# **Project Report Format**

#### 1. INTRODUCTION

## 1.1 Project Overview:

The "Car Purchase Prediction Using ML" project is a cutting-edge application of machine learning that aims to accurately forecast car purchases based on customer data. By utilizing features such as age, income, and historical purchase patterns, the model achieves high predictive accuracy. This innovation assists potential buyers by estimating their likelihood to make a purchase, guiding their decision-making process. The model is integrated into a user-friendly interface for easy use.

Overall, this project has the potential to revolutionize the automotive industry by offering tailored marketing strategies, enhancing customer experiences, and optimizing dealership targeting through data-driven predictions.

## 1.2 Purpose:

The purpose of "Car Purchase Prediction Using ML" is to leverage machine learning techniques to accurately forecast the likelihood of a customer making a car purchase. By analysing various features such as age, income, and historical purchase patterns, the model aims to provide valuable insights to potential buyers, guiding them in their decision-making process.

This technology-driven approach aims to enhance the overall customer experience, optimize marketing efforts, and refine dealership targeting strategies in the automotive industry. Ultimately, the goal is to empower both customers and businesses with data-driven information for more informed and efficient car purchasing decisions.

#### 2. LITERATURE SURVEY

# 2.1 Existing problem:

The existing problem that seeks to address is the uncertainty and lack of personalized guidance that potential car buyers often face. When individuals are in the market for a car, they may struggle to assess their own preferences, budget constraints, and the most suitable options available to them. This uncertainty can lead to prolonged decision-making processes, potential buyer hesitation, and sometimes dissatisfaction with the chosen vehicle after the purchase.

Furthermore, automotive businesses also face challenges in efficiently allocating their marketing resources. Without accurate insights into potential customers' likelihood to make a purchase, they may invest resources in targeting individuals who are less likely to convert, resulting in wasted time and resources.

#### 2.2 References:

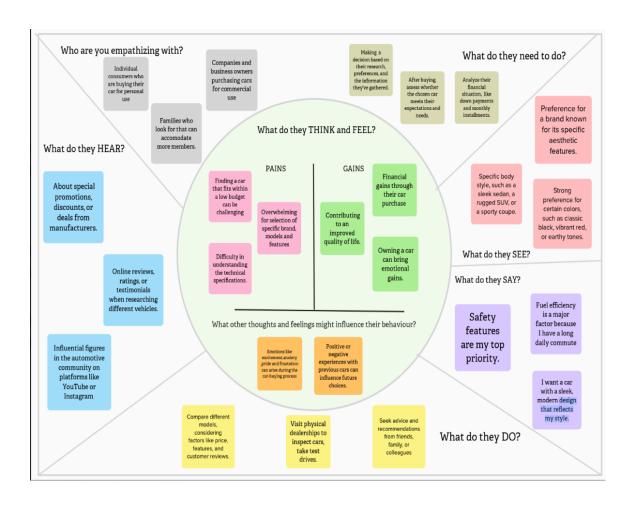
- 1. https://www.temjournal.com/content/81/TEMJournalFebruary2019\_113\_118.pdf
- 2.https://issuu.com/ijraset/docs/price\_prediction\_for\_second\_hand\_cars\_using\_machin
- 3.https://www.researchgate.net/publication/335799148\_Car\_Price\_Prediction\_Using\_Machine Learning
- 4. https://towardsdatascience.com/predicting-car-price-using-machine-learning-8d2df3898f16

### 2.3 Problem Statement Definition:

The goal of this project is to develop a machine learning model that accurately predicts the likelihood of a potential customer making a car purchase. Leveraging features such as age, income, and historical purchase patterns, the model aims to provide valuable insights to assist potential buyers in their decision-making process. The project seeks to revolutionize the automotive industry by offering tailored marketing strategies, enhancing customer experiences, and optimizing dealership targeting through data-powered predictions. This problem statement clearly defines the objective, the features to be used, and the expected outcomes of the project. It provides a clear roadmap for the development and implementation of the machine learning solution.

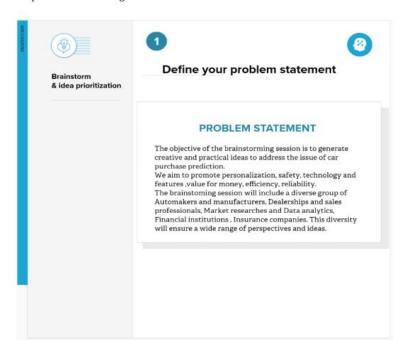
#### 3. IDEATION & PROPOSED SOLUTION

# 3.1 Empathy Map Canvas:

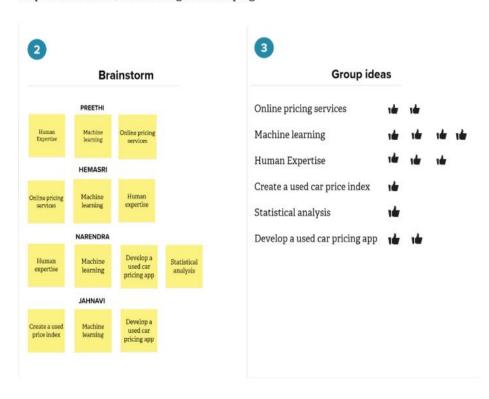


# **Brainstorming for Car purchase Prediction**

Step-1: Team Gathering and Select the Problem Statement.



Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization



#### Description as to why we have chosen ML as the first priority:

The automotive industry generates and collects an enormous amount of data and ML can effectively analyse and make sense of this data to provide valuable insights.ML can provide data-driven insights that go beyond human intuition. By analysing historical data, it can identify patterns and correlations. This can help buyers find the most cost-effective options by considering factors such as long-term ownership costs, fuel efficiency, and resale value. It provide a competitive edge. Predictive models can help dealerships and manufacturers stay ahead of market trends and offer buyers the best options. It can promote eco-friendly car choices by recommending hybrid or electric vehicles, contributing to sustainability efforts. In summary, using Machine Learning for car purchase prediction is a logical choice because it leverages data, personalization, and

advanced algorithms while helping businesses stay competitive in a rapidly evolving industry. It enhances decision-making, takes advantage of vast data resources, and aligns with the growing emphasis on data-driven insights and personalization in the modern world.

## 4. **REQUIREMENT ANALYSIS**

## 4.1 Functional requirements:

In the realm of predicting car purchases, the process of requirement analysis involves a meticulous examination of both functional and non-functional aspects to ensure the development of an effective and efficient predictive model. Functionally, the system must possess the capability to accurately forecast and analyse the factors influencing a customer's decision to purchase a particular car. This involves the identification and integration of key variables such as the customer's financial status, preferences, past purchasing behaviour, and current market trends. The functional requirements extend to the seamless processing of data, employing machine learning algorithms to identify patterns, and providing a user-friendly interface for both customers and administrators.

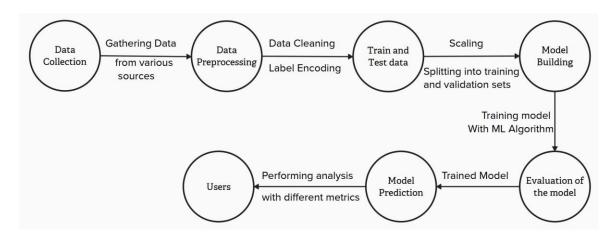
## 4.2 Non-Functional requirements:

On the non-functional side, the analysis delves into aspects that define the system's performance, security, and usability. Performance considerations include the speed and accuracy of the predictive model, ensuring that it can handle a diverse range of input data efficiently. Security measures are paramount, safeguarding sensitive customer information and ensuring compliance with data protection regulations. Usability encompasses the user interface's intuitiveness, making it accessible for users with varying levels of technical expertise. Moreover, reliability and scalability are essential non-functional elements. The predictive model should consistently deliver accurate results, instilling confidence in users. Scalability is crucial to accommodate potential growth in data volume and user interactions, ensuring the system's longevity and adaptability to evolving market dynamics.

## 5. **PROJECT DESIGN**

# 5.1 <u>Data Flow Diagrams & User Stories:</u>

## **Data Flow Diagram:**



## **User Stories:**

User type	Functional	User	User story/task	Acceptan	Priority	Release
	Requirement	Stor		cecriteria		
		y				
		Number				
Car Salesperson	Project	USN -1	Set up the development	Succesful	High	Sprint 1
	setup&		environment with the	ly		
	infrastructu		requiredtools and	configure		
	re		frameworks to start the	d with all		
			car purchase prediction	necessary		
			project.	toolsand		
				frame		
				works.		
Car	Development	USN-2	Gather a diverse dataset of data for	Gathered a	High	Sprint 1

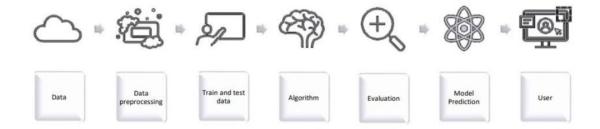
Dealership	environment		training the machine learning model.	diverse datasetof data		
				depicting		
				various		
				factorssuch		
				as		
				age,income etc		
Car Buyers	Data collection	USN-3	Preprocess the collected dataset by	Preproces	High	Sprint 2
			cleaning the data, label encoding	sed the		
			and splitting it into training and validation sets.	dataset		
Researchers	Data	USN-4	Explore and evaluate different	We	High	Sprint 2
and	preprocessing		machine learning architectures	could		
Academics			(e.g., Linear regression) to select	explor		
			the most suitable model for car	e		
			purchase prediction.	various		
				ML		
				models		
System	Model	USN-5	Train the selected machine	We could	High	Sprint 3
administrator	development		learning model using the	do		
			preprocessed dataset	validatio		
			and monitor its performance on	n		
			the validation set.			
Educational	Training	USN-6	Implement data augmentation	We could	Medium	Sprint 3
Institutions			techniques (e.g., rotation, flipping)	do		
			to improve the model's robustness	testing.		
			and accuracy.			
	Model	USN-7	Deploy the trained machine	We	Medium	Sprint 4
	Deployment &		learning model as an API or web	could		
	Integration		service to make it accessible for	check		
			car purchase prediction. Integrate	the		
			the model's API into a user-	scalabi		
			friendly web interface for users to	lity		
			input their data and check for the			
			car purchase prediction results.			

Testing &	USN-8	Conduct thorough testing of	We could	Medium	
quality		themodel and web interface	create web		
assurance		to identify and report any	applicatio		
		issues or	n		
		bugs. Fine-tune the model			
		hyperparameters and optimize			
		itsperformance based on user			
		feedback and testing results.			

## 5.2 Solution Architecture:

It optimizes the car purchase prediction process by leveraging Linear regression for realtime data preprocessing. It not only enhances price prediction but also contributes to interpretability, feature selection, price trend analysis. It also improves the data-driven negotiation and risk mitigation. Our solution leverages Linear Regression to address the car purchase prediction problem effectively:

- Data Gathering.
- Data Preprocessing.
- Model Building.
- Car Purchase Prediction.
- Real Time Analysis.



## 6. PROJECT PLANNING & SCHEDULING

## 6.1 Technical Architecture:

The Deliverable shall include the architectural diagrams below and the information as per the table 1& table 2.

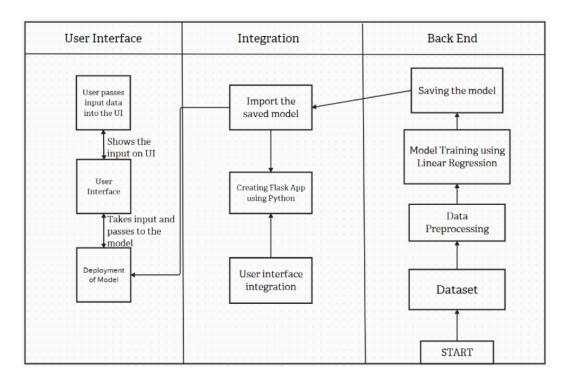


Table-1: Components & Technologies:

S.No	Component	Description	Technology
1.	User Interface	How user interacts	HTML, CSS,
		with application e.g.	JavaScript / Angular
		Web UI	Js /ReactJs etc.
2.	Application Logic-1	Logic for a process in the application	Java / Python
3.	Database	Collect the Dataset	File Manager, MySQL,
		Based on the Problem	NoSQL, etc.
		Statement	
4.	File Storage/ Data	File storage requirements for	Local System, Google
		Storing the dataset	Drive Etc
5.	Frame Work	Used to Create a web	Python Flask,Django etc
		Application, Integrating	
		Frontend and Back End	
6.	Deep Learning Model	Purpose of Model	Logic regression, ridge
			regression etc.
7.	Infrastructure (Server /	Application Deployment on	Local, CloudFoundry,
	Cloud)	Local System / CloudLocal	Kubernetes, etc.
		ServerConfiguration:	
		Cloud Server Configuration :	

**Table-2: Application Characteristics:** 

S.No	Characteristics	Description	Technology
1.	Open-Source Frameworks	The open source frameworks used are scikit-learn,tensor flow,keras,pytorch,pandas,sea born and mathplotlib and jupyter notebook.	Python's Flask
2.	Security Implementations	Security implementations used are Data Encryption, Secure APIs,Authentication and Authorization,Input Validation,Regular Security Audits and Penetration Testing.	e.g. SHA-256, Encryptions, IAM Controls, OWASP etc.
3.	Scalable Architecture	Scalable Architectures used are microservices	Technology used

# 6.2 Sprint Planning & Estimation:

Sprint	Functiona l Requirem ent	User Story Numbe r	User Story/Task	Story points	Priority	Team Member s
Sprint 1	Project setup &Infrastruct ure	USN-1	Set up the development environmentwith the required tools and frameworks to start the car purchaseprediction.	5	High	Jahnavi
Sprint 1	Developm ent environme nt	USN-2	Gather a diverse dataset of data for training the machine learning model.	5	High	Hemasri
Sprint 2	Data collection	USN-3	Preprocess the collected dataset bycleaning the data, label encoding and splitting it into training and validation sets.	5	High	Preethi
Sprint 2	Data preprocessi ng	USN-4	Explore and evaluate different machine learning architectures (e.g., Linear regression) to selectthe most suitable model for car purchase prediction.	5	High	Narendra
Sprint 3	Model developm ent	USN-5	Train the selected machine learning model using the preprocessed dataset and monitor its performance on the validation set.	10	High	Narendra
Sprint 3	Training	USN-6	Implement data augmentation techniques (e.g., rotation, flipping)to improve the model's robustness and accuracy.	5	Medium	Jahnavi
Sprint 4	Model Deployment & Integration	USN-7	Deploy the trained machine learning model as an API or web service to make it accessible for car purchase prediction. Integratethe model's API into a user-friendly web interface for users to	10	Medium	Hemasri

		input their data and check for the car purchase prediction results.			
Sprint 5	Testing & quality assurance	Conduct thorough testing of the model and web interface to identify and report any issues or bugs. Fine-tune the model hyperparameters and optimize its performance based on user feedback and testing results	5	Medium	Preethi

## 7. CODING & SOLUTIONING

## Dataset:

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb

df = pd.read_csv("car_data.csv")
df
```

#### Out[1]:

	User ID	Gender	Age	AnnualSalary	Purchased
0	385	Male	35	20000	0
1	681	Male	40	43500	0
2	353	Male	49	74000	0
3	895	Male	40	107500	1
4	661	Male	25	79000	0
995	863	Male	38	59000	0
996	800	Female	47	23500	0
997	407	Female	28	138500	1
998	299	Female	48	134000	1
999	687	Female	44	73500	0

1000 rows × 5 columns

## Preprocessing:

```
In [9]: #checking for null values
         df.isnull().any()
 Out[9]: User ID
                           False
          Gender
                           False
          Age
                           False
          AnnualSalary
                           False
          Purchased
                           False
          dtype: bool
In [10]:
         df = df.drop(columns = ['User ID'],axis = 1)
         df.head()
Out[10]:
             Gender Age Annual Salary Purchased
               Male
                      35
                               20000
                               43500
                                             0
           1
               Male
                      40
           2
               Male
                      49
                               74000
                                             0
                      40
                              107500
               Male
               Male
                     25
                               79000
```

## **Encoding:**

```
In [12]: from sklearn.preprocessing import LabelEncoder

#Label encoding
le = LabelEncoder()

df.Gender = le.fit_transform(df.Gender)
df
```

#### Out[12]:

	Gender	Age	Annual Salary	Purchased
0	1	35	20000	0
1	1	40	43500	0
2	1	49	74000	0
3	1	40	107500	1
4	1	25	79000	0
995	1	38	59000	0
996	0	47	23500	0
997	0	28	138500	1
998	0	48	134000	1
999	0	44	73500	0

1000 rows × 4 columns

#### Scaling:

```
In [15]: #Scaling on independent variables
             from sklearn.preprocessing import MinMaxScaler
scale =MinMaxScaler()
            X_scaled= pd.DataFrame(scale.fit_transform(X),columns =X.columns)
X_scaled.head()
Out[15]:
                               Age Annual Salary
                     1.0 0.377778
                      1.0 0.488889
                      1.0 0.688889
                                          0.429091
                      1.0 0.488889
                                           0.672727
                     1.0 0.155556 0.465455
In [16]: scalar = scale.fit_transform(X)
             scalar
                                     , 0.37777778, 0.03636364],
, 0.48888889, 0.20727273],
, 0.688888889, 0.42909091],
Out[16]: array([[1.
                       [1.
                      ...,
[0.
                                     , 0.22222222, 0.89818182],
, 0.66666667, 0.86545455],
, 0.57777778, 0.42545455]])
                       [0.
```

#### Splitting data and model building:

#### Prediction:

```
In [20]: d_y_predict = model1.predict(X_test)
d_y_predict
Out[20]: array([0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0,
             0,
                       1, 1, 0,
                              0, 0, 0,
                                      1, 0, 0,
                                             0, 0, 1,
                                                    0, 1, 1,
                  0,
                                                    1, 0,
                                                    1, 1,
               ø,
                  0, 0, 1, 0, 0,
                              1, 0, 1,
                                      0, 1,
                                             0, 0, 1,
                                                    0, 0,
                                                  0,
               0, 1, 1, 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 1, 1, 0, 0, 1, 0, 1, 0, 1, 1,
                                             1, 0,
                                                    1, 0,
                                             1, 0, 1, 0, 0, 1, 0,
               1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0,
                                                    1, 0, 0,
             0, 1, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0,
             1, 0, 1, 1, 1, 0, 1, 1, 0
0, 0, 0, 0], dtype=int64)
                                 0,
                                   0,
                                      0, 0,
                                           1,
                                             0,
                                                1,
                                                    0,
In [21]: d_y_predict_train = model1.predict(X_train)
```

#### 8. **PERFORMANCE TESTING**

## 8.1 Performance Metrics:

```
from sklearn.metrics import accuracy_score,classification_report,confusion_matrix

print('Testing Accuracy = ', accuracy_score(Y_test,d_y_predict))
print('Training Accuracy = ', accuracy_score(Y_train,d_y_predict_train))

Testing Accuracy = 0.895
Training Accuracy = 0.91
```

#### 9. **RESULTS**

## 9.1 Output Screenshots:

## Model prediction:

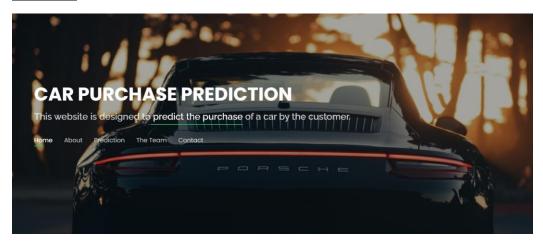
```
In [28]: from sklearn.tree import DecisionTreeClassifier

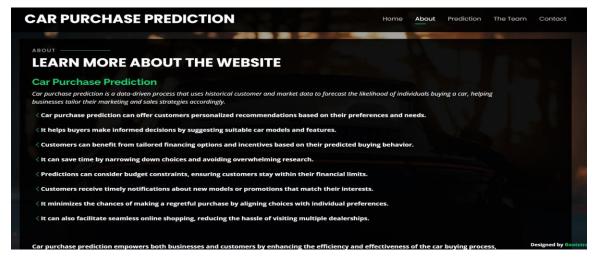
DT_model = DecisionTreeClassifier()
DT_model.fit(X_test,Y_test)

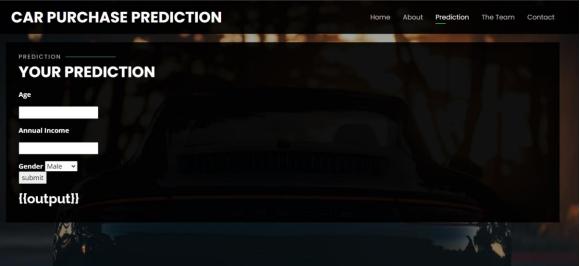
prediction = DT_model.predict((np.array([[1.0,0.377778,0.036364]])))

if prediction[0] == 1:
    print("Purchased")
else:
    print("0")
```

# Website:







#### 10. ADVANTAGES & DISADVANTAGES

## Advantages:

Market Understanding: Car purchase prediction can provide valuable insights into consumer behaviour and preferences. It helps car manufacturers, dealers, and marketers understand which models, features, or price points are in demand.

Financial Planning: Consumers can use car purchase predictions to plan their finances better. They can anticipate the cost of ownership, including monthly payments, insurance, and maintenance, and make informed decisions.

Reduction in Wasted Time and Effort: Car sales personnel can focus their efforts on potential buyers more likely to make a purchase, reducing the time and resources spent on customers who are less likely to buy.

## **Disadvantages:**

Privacy Concerns: Predictive models often rely on personal data, raising concerns about privacy and data security. Consumers may be uncomfortable with the collection and use of their personal information for marketing purposes.

Accuracy and Uncertainty: No predictive model is perfect, and there is always a margin of error. Predictions can be inaccurate, leading to missed opportunities or unfulfilled expectations.

Overreliance on Data: Overreliance on predictive models can lead to overlooking intangible factors, such as a customer's emotional connection to a specific car brand or model.

Cost of Implementation: Developing and maintaining predictive models can be expensive. Small businesses or dealerships with limited resources may find it challenging to implement and update such systems.

#### 11. CONCLUSION

In conclusion, predicting a car purchase is a complex and multi-faceted endeavour that relies on various factors and variables. These factors range from personal preferences and financial capabilities to economic conditions and technological advancements. While data-driven models and algorithms can aid in forecasting car purchases to a certain extent, it is crucial to recognize that human behaviour and decision-making are influenced by a multitude of intricate and often unpredictable factors.

In the long term, the accuracy and reliability of car purchase predictions may improve as technology advances and the availability of data continues to grow. Machine learning algorithms and artificial intelligence will likely play an increasingly significant role in analysing and interpreting the vast amounts of data generated by individuals and the automotive industry. This may lead to more personalized and targeted marketing efforts, as well as more efficient resource allocation for car manufacturers and dealerships.

Nonetheless, it is essential to remember that the human element in car purchasing decisions remains central. Emotions, individual preferences, lifestyle changes, and unforeseen external events can all sway someone's decision to buy a car. Therefore, while predictive models and data analytics can provide valuable insights, they should be used in conjunction with a deep understanding of the customer, their unique needs, and their evolving circumstances.

In summary, the future of car purchase prediction lies in a synergy of data-driven analysis and a nuanced understanding of the human factor. As we move forward, the automotive industry must continue to adapt to changes in consumer behaviour, technology, and economic conditions to make increasingly accurate predictions, ultimately improving the car buying experience for both consumers and industry professionals.

#### 12. FUTURE SCOPE

The future scope of car purchase prediction holds immense promise, driven by a combination of technological advancements, data availability, and evolving consumer preferences. As we look ahead, several key trends and developments are likely to shape the landscape of car purchase prediction.

First and foremost, the proliferation of connected vehicles and the Internet of Things (IoT) will continue to generate an unprecedented amount of data. These data sources, ranging from in-car sensors to smartphone apps, will provide a wealth of real-time information on how individuals use their vehicles, their driving habits, and their preferences. This data will not only enhance the accuracy of car purchase predictions but also enable the development of highly personalized recommendations and offers for potential buyers.

Artificial intelligence and machine learning will play a pivotal role in harnessing this data. Advanced algorithms will be able to process and interpret vast datasets quickly, identifying subtle patterns and correlations that might elude human analysts. As these algorithms become more sophisticated, they will refine their predictive capabilities, considering not only historical data but also real-time factors, such as weather conditions, traffic patterns, and even the driver's current mood, to make more accurate recommendations.

In the long term, we can anticipate a convergence of all these trends, resulting in highly sophisticated car purchase prediction systems that seamlessly integrate AI-driven data analytics, AR/VR experiences, eco-friendly options, and personalized mobility solutions. These systems will not only benefit consumers by offering tailored recommendations but also empower the automotive industry to optimize production, marketing, and distribution, ultimately improving the overall car buying and ownership experience. However, it is essential to remember that with these advancements, issues related to data privacy, security, and ethical considerations must be addressed to ensure that car purchase prediction remains a positive force in the automotive industry's future.

#### 13. APPENDIX

Source Code

HTML:

```
| chody | cheader id="header" | cheader id="header id="header" | cheader id="header id="header id="header id="header id="header id="header id="header id="header" | cheader id="header id="
```

```
About Section id-"about" class="about">

div class="about me container">

div class="cettion-title">

db2aAbout(Np)

goleann more about the website
(div)

div class="cettion-lig-11 pt-4 pt-1g-0 content" data-aos="fade-left">

db2aAbout(Np)

div class="fis-taile">

db3aCar Purchase prediction is a data-driven process that uses historical customer and market data to forecast the likelihood of individuals buying a car, helping businesses tailor their marketing and sales strategies accordingly.

(p)

div class="con">

div class="con">

div class="con">

div class="con">

div class="bi bi-cherron-left">

dls-d cl
```

```
| Color | California | Californ
```

#### CSS:

```
| Sastet | Ses | F | Sylects | Syle
```

```
#header h1 a,
#header h1 a,
#header h1 a,
#header h1 ainover {

color: #fff;
line-height: 1;
display: inline-block;
}

#header h2 {
color: #rgba(255, 255, 0.8);

#header h2 {
color: #rgba(255, 255, 0.8);

#header h2 span {
color: #rgba(255, 255, 0.8);

#header h2 span {
color: #rfff;
border-bottom: 2px solid ##18d26e;
padding-bottom: 6px;

#header img {
padding: 0;
margin: 0;

#header .social-links {

margin-top: 40px;
display: flex;

#header .social-links a {

font-size: 16px;
display: flex;

justify-content: center;
align-items: center;
align-item
```

#### Flask app:

```
app1.py X
app1.py > ...
from flask import Flask,render_template,request
        import pickle
import numpy as np
from sklearn.preprocessing import MinMaxScaler
        app = Flask( name )
        model = pickle.load(open('Car prediction.pkl','rb'))
scale = pickle.load(open('MinMax.pkl','rb'))
        @app.route('/')
def start():
             return render_template('index1.html')
        def login():
              age = request.form["age"]
annual_income = request.form["annualincome"]
gender = request.form["gender"]
           t = np.array([[age,annual_income,gender]])
           t_scaled= scale.fit_transform(t)
output =model.predict(t_scaled)
                   return render_template('index1.html', output="Not purchasble")
return render_template("index1.html")
                annual_income = request.form["annualincome"]
gender = request.form["gender"]
            t = np.array([[age,annual_income,gender]])
           output =model.predict(t_scaled)
print(output)
                   return render_template('index1.html', output="Purchasble")
                   return render_template('index1.html', output="Not purchasble")
return render_template("index1.html")
              app.run(debug=True)
```

#### GitHub & Project Demo Link:

https://github.com/smartinternz02/SI-GuidedProject-600765-1697471673