Used customized Form elements from jQuery and Bootstrap, though media query is not added, the web page will load correctly on the mobile without overlapping each other and the contents will not be misplaced.

8. TESTING

8.1 Test Cases

Test case ID	Feature Type	Component	Test Scenario	Steps To Execute	Test Data	Expected Result	Actual Result	Status	Executed By
HomePage_TC_OO 1	Functional	Home Page	Verify user is able to the home page by clicking on the link	1.Enter URL and click go 2.Verify home page displayed or not	http://ameenul. pythonanywhere.com/	Home page should be shown	Working as expected	Pass	Ameenul
HomePage_TC_OO 2	UI	Home Page	Verify the UI elements are responsive	1.Enter URL and click go 2.Repeat the step 1 in different devices 3.Verify all the Elements are visible and accesible	http://ameenul. pythonanywhere.com/	Application should show below UI elements: a.Description b.Proceed to predict screen button	Working as expected	Pass	Ameenul
HomePage_TC_OO 3	Functional	Home page	User is able to click on go to predict screen button	1.Enter URL and click go 2.Click on Go to predict Screen button	http://ameenul. pythonanywhere.com/	User should navigate to predict screen	Working as expected	Pass	Ameenul
PredictionPage_TC _OO1	UI	Prediction page	Verify the UI elements are responsive	1.Enter URL and click go 2.Click on Go to predict Screen button 3.Check all the UI elements are accessible and visible 4.Repeat the steps in different devices	http://ameenul. pythonanywhere.com/	Application should show below UI elements: a.form element with 10 fields b.A submit button	Working as expected	Pass	Ameenul
PredictionPage_TC _OO2	Functional	Prediction page	Verify user cannot enter null values	1.Enter URL and click go 2.Click on go to predict screen button 3.Submit the form without entering values	http://ameenul. pythonanywhere.com/	Application should show 'Please fill in this field ' validation message.	Working as expected	Pass	Ameenul
PredictionPage_TC _OO3	Functional	Predictoin page	Verify user cannot enter Invalid values (e.g a string for year field)	1.Enter URL and click go 2.Click on go to predict screen button 3.Enter 'Year' in year field, 'ps' in Power ps field and 'KM' in Kilometers field	Registration Year : "random string" Power of car in PS : "random string" Kilometres that car has driven : "random string"	Application should not enter those values into fields	Working as expected	Pass	Ameenul
PredictionPage_TC _OO4	Functional	Prediction page	Verify user is able to get the prediction result on screen	1.Enter URL and click go 2.Click on go to predict screen button 3.Enter the valid values in the form fields 4.Click on submit button	Registration Year : 2001 Power of car in PS : 150 Kilometres that car has driven : 150000	Application should show the prediction result at the bottom	Working as expected	Pass	Ameenul

8.2 User Acceptance Testing

Defect Analysis

This report shows the number of resolved or closed bugs at each severity level, and how they were resolved

Resolution	Severity 1	Severity 2	Severity 3	Severity 4	Subtotal
By Design	2	1	0	1	4
Duplicate	1	0	0	0	1
External	0	3	0	2	5
Fixed	5	1	6	8	20
Not Reproduced	0	1	1	0	2
Skipped	0	1	0	2	3
Won't Fix	0	7	0	2	9
Totals	8	14	7	15	44

Test Case Analysis

This report shows the number of test cases that have passed, failed, and untested

Section	Total Cases	Not Tested	Fail	Pass
Print Engine	3	0	0	3
Client Application	15	0	0	15
Outsource Shipping	3	0	0	3
Exception Reporting	2	0	0	2
Final Report Output	4	0	0	4
Version Control	2	0	0	2

9. RESULTS

9.1 Performance Metrics

```
#predicting the values fo test test
y_pred = regressor.predict(X_test)

mae = metrics.mean_absolute_error(Y_test,y_pred)
mse = metrics.mean_squared_error(Y_test,y_pred)
rmse = np.sqrt(mse)
r2 = metrics.r2_score(Y_test,y_pred)

print("Results of sklearn.metrics:")
print("MAE:",mae)
print("MSE:", mse)
print("RMSE:", rmse)
print("R-Squared:", r2)
```

```
Results of sklearn.metrics:
MAE: 1377.5136604841664
MSE: 9979418.71328512
RMSE: 3159.0217968993375
R-Squared: 0.860497492549767
```

The accuracy of the random forest regressor with max_depth=15 is 86%

Whereas,

- KNN-regressor – 72%
- Xgboost – 47%
- Support vector regressor – 70% (too much memory)

10. ADVANTAGES & DISADVANTAGES

Advantages:

- Very Fast prediction
- Simple UI, just click the link and predict

- Better accuracy without overfitting
- High availability

Disadvantages:

- If the actual value is higher then, the data points may have higher deviation in the predicted value.
- User entered data are not stored in database

11. CONCLUSION

Thus, by using Random Forest regressor, the model can predict the resale value of the car with maximum accuracy without overfitting. It also consumes low memory and faster than other regressors for this dataset. Also, deploying the model in IBM cloud allows us to use the model from the hosted website. In depth analysis and powerful computers can make this model more accurate.

12. FUTURE SCOPE

This project will be more useful in future, as renting and reselling of a car is becoming more common, there are some services such as cars24.com where we can buy and sell second hand cars, the customers of these services may need to estimate or predict the resale value of the cars based on its configuration.

13. APPENDIX

source code

demo

Live website

Github