CAR PURCHASE PREDICTION USING MACHINE LEARNING

Team ID -593155

Nysa Singh
Harsh Agrawal
Utkarsh Shukla
Soumya Darshan Sukla

1. INTRODUCTION

1.1 Project Overview

In the dynamic landscape of the automotive industry, understanding customer behaviour is paramount for manufacturers and dealerships alike. The ability to predict whether a customer will make a purchase decision is not only instrumental in optimising sales strategies but also in enhancing customer experience and satisfaction. Leveraging the potential of machine learning, businesses can gain valuable insights into customer preferences, enabling them to tailor their offerings and services to individual needs.

Our project, titled "Car Purchase Prediction using Machine Learning," delves into the core of this challenge. Focused on predicting customer purchase decisions, our endeavour involves an in-depth analysis of historical data encompassing diverse customer demographics, car specifications, and purchase outcomes. By employing advanced machine learning algorithms, we aim to develop a robust predictive model capable of classifying customers into two distinct categories: those likely to make a purchase and those who are not.

1.2 Purpose

- Data-Driven Decision Making:
  - By harnessing machine learning techniques, our goal is to assist automotive businesses in making data-driven decisions. Through rigorous analysis of customer data, we aim to uncover patterns and trends that influence purchase decisions. This knowledge empowers businesses to optimise their marketing efforts, refine inventory management, and deliver personalised customer experiences.
- Enhancing Customer Engagement:
  - Understanding customer preferences at a granular level allows businesses to enhance customer engagement significantly. By tailoring offerings based on individual needs and preferences, businesses can create a seamless and personalised purchasing journey. This not only fosters customer loyalty but also ensures a higher rate of successful conversions.

2. LITERATURE SURVEY

2.1 Existing problem

The automotive industry faces the challenge of efficiently targeting potential customers and optimizing marketing strategies. To address this, our project aims to predict car purchases using customer data. The problem is to provide precise purchase likelihoods to guide decision-making and revolutionize marketing strategies, enhancing overall efficiency and customer experiences in the automotive sector.

2.2 References

Enis Gegic,Dino keco,International Burch University,Car Price prediction using ML Technique 2019,Tem Journal volume 8

https://towardsdatascience.com/predicting-used-car-prices-with-machine-learning-techniques-8a9d8313952

Prajwal Ganorkar,Syam Sharma,International Journal of computer science and engineering,May 2019,volume 7

https://www.researchgate.net/publication/343878698_Used_Cars_Price_Prediction_using_Supervised_Learning_Techniques

Rithvik Raj Mekala1 , Gouru Laxmi Sevitha,Prediction of Price forCars Using Machine Learning,Volume 10 Issue VI June 2022

https://www.ijraset.com/research-paper/prediction-of-price-for-cars-using-machine-learning

2.3 Problem Statement Definition

Our project is focused on helping the automotive industry better identify potential customers and improve their marketing tactics. We aim to predict the probability of customers buying cars by analyzing their data. Our goal is to provide accurate purchase predictions that can assist in decision-making and transform marketing strategies, ultimately leading to increased efficiency and improved customer experiences in the automotive sector.

A car purchase prediction model using machine learning is a data-driven approach to estimate the likelihood of an individual or group of individuals buying a car. It leverages advanced algorithms and statistical techniques to analyze customer data and make predictions based on historical patterns and trends. Here's a more detailed explanation of how such a model typically works:

1. Data Collection: The first step is to gather relevant data from various sources. This data can include customer demographics, past purchase history, online behavior, financial information, and any other relevant variables that may influence a car purchase decision.

2. Data Preprocessing: Once the data is collected, it needs to be cleaned and preprocessed. This involves handling missing values, normalizing or scaling features, and encoding categorical variables. The data is then split into a training set and a testing set for model training and evaluation.

3. Feature Engineering: Feature engineering involves selecting the most relevant features (attributes) for the prediction model. This step may also involve creating new features that provide additional insights into customer behaviour and preferences.

4. Model Selection: Machine learning algorithms are selected based on the nature of the problem and the dataset. Common algorithms for car purchase prediction include logistic regression, decision trees, random forests, support vector machines, and neural networks.

5. Model Training: The selected algorithm is trained on the training dataset, where it learns to make predictions based on the input features. The model tries to find patterns and relationships between customer data and car purchase decisions.

6. Model Evaluation: After training, the model is evaluated using the testing dataset to assess its performance. Common evaluation metrics include accuracy, precision, recall, F1 score, and area under the receiver operating characteristic (ROC AUC) curve.

7. Hyperparameter Tuning: To improve the model's performance, hyperparameter tuning may be performed. This involves adjusting the settings of the algorithm to find the best combination of parameters.

8. Deployment: Once the model demonstrates satisfactory performance, it can be deployed in a production environment. This typically involves integrating it with the organization's marketing and customer relationship management (CRM) systems.

9. Real-time Predictions: In practice, the model can make real-time predictions for individual customers or customer segments. These predictions provide insights into the likelihood of a particular customer making a car purchase.

10. Marketing Strategy Optimization: The predictions generated by the model can guide marketing strategies. For example, marketing campaigns can be targeted towards customers with a high likelihood of making a purchase, personalized recommendations can be offered, and resources can be allocated more efficiently.

11. Continuous Improvement: The model should be periodically retrained and updated as new data becomes available. This ensures that it remains accurate and relevant in a changing market environment.

Overall, a car purchase prediction model using machine learning aims to enhance decision-making in the automotive industry by providing insights into customer behavior and preferences, enabling more effective marketing strategies and improving the overall customer experience.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

3.2 Ideation & Brainstorming

Brainstorm & idea prioritization

Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.

- 10 minutes to prepare
- 1 hour to collaborate
- 2-8 people recommended

Harsh Agrawal - 21BIT0412
Utkarsh Shukla -21BIT0395
Nysa Singh - 21BIT0376
Soumya Darshan Sukla - 21BIT0659

Before you collaborate

A little bit of preparation goes a long way with this session. Here's what you need to do to get going.

10 minutes

Team gathering
Define who should participate in the session and send an invite. Share relevant information or pre-work ahead.

Set the goal
Think about the problem you'll be focusing on solving in the brainstorming session.

Learn how to use the facilitation tools
Use the Facilitation Superpowers to run a happy and productive session.

Open article →

1

Define your problem statement

What problem are you trying to solve? Frame your problem as a How Might We statement. This will be the focus of your brainstorm.

5 minutes

Problem

The automotive industry faces the challenge of efficiently targeting potential car buyers and optimizing marketing strategies. We aim to develop an ML model to predict car purchases using demographic and historical data. This will help automakers and dealerships make data-driven decisions, enhance customer experiences, and improve sales by offering a user-friendly interface for potential buyers to estimate their purchase likelihood.

Key rules of brainstorming

To run an smooth and productive session

- Stay in topic.
- Encourage wild ideas.
- Defer judgment.
- Listen to others.
- Go for volume.
- If possible, be visual.

2

Brainstorm

Write down any ideas that come to mind that address your problem statement.

10 minutes

TIP

You can select a sticky note and hit the pencil [switch to sketch] icon to start drawing!

Harsh Agrawal

- Customer profiling for precise car purchase predictions
- Personalized insights to boost car sales and satisfaction
- Predictive models for data-informed marketing decisions
- Personalized insights to boost car sales and satisfaction
- Auction Based Pricing

Soumya Darshan

- Re-sale Priority
- Current Purchase Value
- Historical Market Trends
- Social Media Sentiment Analysis
- Seasonal Purchase

Utkarsh Shukla

- Customer purchase propensity modeling
- Customer journey optimization
- AI-driven sales recommendations
- Customer Demand
- Smart dealership targeting

Nysa Singh

- Machine learning-driven CRM.
- Automated ad spend allocation
- Loyalty and incentive Programs
- Behavioral recommendation engines.
- Government Regulations: taxes & import duties which influence car prices.

3

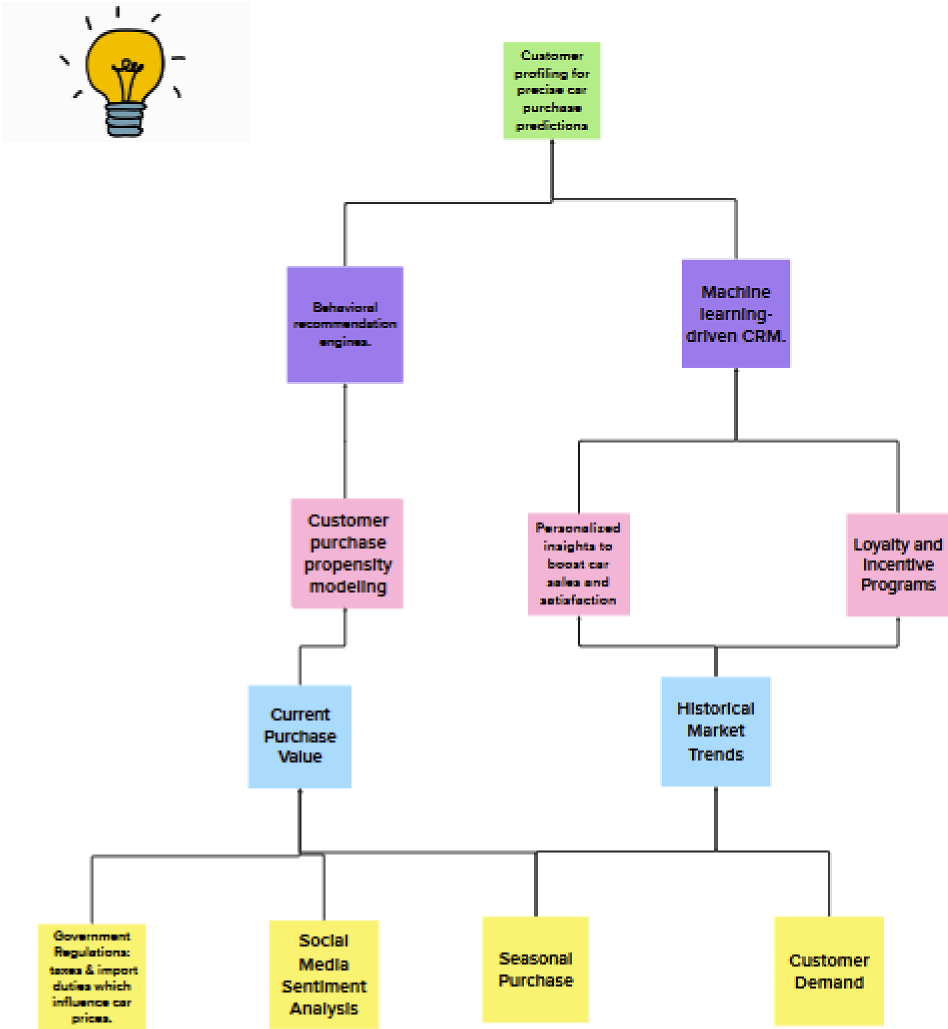
Group ideas

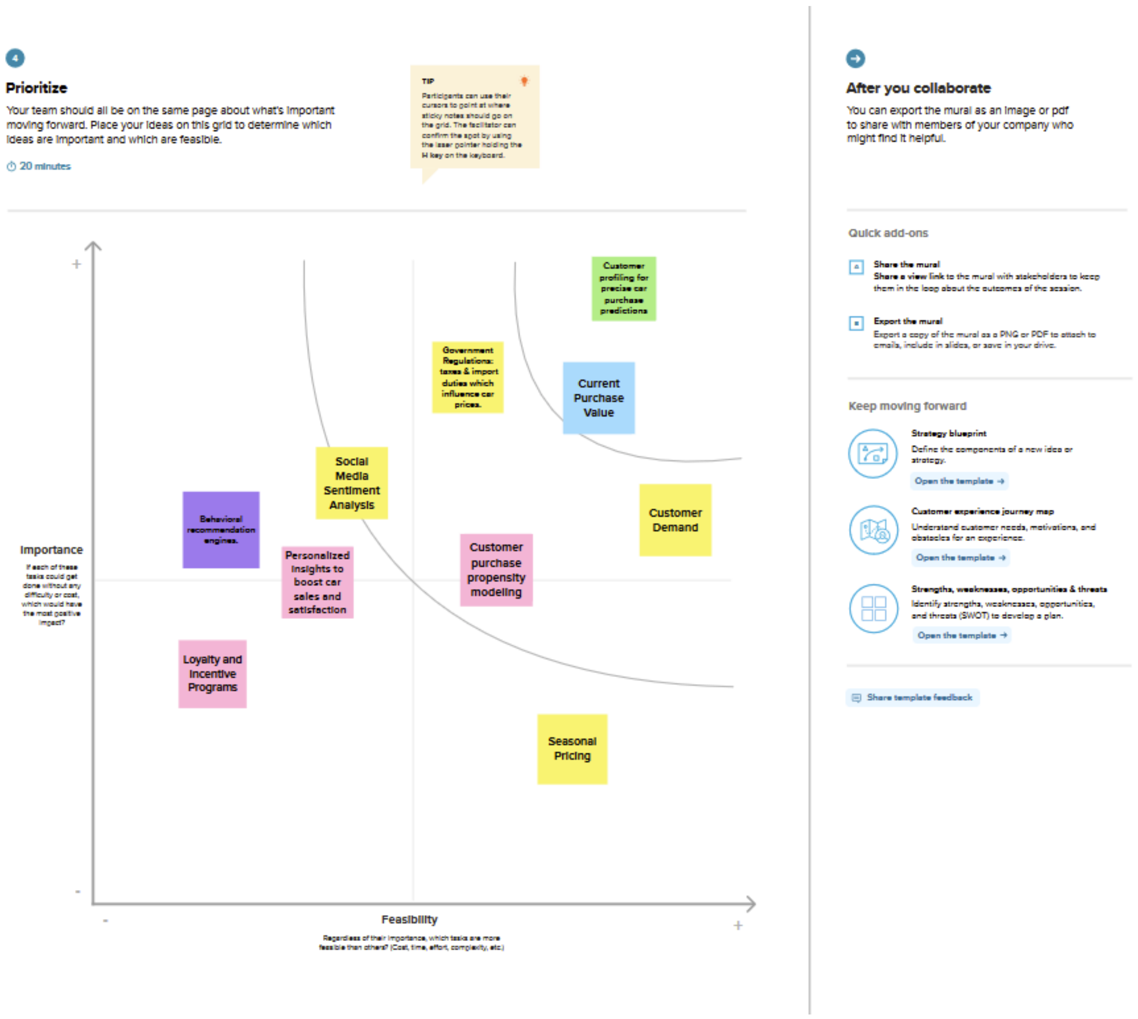
Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

20 minutes

TIP

Add customizable tags to sticky notes to make it easier to find, browse, organize, and categorize important ideas as themes within your mural.





4. REQUIREMENT ANALYSIS

4.1Functional Requirements:

- Data Collection:
    - The system must proficiently collect relevant data from diverse sources, encompassing customer demographics, car specifications, and historical purchase records.
- Data Preprocessing:
    - The system must employ robust techniques to clean and preprocess the collected data, addressing missing values, outliers, and ensuring data consistency and integrity.
- Feature Selection:
    - The system should autonomously identify pertinent features with significant impact on predicting customer purchase decisions, optimizing the model's performance.
- Machine Learning Model:

    - Implementation of a sophisticated machine learning algorithm for binary classification, ensuring the model is trained meticulously on the preprocessed data.
- Prediction:
    - The system should accurately predict whether a customer is inclined to make a car purchase based on the provided input features, providing reliable and insightful results.

4.2 Non-Functional Requirements:

The non-functional requirements essential for our car purchase prediction system encompass the following aspects:

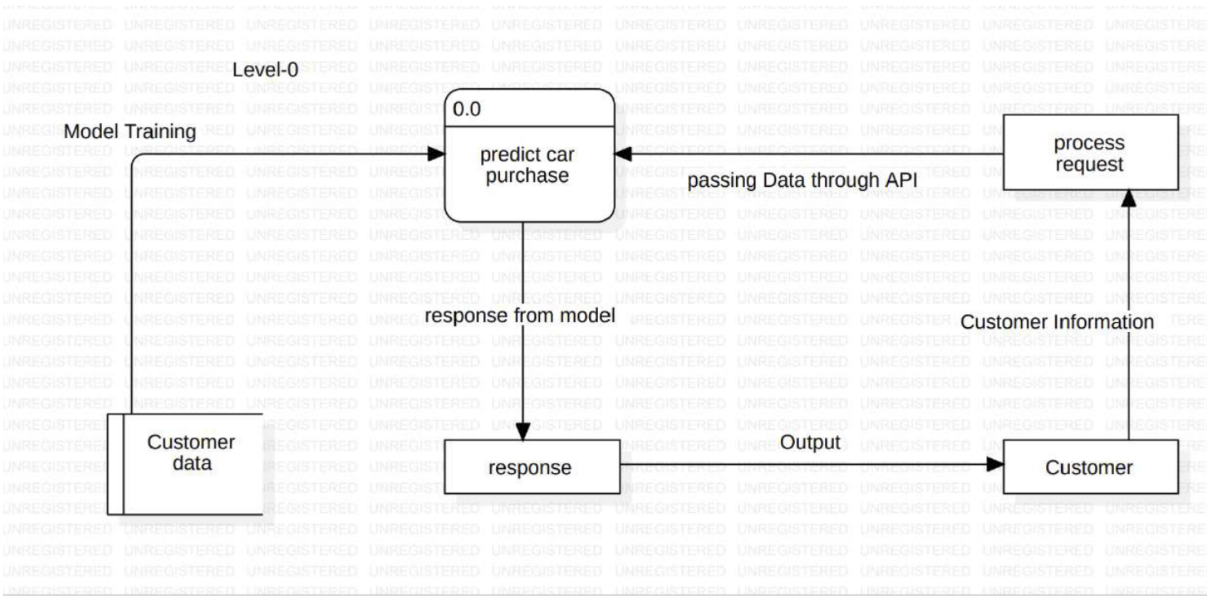
- Performance:
    - The system must demonstrate exceptional performance, handling substantial data volumes efficiently, and delivering predictions with minimal latency.
- Accuracy:
    - The model should exhibit a high level of accuracy, minimizing false positives and false negatives to enhance the reliability of predictions.
- Scalability:

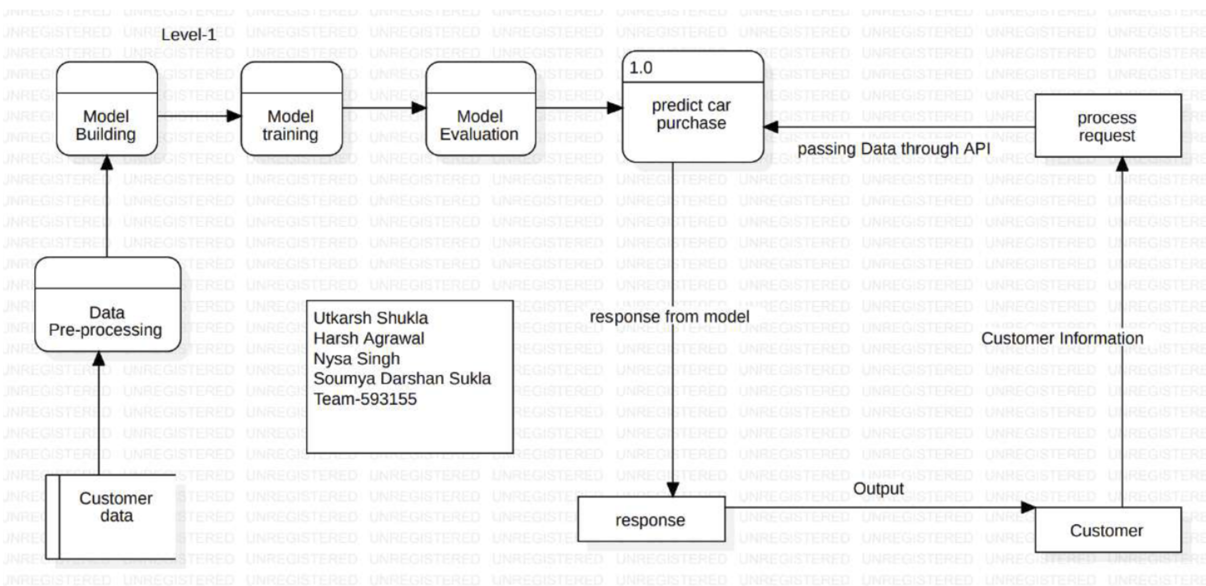
    The system must be scalable, allowing seamless integration with diverse data sources and accommodating increasing data volumes while maintaining optimal performance levels.

- Security:
    - Stringent security measures must be in place to safeguard customer data, preventing unauthorized access, and ensuring data integrity and confidentiality.
- Usability:
    - The user interface, if applicable, must be intuitively designed, facilitating seamless data input and enabling users to interpret prediction results easily.
- Reliability
    - The system should be reliable and consistently available, minimizing downtime and ensuring uninterrupted performance even under varying operational conditions.

5. PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories



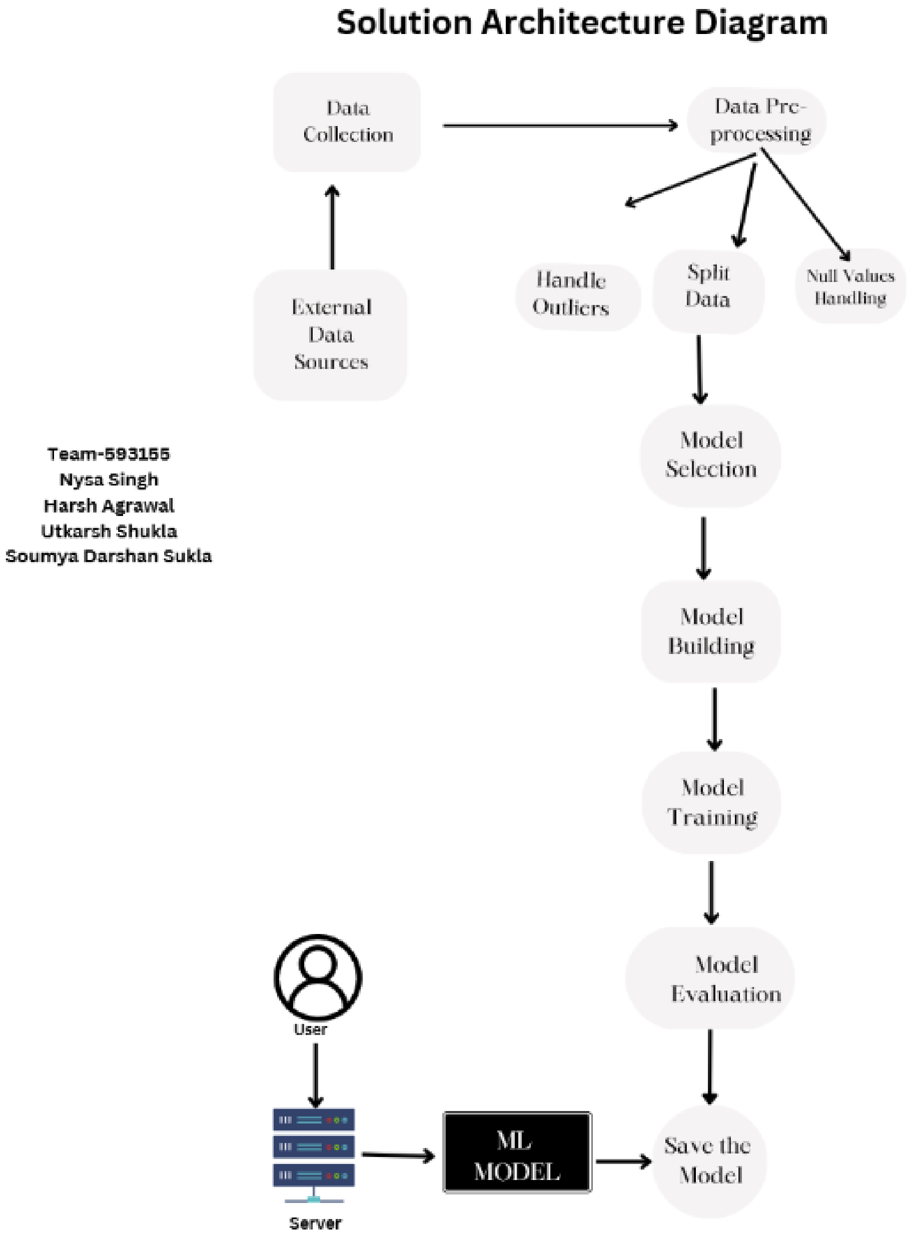


User Stories

Use the below template to list all the user stories for the product.

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Customer (Mobile user)	Registration	USN-1	As a user, I can provide additional information during registration, such as age, income, and purchase history.	I can access my account / dashboard	Medium	Sprint-2
	Dashboard	USN-2	As a user, I can log in to my dashboard to view my estimated purchase likelihood.	I can access my dashboard with my estimated purchase likelihood.	High	Sprint-2
	Purchase Likelihood	USN-3	As a user, I can see a clear and understandable representation of my purchase likelihood score.	The score is presented in a user-friendly format.	High	Sprint-2
	Feedback and Assistance	USN-4	As a user, I can request assistance or provide feedback through the application.	I can submit assistance requests or feedback, and it is acknowledged.	Low	Sprint-3
	Purchase Recommendatio ns	USN-5	As a user, I can receive car purchase recommendations based on my likelihood score and preferences.	Recommendations are tailored to my purchase likelihood and preferences.	Medium	Sprint-3
	User Profile Management	USN-6	As a user, I can update my profile information, including email, password, and additional details.	Changes to profile information are saved and reflected in my dashboard.	Low	Sprint-3
Customer (Web user)	Registration	-	-	-	-	-
	Login	-	-	-	-	-
	Dashboard	-	-		-	-
Customer Care Executive	Dashboard	USN-7	As a customer care executive, I can access a dashboard to view customer purchase likelihood data.	I can view customer data and purchase likelihood information.	High	Sprint-2
Administrator	Dashboard	USN-8	As an administrator, I can access a dashboard to manage system settings and user accounts.	I can configure system settings and manage user accounts.	High	Sprint-4

5.2 Solution Architecture



6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture

The technical architecture is meticulously crafted to ensure a seamless and effective implementation of the car purchase prediction model. Adopting a microservices approach, the user interface is developed using HTML, CSS, and JavaScript, providing an intuitive and responsive experience for potential buyers. Python, along with its powerful libraries, is employed for the application logic, optimizing the model's performance.

Customer data, a cornerstone of the prediction model, is stored in MS Excel CSV format, chosen for its simplicity and compatibility. This format facilitates easy data manipulation and integration into the machine learning pipeline. The heart of the system lies in the Python Jupyter Notebook, where the ML model is developed and fine-tuned through iterative processes of training and feature engineering.

6.2 Sprint Planning & Estimation

Phase 1: Ideation Phase (5 days)

- Day 1-2: Brainstorming Map
- Day 3-4: Empathy Map
- Day 5: Documentation and Review

Phase 2: Project Design Phase (5 days)

- Day 1-2: Dataflow Diagram
- Day 3-4: Proposed Solution and Solution Architecture Diagram
- Day 5: Documentation and Review

Phase 3: Project Planning Phase (4 days)

- Day 1: Define Technology Stack
- Day 2-3: Create User Stories
- Day 4: Documentation and Review

Phase 4: Project Development Phase (10 days)

- Day 1-2: Setup Development Environment
- Day 3-6: Model Development and Testing
- Day 7-8: User Interface Integration
- Day 9-10: Documentation and Review

Phase 5: Performance and Final Submission Phase (3 days)

- Day 1: Solution Performance Evaluation
- Day 2: Finalize Project Documentation
- Day 3: Submission and Review

6.3 Sprint Delivery Schedule

The phased approach to project delivery ensured a well-orchestrated progression from ideation to submission. In the Ideation Phase, a dynamic brainstorming session spanned the first two days, followed by the creation of an Empathy Map, concluding with comprehensive documentation on day 5. The subsequent Project Design Phase seamlessly transitioned from a detailed Dataflow Diagram to a Proposed Solution and Solution Architecture Diagram within the allocated 5 days. In the Project Planning Phase, a robust technology stack was defined on day 1, followed by two days dedicated to user stories, culminating in comprehensive documentation on day 4. The 10-day Project Development Phase intricately weaved coding, testing, and integration tasks, with the final two days reserved for documentation and a thorough review. The Performance and Final Submission Phase evaluated the solution on day 1, finalized documentation on day 2, and concluded with submission and review on day 3. The collaborative efforts of the 4-member team ensured the successful and timely completion of each sprint, paving the way for a cohesive and impactful project outcome.

7. CODING & SOLUTION

7.1 Features:

- Age: Age of the individuals inquiring about the car.
- Annual Salary: Annual salary of the individuals.
- Gender: Gender of the individuals (encoded as 0 for male and 1 for female).

7.2 Machine Learning Model:

- Algorithm Used: KMeans Clustering
- Number of Clusters (K):
- The optimal number of clusters can be determined using the Elbow Method or other similar techniques. In the provided code, the Elbow Method was used to find the optimal number of clusters.
- Training Data: Scaled and preprocessed data including Age, Annual Salary, and Gender.
- Training Method:
- Model File: The trained KMeans model was saved as 'model.pkl' using joblib for future use.

7.3 Data Exploration and Analysis:

- Univariate Analysis: Explored individual features like Age, Annual Salary, and Gender to understand their distributions and characteristics.
- Bivariate Analysis: Explored relationships between features, such as the correlation between Age and Annual Salary, and visualized these relationships using various plots.
- Outlier Detection: Checked for outliers in Age and Annual Salary columns using box plots.
- Data Visualization: Utilized various plots like bar plots, scatter plots, and heatmaps to visualize the relationships between different features and the target variable (Purchased).

7.4 Machine Learning Algorithms:

- Logistic Regression: Applied logistic regression for classification, although the data seemed not linearly separable.
- Naive Bayes: Used Gaussian Naive Bayes classifier for predicting car purchases based on Age, Annual Salary, and Gender.
- Support Vector Machine (SVM): Applied SVM with a linear kernel for classification.
- Decision Tree: Utilized a decision tree classifier with entropy criterion for making predictions.
- K-Nearest Neighbors (KNN): Used KNN algorithm with k=3 and Minkowski distance metric for classification.

  The KNN model was chosen for web application development as it gave the best possible results.

7.5 Model Evaluation:

- Accuracy: Evaluated the accuracy of the models using the test dataset.
- Confusion Matrix: Computed confusion matrices for the KNN classifier to evaluate true positive, true negative, false positive, and false negative predictions.

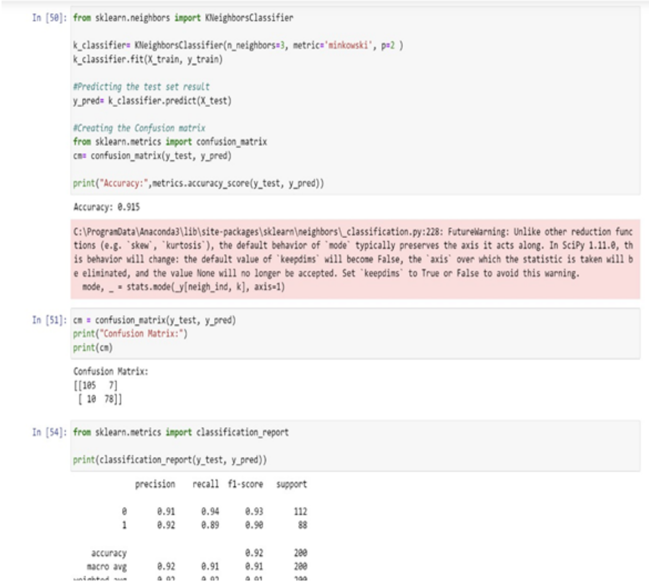
8. PERFORMANCE TESTING

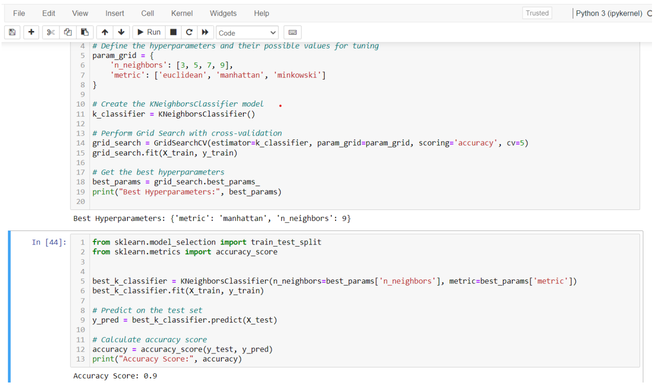
8.1 Performance Metrics

S.No.	**Parameter**	**Values**	**Screenshot**

<table>
<tr>
<td>1</td>
<td>Metrics</td>
<td>Classification Model:

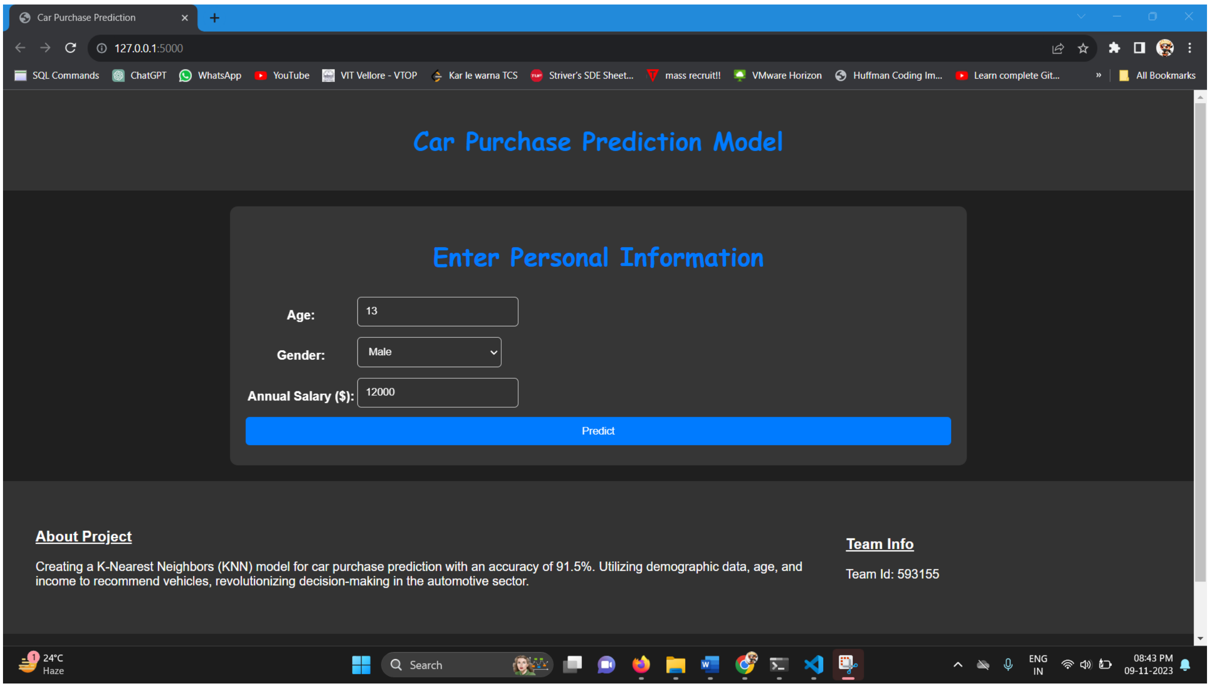
Confusion Matrix [[105,7],[10,78]]

Accuracy Score-0.915</td>
<td>
</td>
</tr>
<tr>
<td>2</td>
<td>Tune the Model</td>
<td>Hyperparameter Tuning: GridSearchCV, RandomizedSearchCV

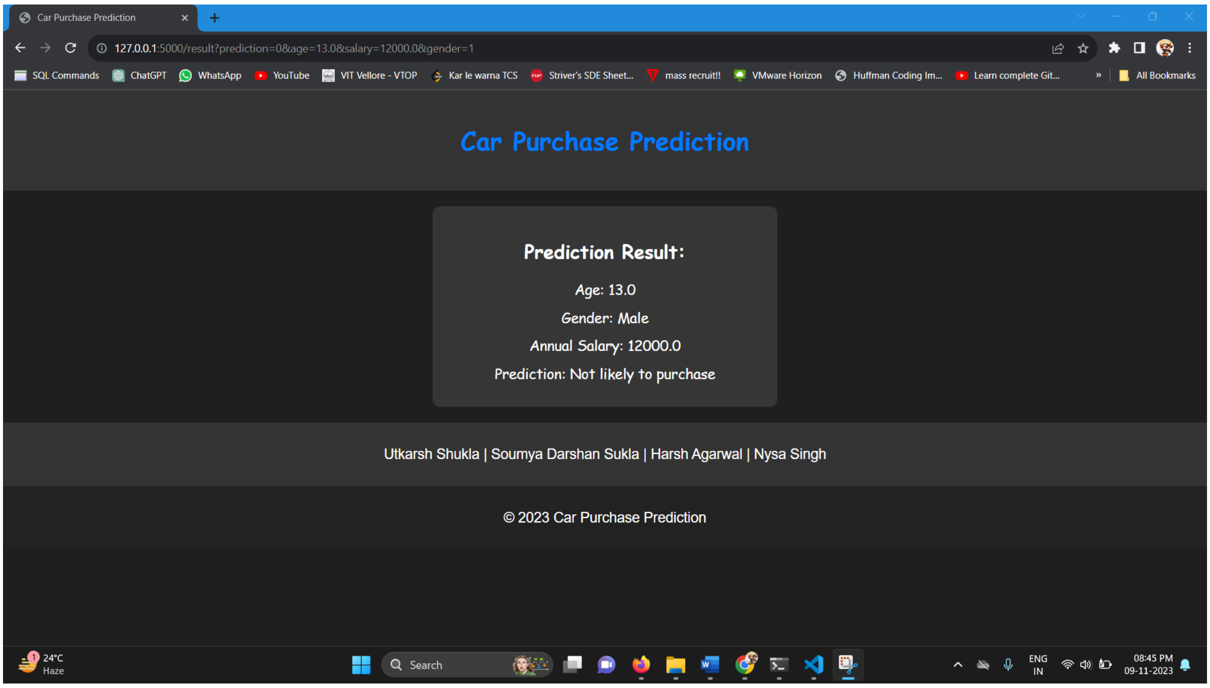
Validation Method: Cross-validation</td>
<td>
</td>
</tr>
</table>

9. RESULTS

9.1 User Info Form



9.2 Result Screen



10. ADVANTAGES & DISADVANTAGES

10.1 Advantages:

- Informed Decision Making:
    - Predictive models empower businesses to make informed decisions, optimizing marketing strategies and inventory management based on customer preferences and behavior patterns.

- Personalized Customer Experience:
  - By predicting customer preferences, businesses can offer personalized recommendations and services, enhancing customer engagement and loyalty.
- Increased Sales Efficiency:
  - Sales teams can focus their efforts on leads more likely to convert, leading to increased efficiency and higher conversion rates.
- Competitive Edge:
  - Companies leveraging predictive analytics gain a competitive advantage by understanding market trends and customer demands, allowing them to stay ahead of competitors.
- Data-Driven Insights:
  - Predictive modeling provides valuable insights into customer behavior, enabling businesses to anticipate market trends and adapt their strategies accordingly.

5.2 Disadvantages:

- Data Dependency:
  - Predictive models heavily rely on the quality and quantity of data available. Inadequate or biased data can lead to inaccurate predictions.
- Overfitting:
  - Models may become overly complex, capturing noise in the training data rather than genuine patterns. Overfitting can reduce the model's effectiveness on new data.
- Ethical Concerns:
  - Predictive models, if not carefully developed, can perpetuate biases present in the training data, leading to ethical concerns related to discrimination and fairness.
- Cost and Resource Intensiveness:
  - Developing accurate predictive models often requires substantial computational resources and skilled data scientists, leading to high development and maintenance costs.
- Changing Market Dynamics:
  - Predictive models may become less effective if market dynamics change significantly, requiring continuous monitoring and adaptation to maintain accuracy.
- Privacy Issues:
  - The collection and analysis of customer data raise privacy concerns. Ensuring compliance with data protection regulations is crucial to avoid legal issues.

11. CONCLUSION

In conclusion, "Car Purchase Prediction using Machine Learning" represents a significant stride towards transforming the automotive industry. By accurately anticipating customer behaviour, businesses can forge meaningful connections with their clientele, resulting in mutually beneficial outcomes. As we move forward, the insights gained from this project serve as a foundation for continued research and innovation, paving the way for a future where businesses can navigate the complex landscape of consumer choices with confidence and precision.

12. FUTURE SCOPE

1. User Experience Optimization:

Continuously gather user feedback and conduct usability testing to enhance the user interface and overall user experience of the web application.

Implement interactive and dynamic features such as filters, sorting options, and personalized dashboards to engage users effectively.

2. Predictive Model Improvement:

Explore advanced machine learning techniques and algorithms to improve the accuracy and reliability of the car purchase prediction model.

Investigate ensemble methods, neural networks, or deep learning architectures to capture complex patterns in the data.

3. Real-Time Data Integration:

Integrate real-time data sources, such as market trends, customer reviews, or social media sentiments, to enhance the accuracy of predictions and provide users with up-to-date information.

4. Personalization and Recommendations:

Implement personalized recommendation features based on user preferences, browsing history, and demographic information. Utilize techniques like collaborative filtering or content-based filtering for personalized suggestions.

5. Mobile Application Development:

Extend the functionality of your project by developing a mobile application for both Android and iOS platforms. Mobile apps can reach a broader user base and provide on-the-go access to car purchase predictions.

6. Integration with External Services:

Integrate the web application with external services such as online car marketplaces, dealerships, or financial institutions to offer additional services like loan calculations, insurance quotes, or test drive scheduling.

7. Enhanced Analytics and Reporting:

Implement advanced analytics features such as data visualization tools, trend analysis, and detailed reporting functionalities. Users can benefit from insightful data representations and make more informed decisions.

8. AI Chatbot Integration:

Integrate an AI-powered chatbot within the web application to provide instant customer support, answer user queries, and assist users in exploring car options based on their preferences.

9. Feedback Mechanism:

Implement a feedback mechanism within the application to collect user feedback on predicted car options. Analyze this feedback to improve the accuracy of predictions and enhance user satisfaction.

10. Accessibility and Compliance:

Ensure that the web application adheres to accessibility standards, making it usable for individuals with disabilities. Compliance with accessibility guidelines ensures a wider user base.

11. Monetization Strategies:

Explore various monetization strategies such as premium features, subscription plans, or partnerships with dealerships and manufacturers. Monetization can support the sustainability of the project.

12. Continuous Monitoring and Updates:

Regularly monitor the application's performance, user engagement, and predictive accuracy. Stay updated with the latest technologies and industry trends to incorporate relevant features and improvements.

13. APPENDIX

13.1. Source Code:
https://github.com/smartinternz02/SI-GuidedProject-600923-1697626304/tree/main/Phase4-%20Project%20Development%20Phase

13.2. GitHub & Project Demo Link:
https://github.com/smartinternz02/SI-GuidedProject-600923-1697626304