iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: https://doi.org/10.22214/ijraset.2022.43870

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue VI June 2022- Available at www.ijraset.com

End to End Car Selling Portal By Loan Prediction Using Machine Learning

P Sathya Sai Ram[1], C Sravan Kumar[2], Mukund Pandey[3], D Rakesh[4], V Naveen[5], K Prem Kumar[6]
[1,2,3,4,5]*CSE,* [6] *Assistant Professor, Department Of Computer Science & Engineering, ACE Engineering College, Ghatkesar, Hyderabad, Telangana, India.*

Abstract: ***Cars have become an asset even though it's a liability for common use because of the comforts it provides, users wouldn't want to miss out on the different brands and luxuries it gives but jolting down the types , prices, and the finance part becomes hectic and a consumer usually buys it offline, but what if processes can be done faster digitally?. Buying online can significantly give many options for users/consumers .We are creating a solution by integrating a web application created using Express, React, Node.js, Google Firebase with Machine learning using the Random Forest Classifier after analysis of different models like KNN, SVC, Logistic Regression, Decision Tree Classifier, Extra Trees Classifier.***
Keywords: Web app, React, Node.js, Express, Firebase, Machine Learning, KNN, SVC, Logistic Regression, Decision Tree Classifier, Extra Trees Classifier, Random Forest Classifier

II. PROBLEM STATEMENT AND OBJECTIVE

For a car to reach the end user it is only being done offline i.e through authorized dealers . The user has to go through an old system of visiting the showroom and then has to purchase the car, this can be digitized so as to make the experience better for them. If a user opts for a loan he should submit the documents at the showroom and do all the tasks there which is tiresome.
The main objective of this project are:

- Create user interfaces for seller and customer
- Predict Loan Eligibility of the customer based on the input by the user
- To achieve good accuracy
- To handle payments , orders

III. INTRODUCTION

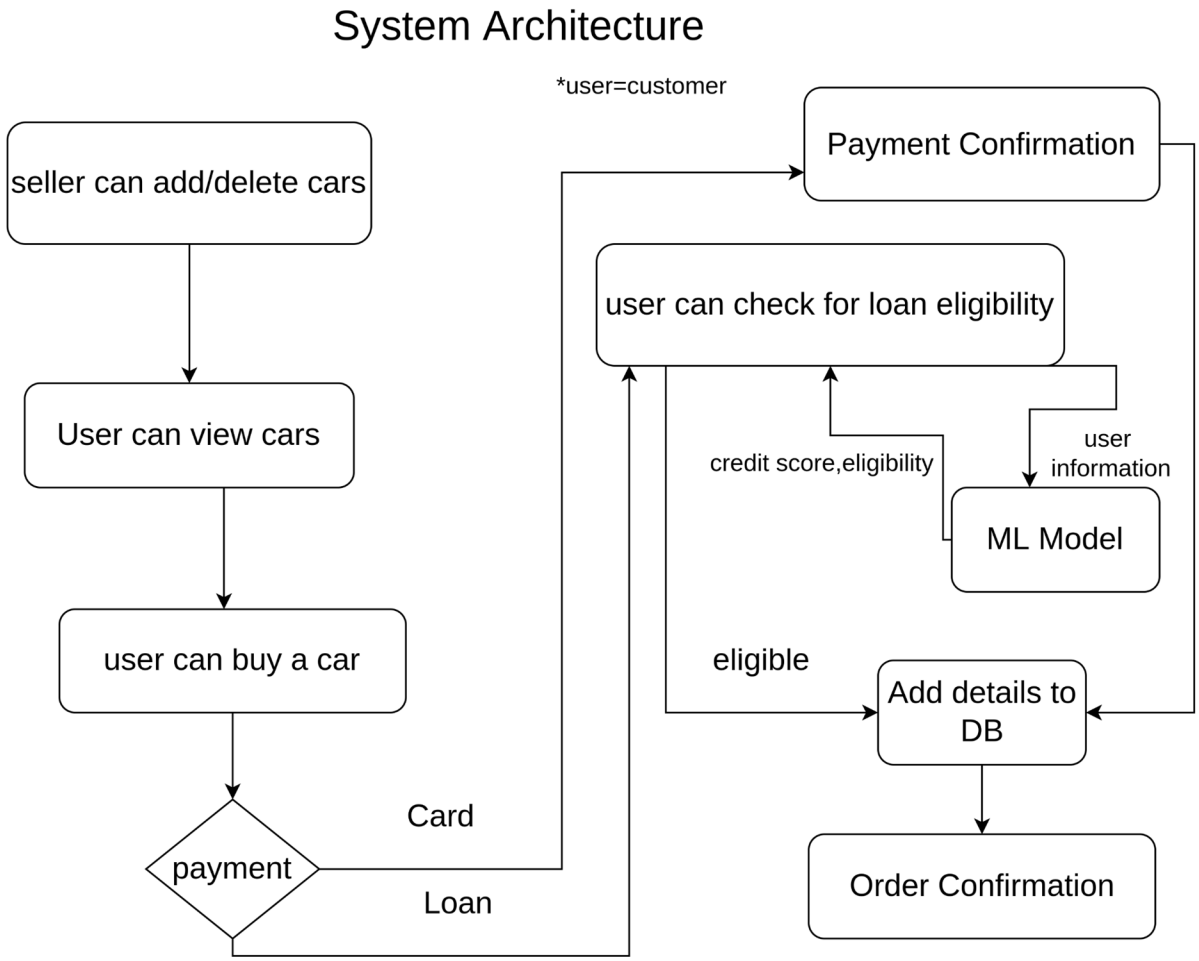
Basic human nature is to go for bargains be it online or offline, but in recent years the trend of offers and online shopping has taken its boom, from 1980 i.e inception of online shopping till
Present it has evolved so much that shopping is key factor in determining a personality, the modern ways of shopping online is simple and that is by the use of technology than telephones which was the main element, the world is moving fast and people are trying to catch up to it. By gadgets and what not , In this chaos consumers are only looking for product quality, delivery time, cost, time taken for the process. Offline stores may give 80 percent of the above mentioned qualities but the biggest disadvantage it still holds is 'taking time'. Consumers are happy if they are getting products delivered at home . The trend is showing that there will be a decrease in offline shopping in the upcoming years , the sectors which are dominantly offline and time consuming are car selling and automobiles in general . The US has seen good startups which are taking their automobile businessess online , but in India it's yet to take off . There is always the fear of supply chain management, shipping issues and warehousing but it's also more cost effective than giving it to dealers and decreasing their margins . Buying a car in India has become a hassle in the last few years with many factors to take into consideration, the research on cars takes lots of time and effort which is better than being blinded by advertisements. Selection of models, financing options , shortlisting alternatives can be tiresome and grueling ,
If the world is all about being progressive and moving forward, why take the traditional methods which are time consuming. Although we can't totally remove the dealer system, this is a step forward in cutting the middlemen and directly connecting to the customer.so web, mobile applications become essential which can help users to buy cars easily , give loan eligibility and loans .

IV. LITERATURE SURVEY

In [1] the author talks about creating a system for buying , selling , and renting a car in Bangladesh .They specify about the lack of apps and websites which are doing this . Their system focused on providing both trading and rental service which has given us some inputs on how to approach our system and possible implementational ideas in future if any

From the patent of [2] we understood the long process of creating a loan application and providing the user the satisfaction of completing the entire process at one place through many interfaces working in realtime and making things easy for users.We have taken notes on how to better design our system based on future possibility of adding banks as lenders .
Through [3] and [4] we observed how our project could be extended further by adding Machine learning , we understood the key variables which are mostly used and [4[gave us ideas of which type of algorithms to be used , they have used lasso regression , multiple regression and regression tree and their analysis was mentioned in the paper .

V. SYSTEM ARCHITECTURE



VI. ALGORITHMS USED

The algorithms we assessed for creating models were

- Random Forest Classifier
- Logistic Regression
- Kneighborsclassifier
- Decision Tree Classifier

A. Decision Tree Classifier

It's a supervised learning algorithm. Even though it's used for both classification and regression , it's widely used in classification problems. In short, A Tree is formed from the features given . It consists of Root Node, Internal Nodes, Leaf Nodes. Root Node is the starting point of the decision tree , it contains all the features of the dataset in consideration . Splitting is done on the features based on conditions to form internal nodes, the branches denote the condition on which it is split. Finally the Leaf nodes are created which act as the output nodes which give classification.

The initial root node features are selected based on attribute selection measure(ASM).This process goes on until there can be no attributes selected.

There are many ASM's but mostly we use gini index and information gain. The gini index is the probability of an instance being wrongly classified if a particular feature is selected, hence we take the feature which has the least gini index.

$$\text{Gini} = 1 - \sum_{i=1}^{n} (p_i)^2$$

Another ASM widely used is the information gain whose base is entropy , we select those features which try to reduce the entropy , thereby giving us some gain in information. So we select those features which give us most information gain

$$E(S) = \sum_{i=1}^{c} - p_i \log_2 p_i$$

B. K Neighbors Classifier

It's a supervised ML classification algorithm. It initially memorizes all the dataset and which features helped in resultant classification. It defines K most similar instances where K is user defined. For a dataset to be classified it checks the nearest instances which possess similar features and tries to classify the dataset based on max similarities as that particular class

C. Logistic Regression

Logistic regression is the appropriate regression analysis to conduct when the dependent variable is binary. Like all regression analyses, logistic regression is a predictive analysis. Logistic regression is used to describe data and to explain the relationship between one dependent binary variable and one or more nominal, ordinal, interval or ratio-level independent variables.

D. Random Forest Classifier

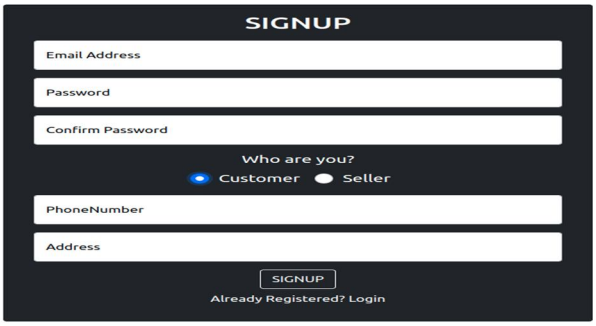
The fundamental idea behind a random forest is to combine many decision trees into a single model. Individually, predictions made by decision trees (or humans) may not be accurate, but combined together, the predictions will be closer to the mark on average.

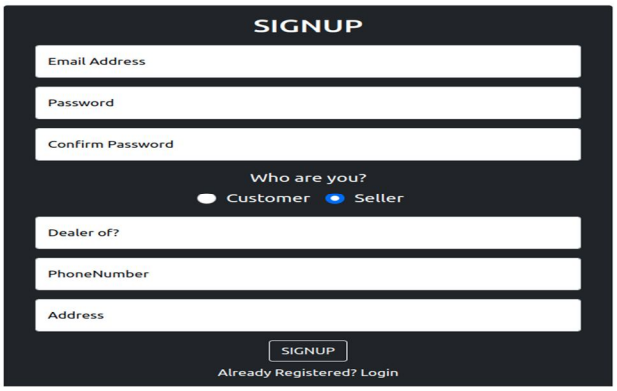
It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

VII. METHODOLOGY

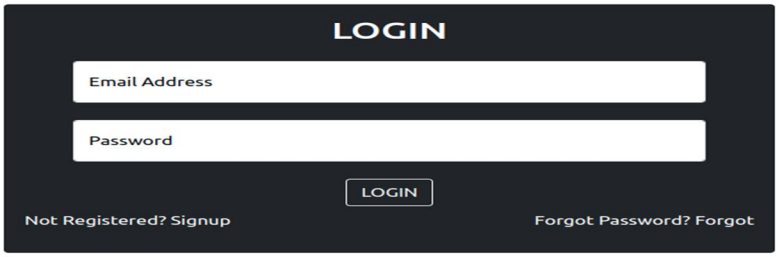
A. Web Application

1) SignUp: The user has to register to get through to the dashboard. The user could be a customer or seller, based on the selected option the fields change. If any field is left blank, email id already exists or the password and password confirmation doesn't match errors will be thrown.





2) *Login:* The user has to login with email and password. Based on their credentials that are stored in Firebase Firestore . They will be redirected to Customer Interface or Seller Interface. If the user has forgotten their password , there is a password reset component which sends mails to users after they enter their mail ids, if any error occurs it will be prompted , else the users will receive success message along with mails

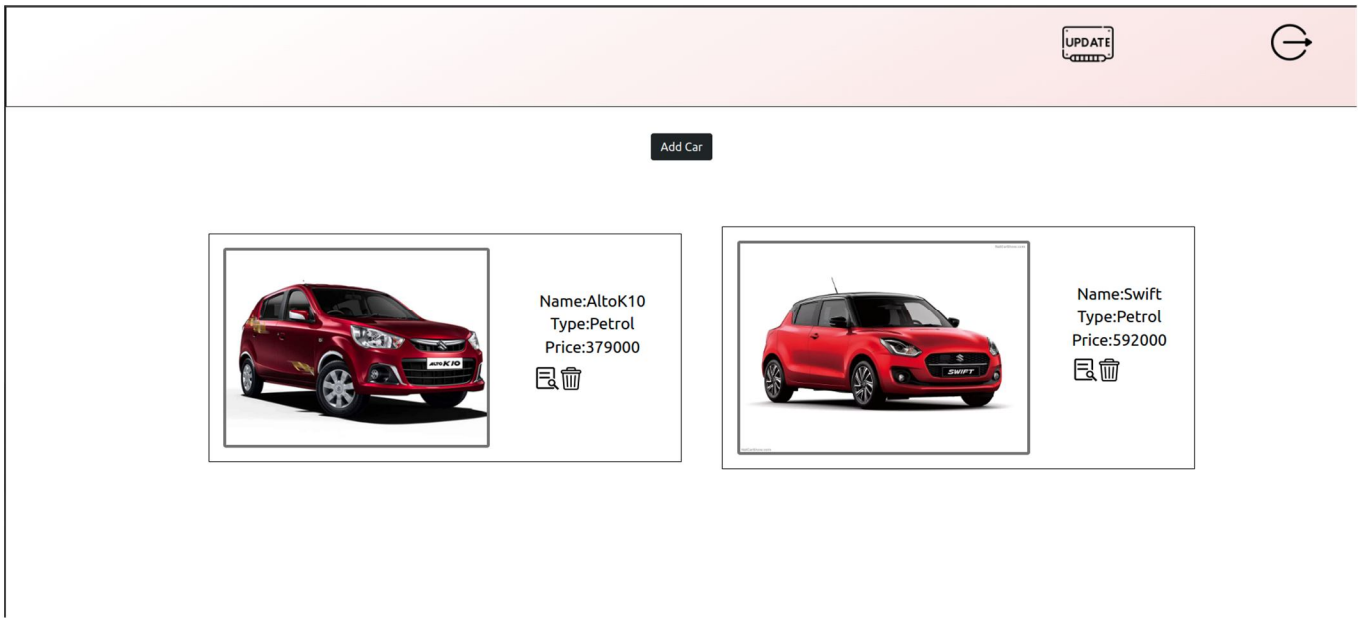


3) The Dashboards can be Customer dashboard or Seller dashboard.

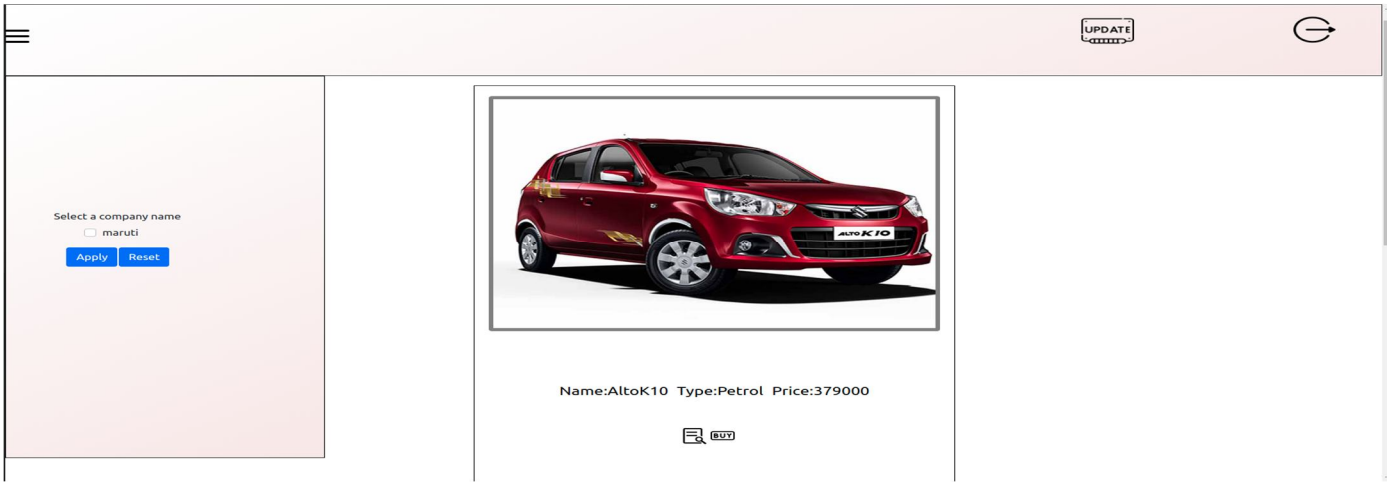
- Customer Dashboard has functionalities of Car selection Filters , Viewing Features of Car, Buying the car
- Seller Dashboard has functionalities of Adding the car , deleting the car

Both the dashboards have common features of Updating information and Logout

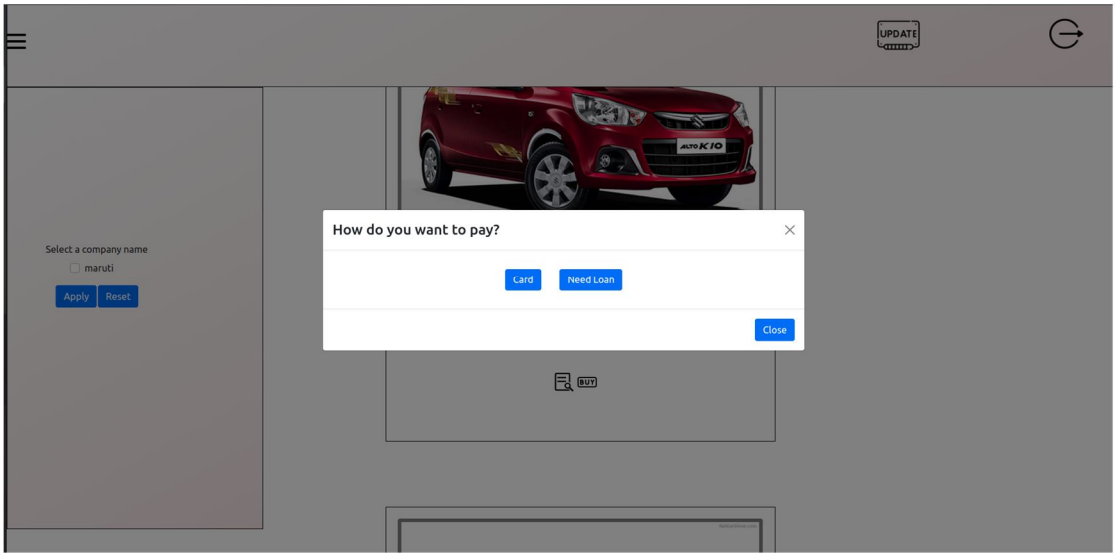
Seller dashboard

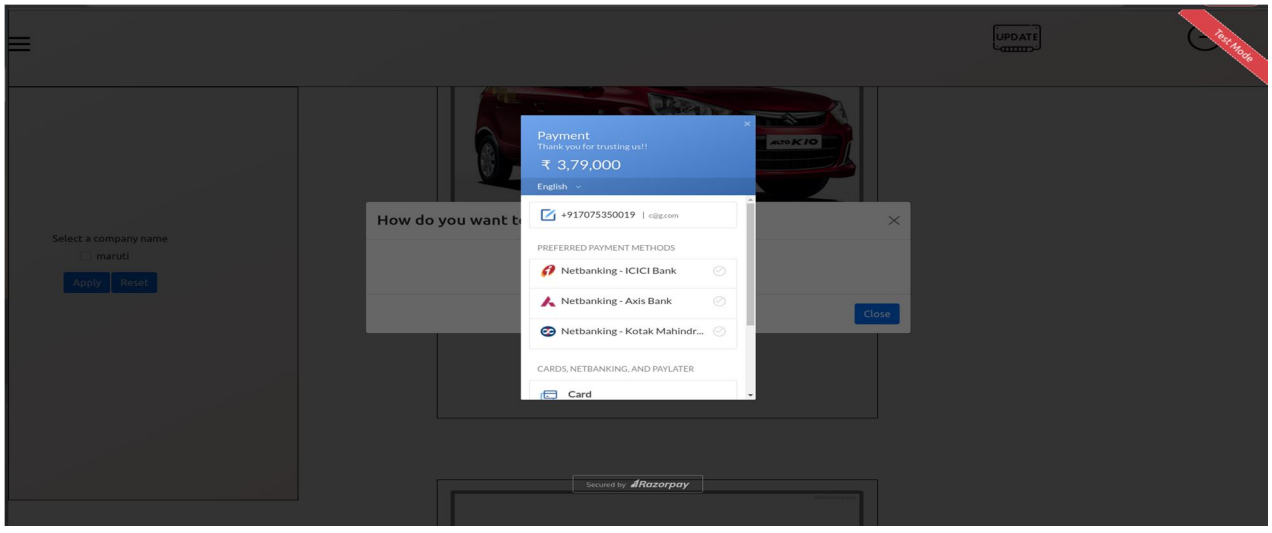


Customer Dashboard

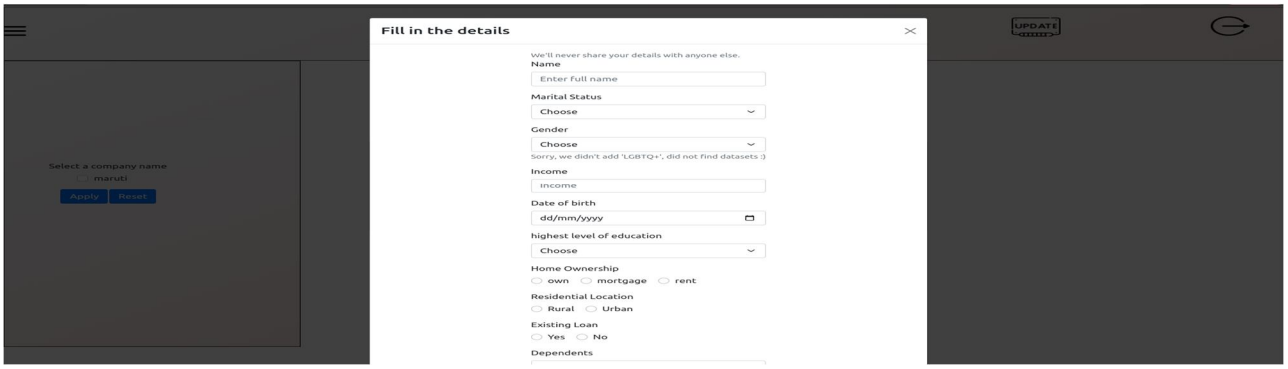


4) If a customer wants to buy a car, he can make payment through card or can check for loan eligibility

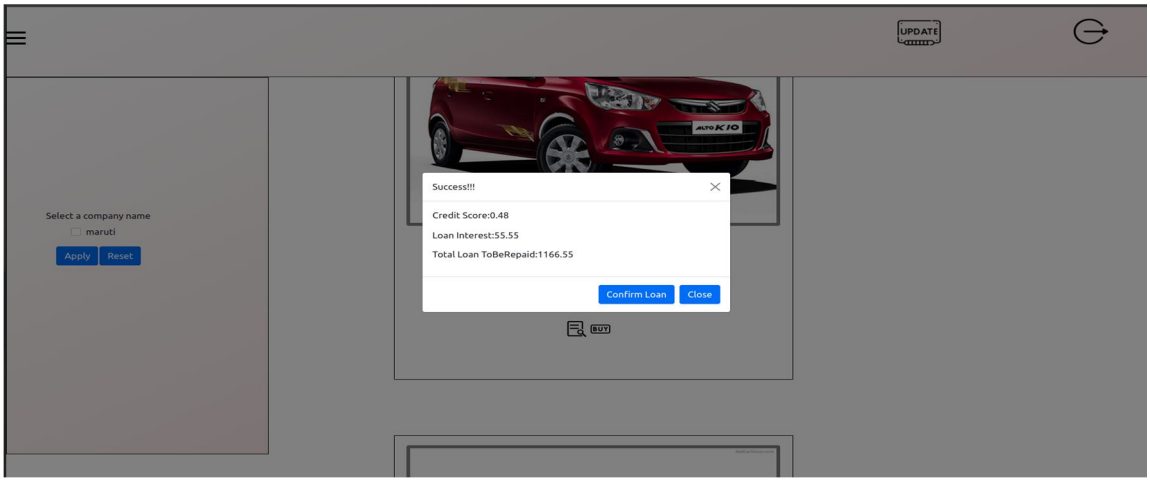


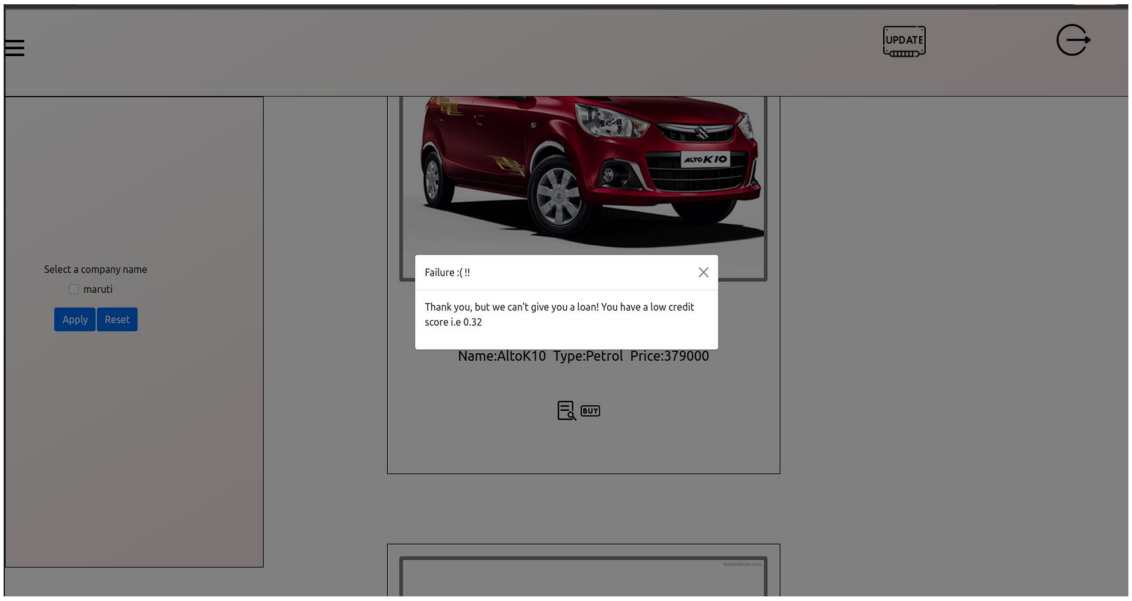


5) If the customer opts for loan eligibility then the necessary information required should be filled , the information is then sent to the ML model which is created using random forest classifier. This then based on the information returns the eligibility, credit score , amount to be repaid by the customer. If the customer is less than 18 , that case is handled too.



6) The Success or failure of the Transaction is determined by payment or opting for loan . The final details gets added to the database.





VIII. ML MODEL

A. Import libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

B. Load data

	0	1	2	3	4
Loan_ID	LP001002	LP001003	LP001005	LP001006	LP001008
Gender	Male	Male	Male	Male	Male
Married	No	Yes	Yes	Yes	No
Dependents	0	1	0	0	0
Education	Graduate	Graduate	Graduate	Not Graduate	Graduate
Self_Employed	No	No	Yes	No	No
ApplicantIncome	5849	4583	3000	2583	6000
CoapplicantIncome	0	1508	0	2358	0
LoanAmount	NaN	128	66	120	141
Loan_Amount_Term	360	360	360	360	360
Credit_History	1	1	1	1	1
Property_Area	Urban	Rural	Urban	Urban	Urban
Loan_Status	Y	N	Y	Y	Y

C. Summarize data

Fill missing values if any, for both categorical and numerical

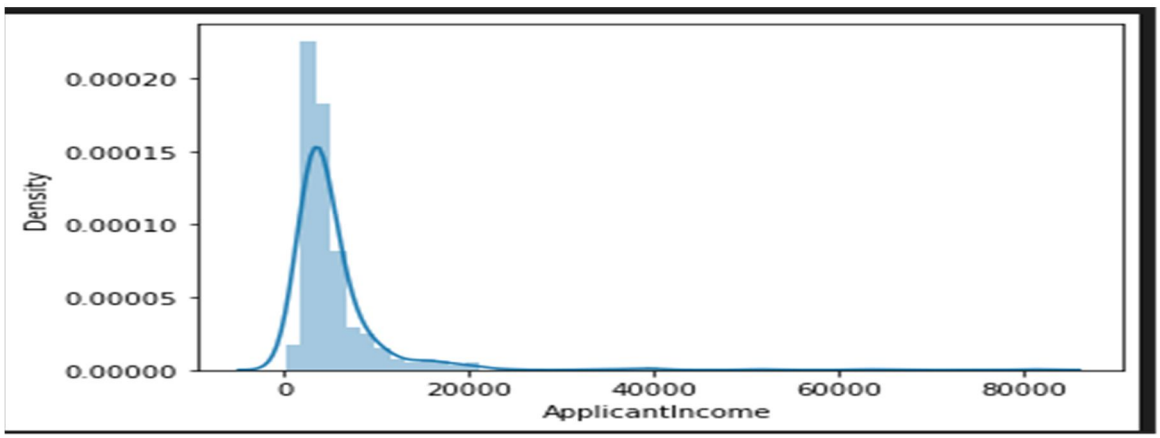
1) Categorical: Mode

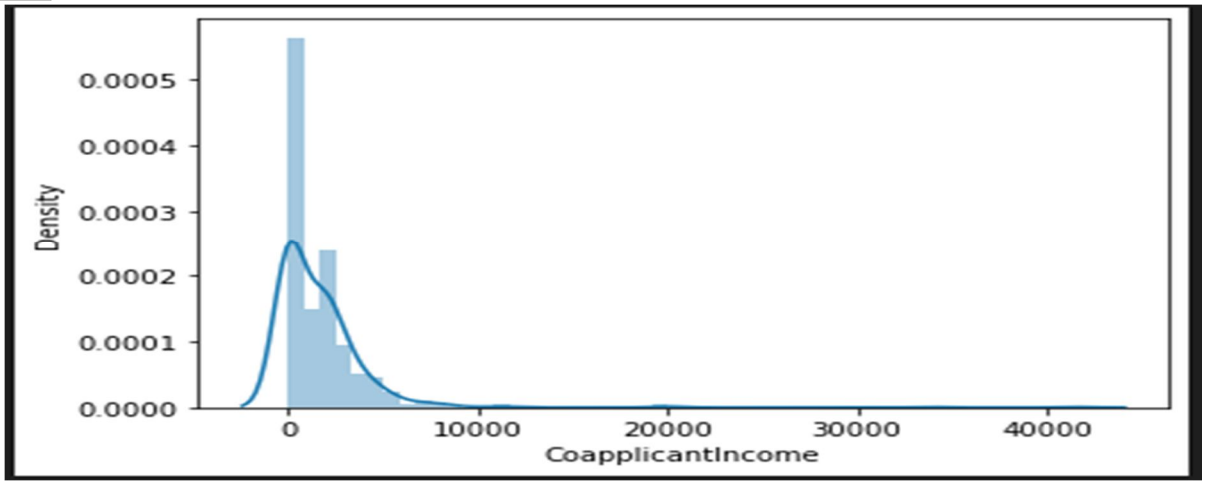
2) Numerical: Median/Mean/Bfill

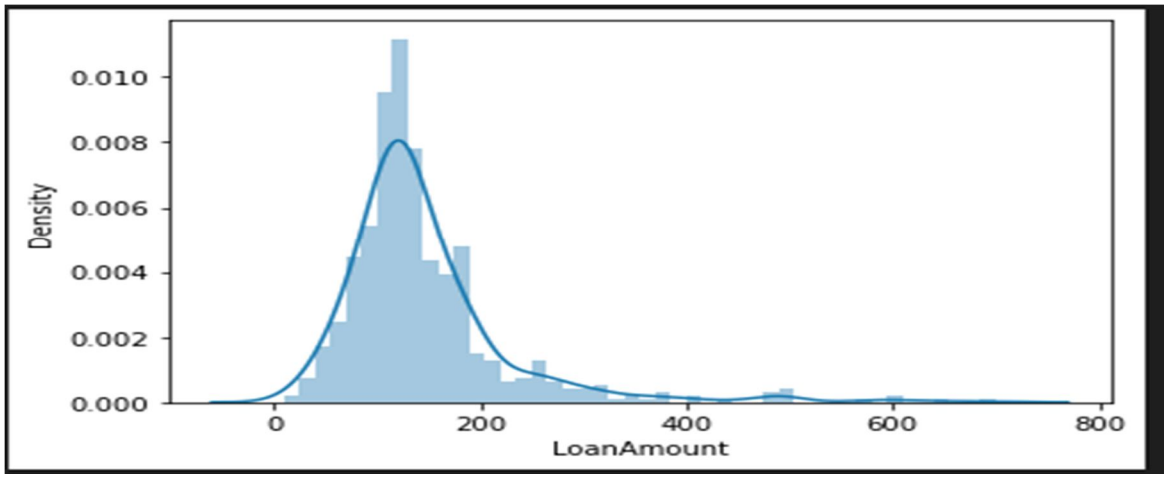
E. Exploratory data analysis

1) Data visualization

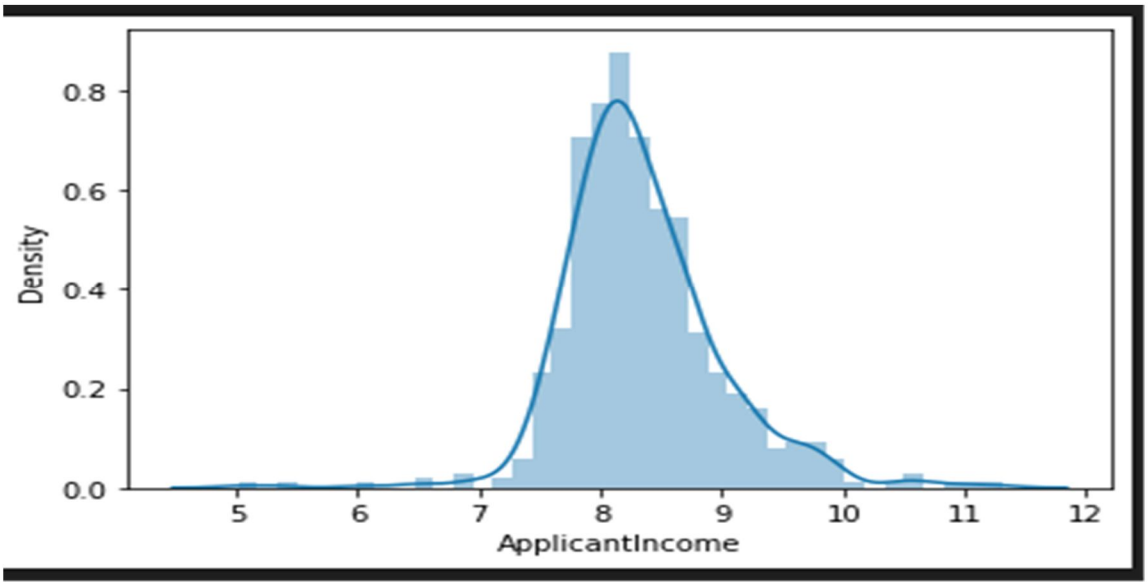
Numerical Data Visualization

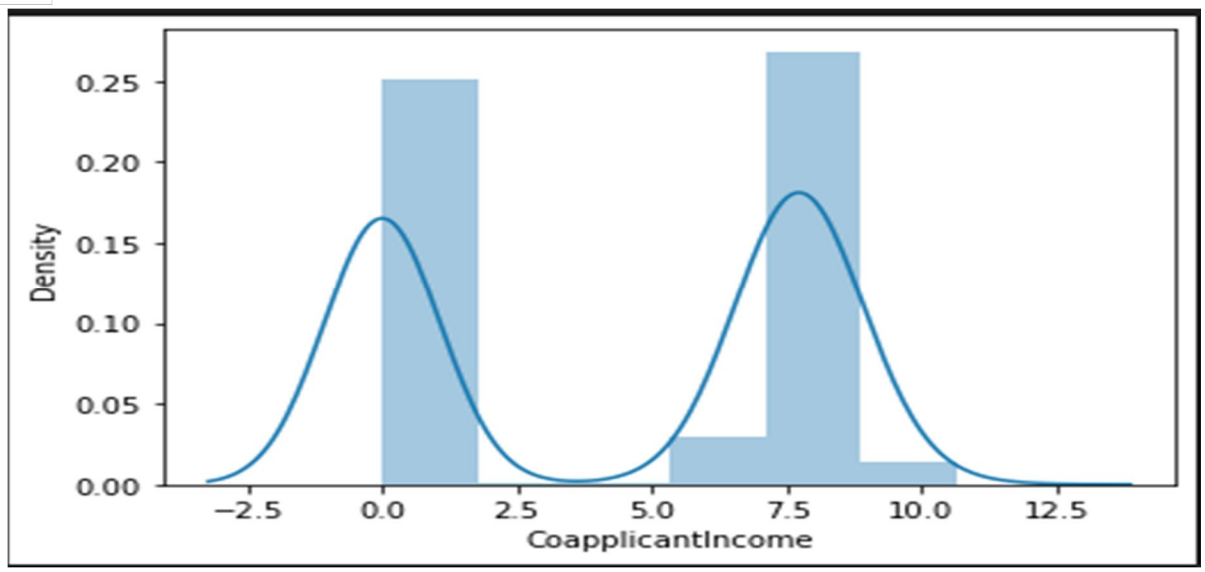


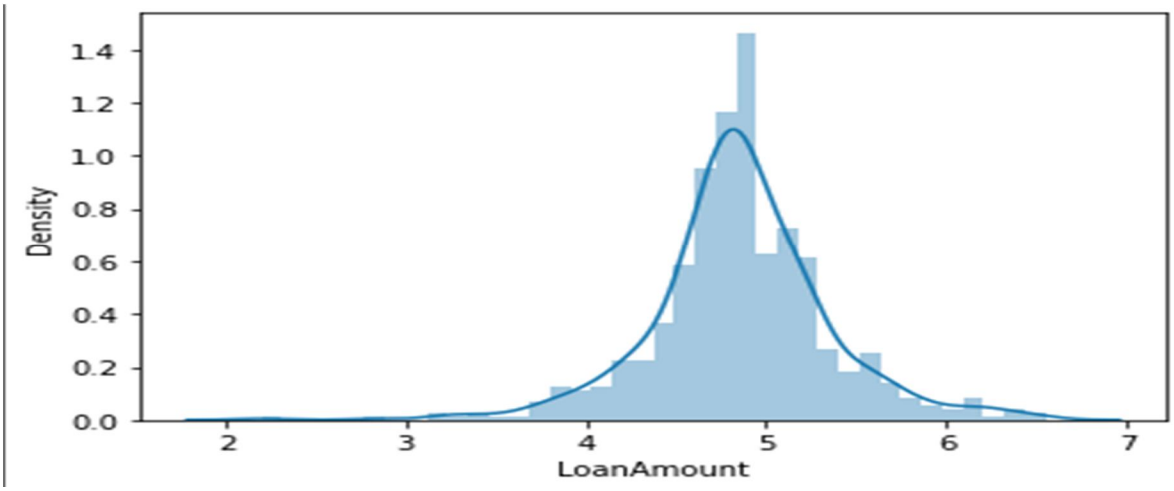




F. Normalization of Data if any Outliers Found



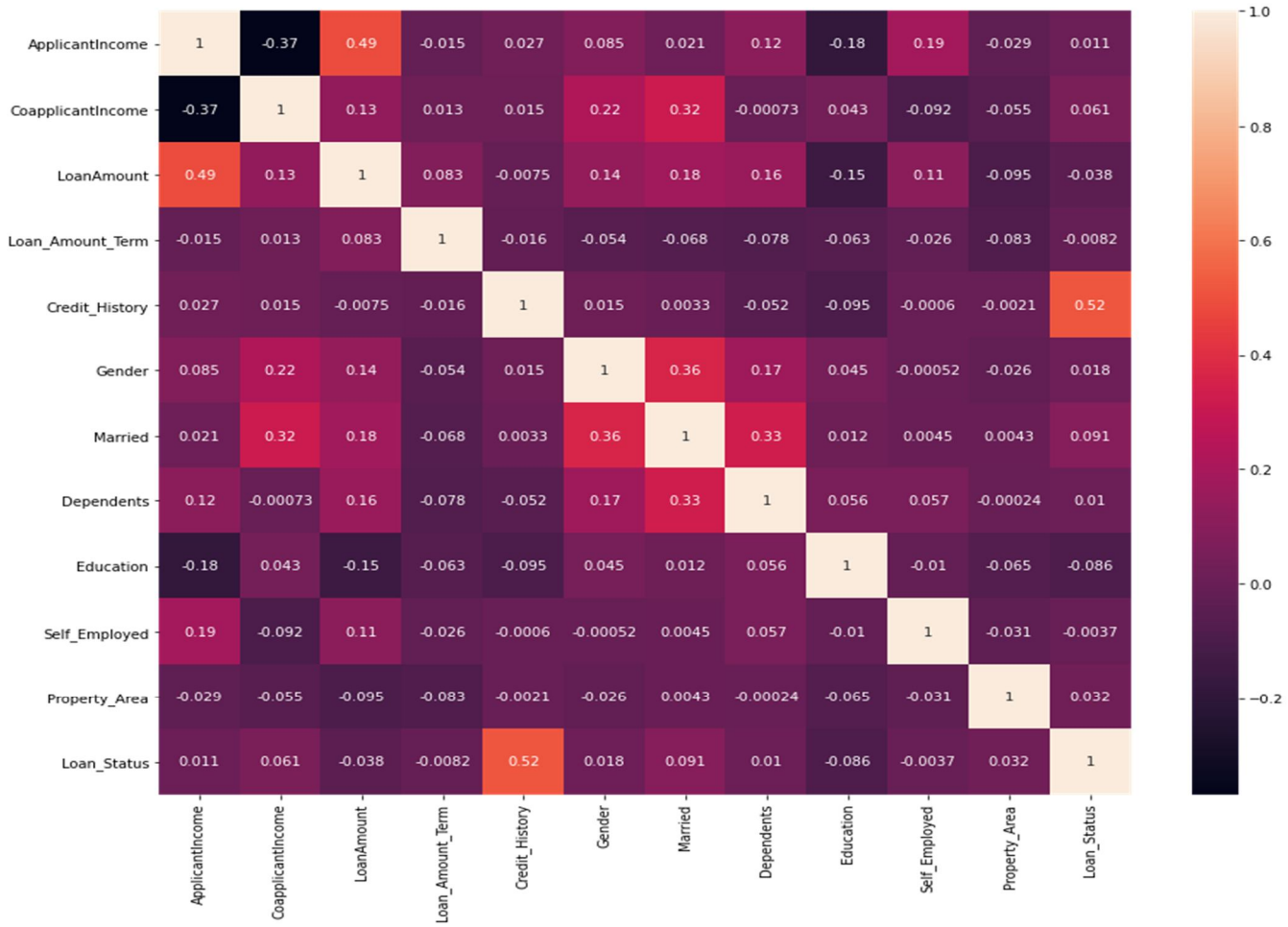




G. Conversion of Object Data Type to Numbers

	0	1	2	3	4
ApplicantIncome	8.674026	8.430109	8.006368	7.856707	8.699515
CoapplicantIncome	0.000000	7.319202	0.000000	7.765993	0.000000
LoanAmount	4.852030	4.852030	4.189655	4.787492	4.948760
Loan_Amount_Term	5.886104	5.886104	5.886104	5.886104	5.886104
Credit_History	1.000000	1.000000	1.000000	1.000000	1.000000
Gender	1.000000	1.000000	1.000000	1.000000	1.000000
Married	0.000000	1.000000	1.000000	1.000000	0.000000
Dependents	0.000000	1.000000	0.000000	0.000000	0.000000
Education	0.000000	0.000000	0.000000	1.000000	0.000000
Self_Employed	0.000000	0.000000	1.000000	0.000000	0.000000
Property_Area	2.000000	0.000000	2.000000	2.000000	2.000000
Loan_Status	1.000000	0.000000	1.000000	1.000000	1.000000

H. Correlation between Different Variables



I. Evaluating different Models based on Different Metrics (Cross Validated Accuracy, Precision, f1 Score, Recall, AUC Curve, Confusion Matrix)

	Accuracy	cv_acc
RandomForestClassifier	0.718919	0.781794
ExtraTreesClassifier	0.708108	0.749220
DecisionTreeClassifier	0.648649	0.688951
LogisticRegression	0.778378	0.802972
KNeighborsClassifier	0.718919	0.749207
SVC	0.686486	0.713301

J. Tuning the Hyper Parameters of the Best Models

K. Feature Importance

IX. RESULT

We have done tests with many users who were satisfied with the interface which was built and were happy to see their eligibility for loans at one place. Many suggested improvements in the filters which are currently being worked upon. The payments from the payment gateway were successful, the information was also stored in the database. The ML model gave almost accurate predictions based on the loan amount , the income of the user and other key factors necessary . The details of the car information being updated reflected on the customer's end realtime. There were no authentication,security issues for the registered users as we used google's firebase. The current filter was able to show the selected cars without hassle. The application was responsive and was also tested on mobile and tablets which had no issues . Required information from the users were properly stored in the Firestore Database without redundancy and in proper collections.

X. CONCLUSION

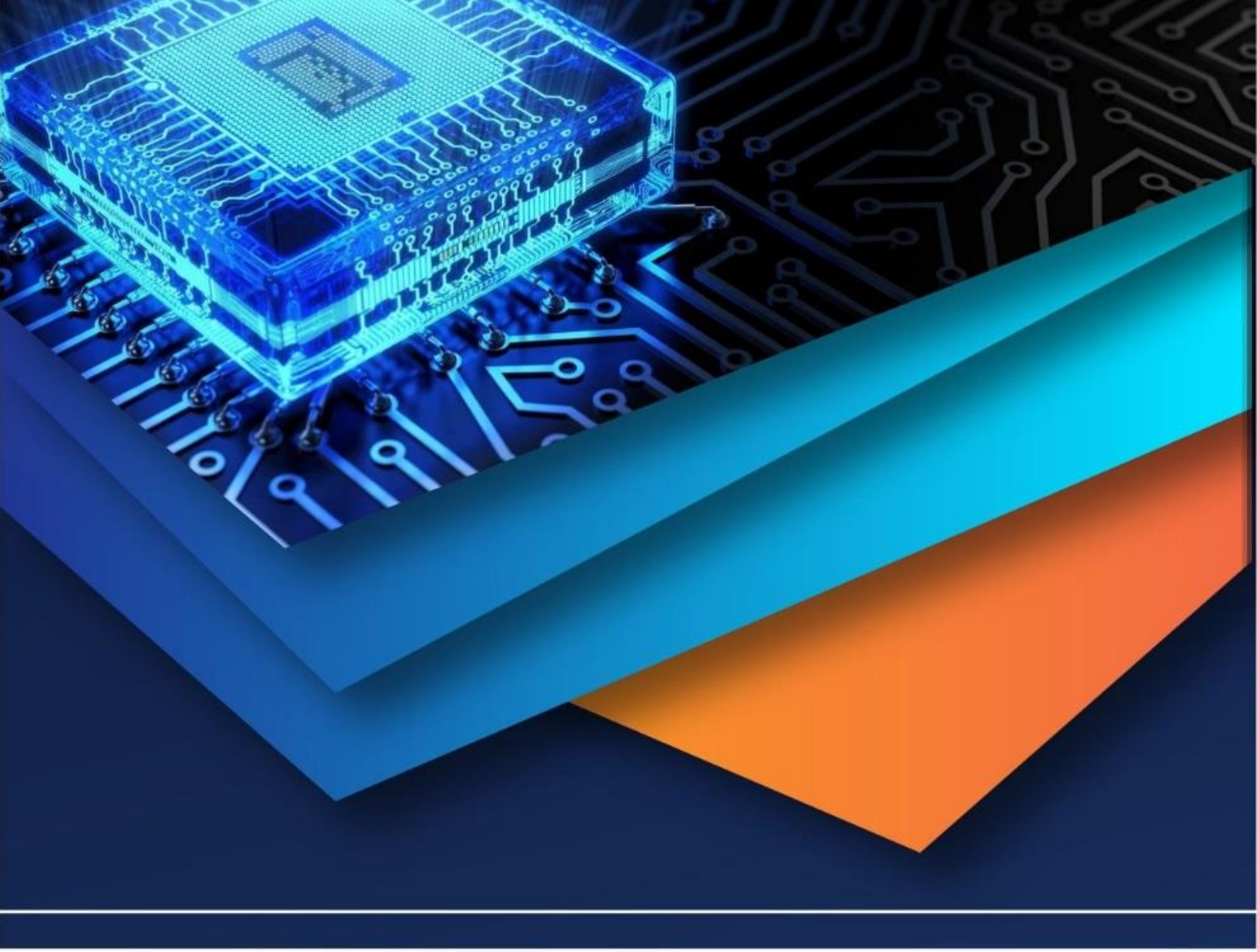
In evolving environments sticking to traditional methods might be good in the short-term but someday they have to update and move towards providing services with the help of technology. Through our project we were able to observe that the majority were happy getting resources at one place. The emphasis of this paper is to connect the users, be it customer or the seller directly with each other and get the customer happy without burdening him and wasting his time.Automobile manufacturers should come together creating better services in reaching out to customers,Tesla in the US has done a great job in doing so. This application that we've built has solved most of the problems the existing system is facing and in the upcoming years we will try to improve on other factors if required .

XI. ACKNOWLEDGEMENT

We would like to thank our guides Associate Prof. Mr .K. Prem Kumar and Associate Prof. Mrs.Soppari Kavitha for their continuous support and guidance. Due to their guidance, we were able to complete our project successfully. Also, we are extremely grateful to Dr. M. V. VIJAYA SARADHI, Head of the Department of Computer Science and Engineering, Ace Engineering College for his support and invaluable time.

REFERENCES

[1] Development of Web and Mobile Application Based Online Buy, Sell and Rent Car System (Shakhawat Hossain Mahi, Umme Habiba Maliha, Sadman Sakib),an IEEE paper
[2] Online car buying(Andrew Joseph IvankovichSteven Dale ZettnerMiguel Antonio Arguello,United States Patent,Patent No.: US 8.204,788 B1)
[3] Car Value Prediction Using Machine Learning(Junaid Mandviwala, Mohd Mubin Girach, Yash Darekar, Yash Patel, Sarang Pandit},2022 JETIR April 2022, Volume 9, Issue 4
[4] Pattabiraman Venkatasubbu, Mukkesh Ganesh, 'Used Cars Price Prediction using Supervised Learning Techniques' (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-1S3, December 2019



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY