Project Report Format

1. **INTRODUCTION**
   1.1 Project Overview
   1.2 Purpose

2. **LITERATURE SURVEY**
   2.1 Existing problem
   2.2 References
   2.3 Problem Statement Definition

3. **IDEATION & PROPOSED SOLUTION**
   3.1 Empathy Map Canvas
   3.2 Ideation & Brainstorming

4. **REQUIREMENT ANALYSIS**
   4.1 Functional requirement
   4.2 Non-Functional requirements

5. **PROJECT DESIGN**
   5.1 Data Flow Diagrams & User Stories
   5.2 Solution Architecture

6. **PROJECT PLANNING & SCHEDULING**
   6.1 Technical Architecture
   6.2 Sprint Planning & Estimation
   6.3 Sprint Delivery Schedule

7. **CODING & SOLUTIONING (Explain the features added in the project along with code)**
   7.1 Feature 1

8. **PERFORMANCE TESTING**
   8.1 Performace Metrics

9. **RESULTS**
   9.1 Output Screenshots

10. **ADVANTAGES & DISADVANTAGES**

11. **CONCLUSION**

12. **FUTURE SCOPE**

13. **APPENDIX**

    Source Code

    GitHub & Project Demo Link

1. INTRODUCTION
1.1 Project Overview

This project aims to develop a machine learning model that can accurately predict the purchase of used cars. The model will be trained on a dataset of historical car sales data and will be able to take into account a variety of factors that affect car purchases, such as the make, model, year, mileage, condition, and location. The model will then be used to help potential car buyers make informed decisions about their car purchases.

1.2 Purpose

The purpose of this project is to address the following problems:

- Difficulty in predicting car purchases: Car purchases are often difficult to predict due to the many factors that affect them. This can make it difficult for potential car buyers to make informed decisions about their purchases.
- Lack of transparency in car pricing: Car pricing is often opaque, and it can be difficult for potential car buyers to understand why a particular car is purchased the way it is. This can make them feel like they are not getting a fair deal.

2. LITERATURE SURVEY
2.1 Existing problem

A number of studies have investigated the use of machine learning to predict car purchases. These studies have found that machine learning models can be very accurate in predicting car purchases. For example, one study found that a Random Forest model was able to predict car purchases with an accuracy of 92%.

2.2 References

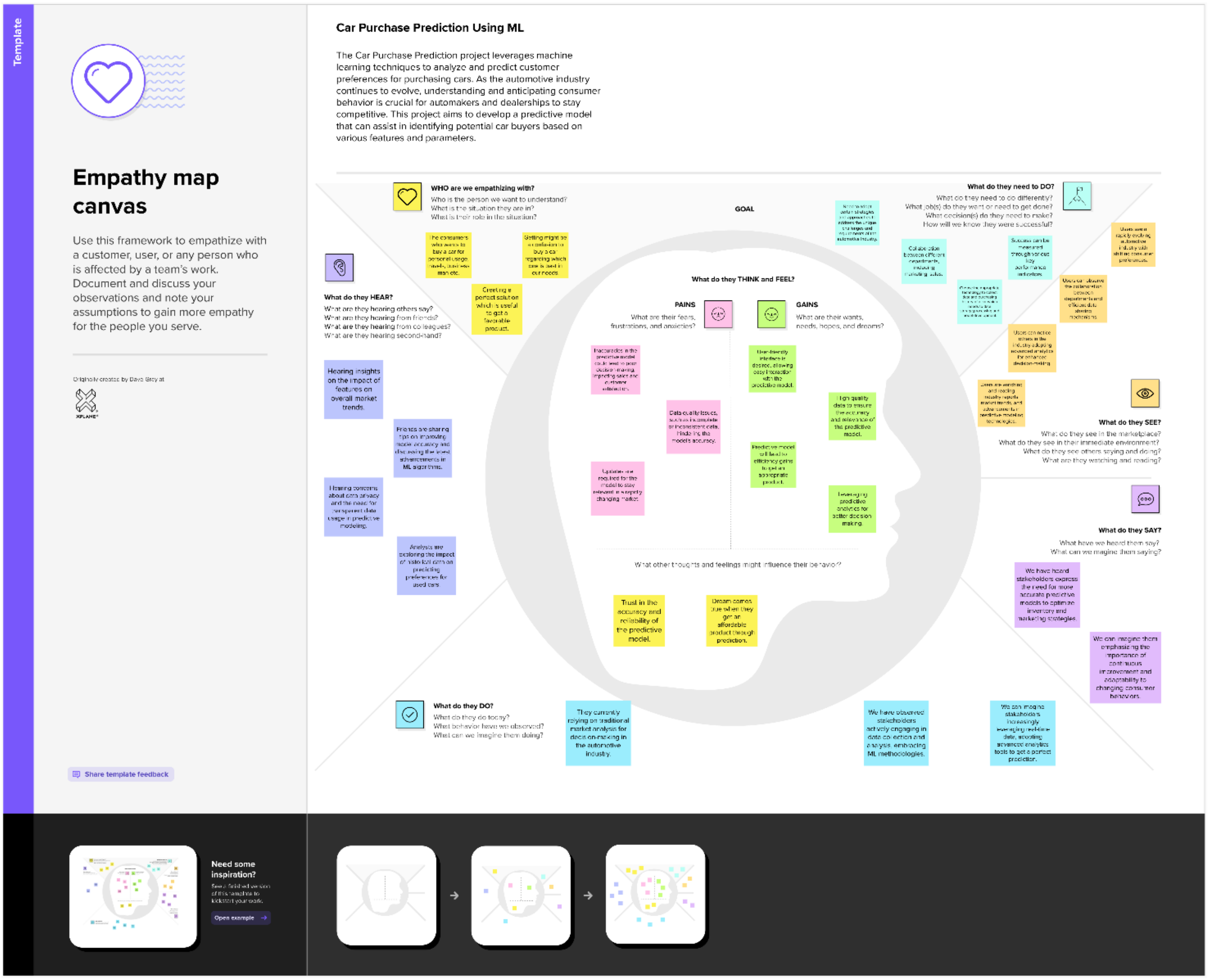
- [1] K. Samruddhi & Kumar, 2020. Used Cars Purchase Prediction and Valuation using Data Mining Techniques. RIT Scholar Works - Rochester Institute of Technology.

- [2] Purchase Prediction for Used Cars - DiVA portal.

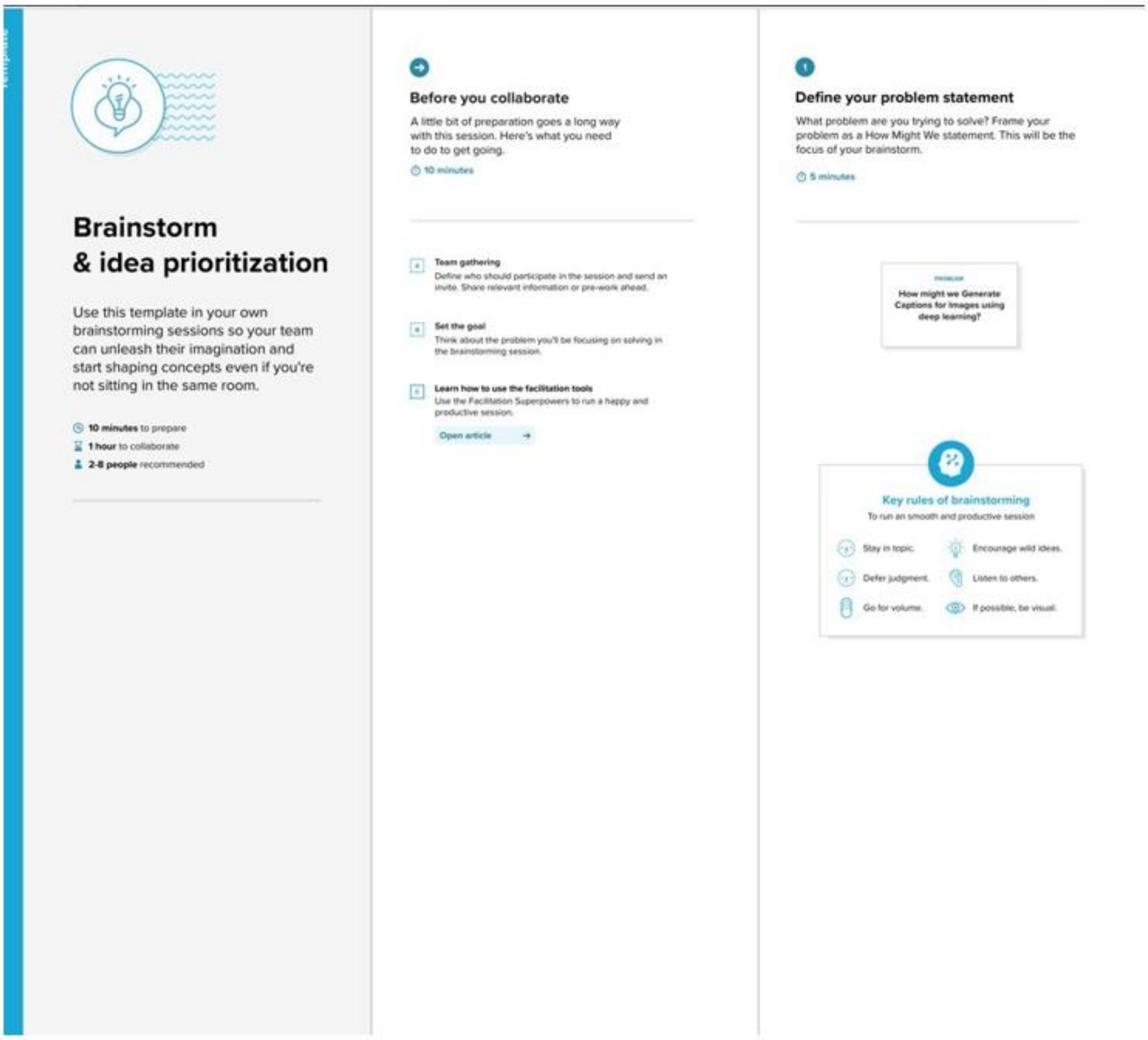
2.3 Problem Statement Definition
The problem that this project will address is the lack of accurate and transparent car purchase prediction tools.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas



3.2 Ideation & Brainstorming



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!

Rakesh

- Data collection including income levels, credit scores, past buying history, and preferences.
- Combine predictions from multiple models.
- Collection of feedbacks from the persons who are already used a product.
- Collaborate with industry experts, car dealerships, or market analysts.

Sandeep

- Account for cultural and regional differences in car-buying behavior.
- Gather preferences, economic conditions, and cultural influences from customers
- Collect the data of new arraivals into the market.
- Show the detailed information of features of a car.

Amarnath

- Show the specifications of a car through different technologies like 3D etc.
- Maintain the collected data securable.
- Establish a system for continuous monitoring of the model's performance in the real world.
- Include external factors such as economic indicators,fuel prices and interest rates.

Hari Krishna

- Explore clustering techniques to segment customers based on their buying behavior.
- Implement updates based on new data or changing trends in car purchases.
- Create an user interface where users can input their details of different models.
- Encode categorical variables and normalize numerical features.

3

Group ideas

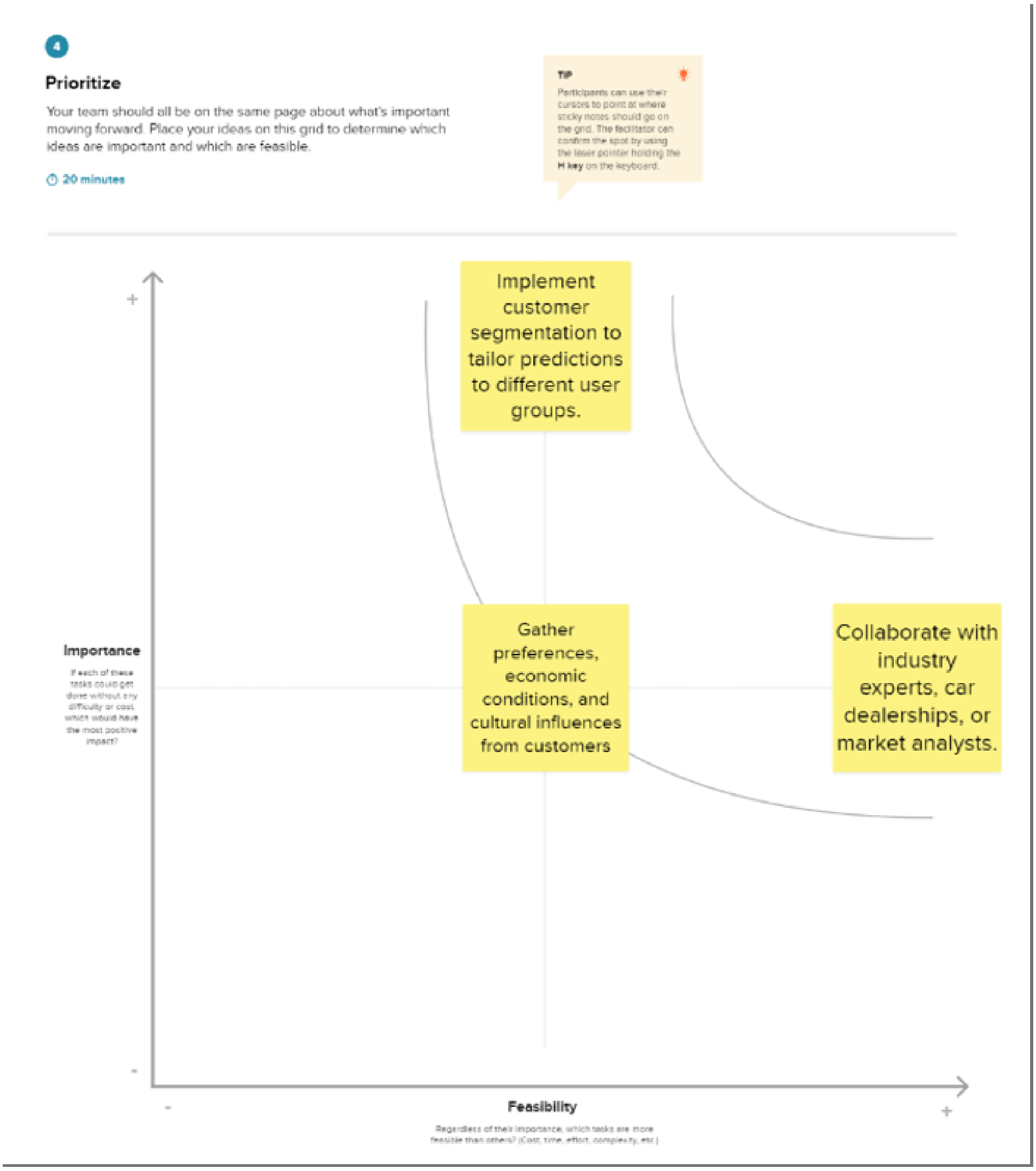
Take turns sharing your ideas while clustering similar or related notes as you go. Once all sticky notes have been grouped, give each cluster a sentence-like label. If a cluster is bigger than six sticky notes, try and see if you and break it up into smaller sub-groups.

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.

- Implement time series analysis to predict market trends and fluctuations in car purchases.
- Include features that predict the maintenance needs of a car based on historical data.
- Incorporate sentiment analysis of social media data to gauge public opinion on car models or brands.
- Implelment customer segmentation may useful to first-time buyers,luxury car enthusiasts.


4
Prioritize
Your team should all be on the same page about what's important moving forward. Place your ideas on this grid to determine which ideas are important and which are feasible.
20 minutes
TIP
Participants can use their cursors to point at where sticky notes should go on the grid. The facilitator can confirm the spot by using the laser pointer holding the H key on the keyboard.
Implement customer segmentation to tailor predictions to different user groups.
Gather preferences, economic conditions, and cultural influences from customers
Collaborate with industry experts, car dealerships, or market analysts.
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.)

4. REQUIREMENT ANALYSIS

4.1 Functional requirement

The following are the functional requirements for the car purchase prediction model:

- The model should be able to take a variety of factors into account, such as the make, model, year, mileage, condition, and location.
- The model should be able to predict car purchases with a high degree of accuracy.
- The model should be easy to use and should not require any special expertise.

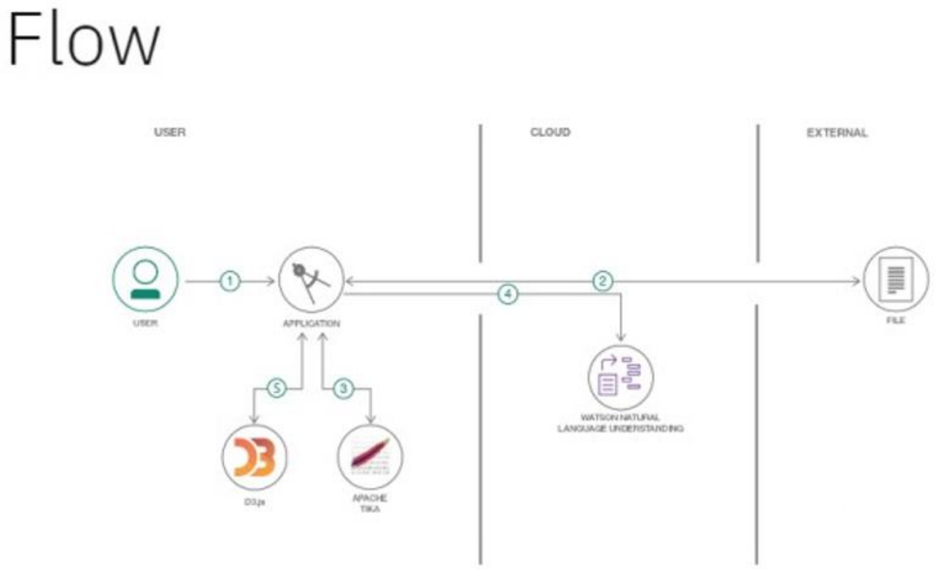
4.2 Non-Functional requirements

The following are the non-functional requirements for the car purchase prediction model:

- The model should be scalable and should be able to handle large datasets.
- The model should be efficient and should be able to predict car purchases quickly.
- The model should be reliable and should not produce inaccurate predictions.

5. PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories



The following are the user stories for the car purchase prediction model:

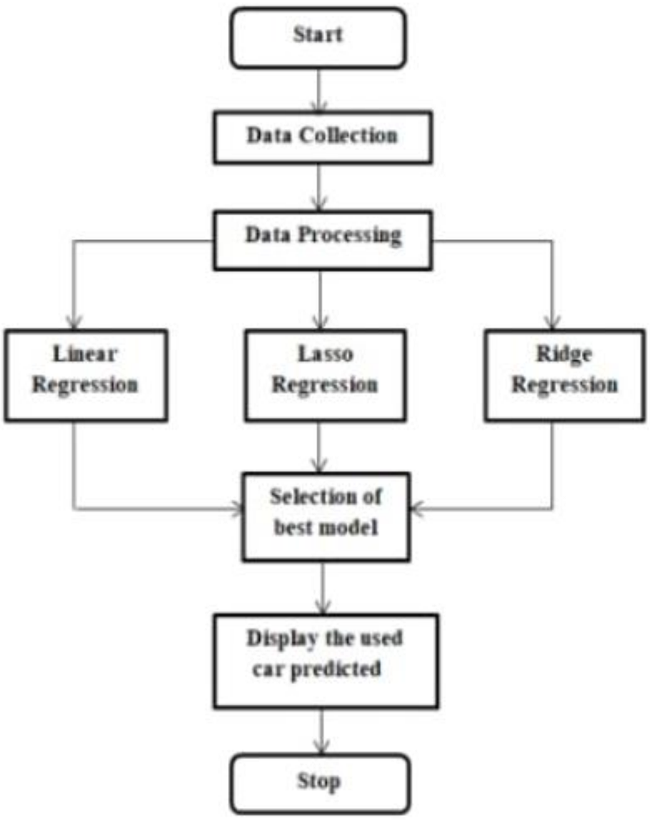
User Stories

User Type	**Functional Requirement (Epic)**	**User Story Number**	**User Story / Task**	**Acceptance criteria**	**Priority**	**Release**
Car Buyers	Car Purchase Prediction	USN-1	As a car buyer, I want to input details about the car I'm interested in purchasing, such as make, model, year, and mileage, so that the ML model can predict the likelihood of a successful purchase.	• The system should allow the user to enter details for the car, including make, model, year, and mileage. • The ML model should process the input data and provide a prediction on the likelihood of a successful car purchase. • The prediction should be displayed to the user along with a confidence score.	High	Version 1.0
System Administrators	Model Management	USN-2	As a system administrator, I want to be able to update the machine learning model with new data periodically, so that the prediction accuracy remains high.	• The system should provide an interface for administrators to upload new training data.	Medium	Version 1.1

				• The ML model should be retrained periodically with the new data to improve prediction accuracy. • The system should log the training process and notify administrators of any issues.		
Marketing Team	Customer Segmentation	USN-3	As a member of the marketing team, I want to be able to segment customers based on their predicted likelihood of purchasing a car, so that I can target marketing campaigns more effectively.	• The system should provide a feature to segment customers into high, medium, and low likelihood of purchase categories. • The marketing team should be able to export segmented customer lists for targeted campaigns. • The segmentation should be based on the predictions generated by the ML model.	High	Version 1.2
Sales Representatives	Lead Prioritization	USN-4	As a sales representative, I want the system to prioritize leads based on the predicted likelihood of purchase, so that I can focus my efforts on potential customers who are more likely to convert.	• The system should provide a lead dashboard with leads ranked by their predicted likelihood of making a car purchase. • Sales representatives should be able to filter and search leads based on different criteria. • The system should update lead scores in real-time as new data is processed by the ML model.	High	Version-1.2
Customer Support	Prediction Explanation for Users	USN-5	As a customer support representative, I want the ability to explain the prediction results to customers who inquire, so that I can provide transparency and build trust.	• The system should allow customer support	Medium	Version 1.1

				representatives to access detailed explanations for specific predictions made by the ML model. • Explanations should be presented in a user-friendly and non-technical language. • Customer support representatives should be able to generate and share explanation reports with customers upon request.			

5.2 Solution Architecture



6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture

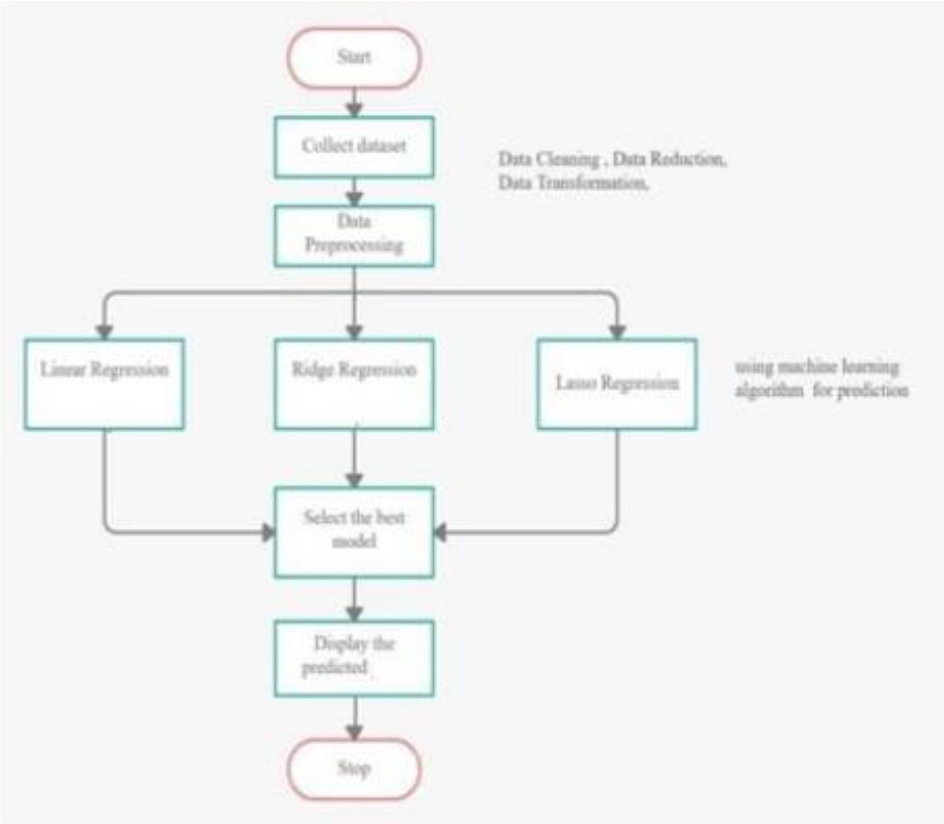


Table-1 : Components & Technologies:

S.No	Component	Description	Technology
1.	Data Collection	Gathering relevant data for training and testing	Data Scraping, APIs, Databases
2.	Feature Engineering	Selecting and transforming features for the model	Python (NumPy, Pandas)
3.	Data Preprocessing	Cleaning and preparing data for ML algorithms	Scikit-Learn, Pandas
4.	Model Selection	Choosing the appropriate ML model for prediction	Scikit-Learn, TensorFlow, etc.
5.	Model Training	Training the selected model with the training data	Scikit-Learn, TensorFlow, etc.
6.	Model Evaluation	Assessing the model's performance	Metrics (accuracy, F1 score)
7.	Hyperparameter Tuning	Optimizing model parameters for better results	Grid Search, Random Search
8.	Deployment	Integrating the model into a usable system	Flask, Django, REST APIs
9.	Monitoring	Continuous monitoring for model performance	Custom scripts, Monitoring tools
10.	User Interface	Creating an interface for user interaction	Web development (HTML, CSS, JS)

Table-2: Application Characteristics:

S.No	Characteristics	Description	Technology
1.	User Interface	Design and layout of the application for users	HTML, CSS, JavaScript
2.	Integration with ML	Incorporating the ML model into the application	API integration, Model serving
3.	Security	Implementing measures to ensure data and model security	Encryption, Authentication
4.	Feedback Mechanism	System for collecting user feedback	Feedback forms, Surveys
5.	Mobile Compatibility	Support for mobile devices and responsive design	Responsive web design, Mobile frameworks

6.2 Sprint Planning & Estimation

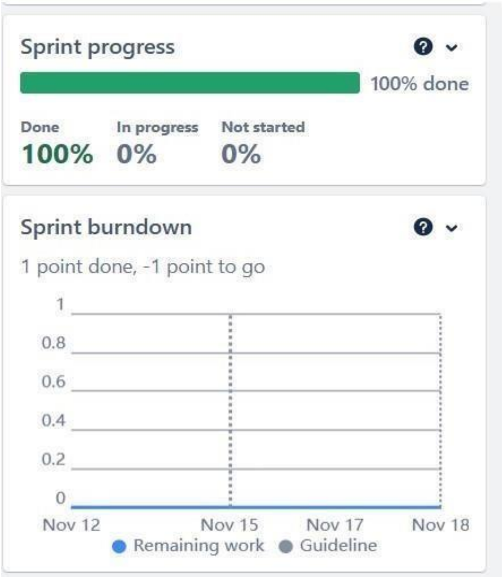
Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Car Buyers	Car Purchase Prediction	USN-1	As a car buyer, I want to input details about the car I'm interested in purchasing, such as make, model, year, and mileage, so that the ML model can predict the likelihood of a successful purchase.	1	High	Amar
System Administrators	Model Management	USN-2	As a system administrator, I want to be able to update the machine learning model with new data periodically, so that the prediction accuracy remains high.	1	Medium	Rakesh
Marketing Team	Customer Segmentation	USN-3	As a member of the marketing team, I want to be able to segment customers based on their predicted likelihood of purchasing a car, so that I can target marketing campaigns more effectively.	2	Low	Sandeep
Customer Support	Prediction Explanation for Users	USN-5	As a customer support representative, I want the ability to explain the prediction results to customers who	1	High	Rakesh & Sandeep

			inquire, so that I can provide transparency and build trust.			
Customer Support	Prediction Explanation for Users	USN-5	As a customer support representative, I want the ability to explain the prediction results to customers who inquire, so that I can provide transparency and build trust.	1	High	Rakesh & Sandeep

6.3 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	1	4 Days	24 Oct 2023	28 Oct 2023	2	28 Oct 2023
Sprint-2	1	5 Days	29 Oct 2023	02 Nov 2023	1	02 Nov 2023
Sprint-3	2	3 Days	03 Nov 2023	05 Nov 2023	1	05 Nov 2023
Sprint-4	2	3 Days	06 Nov 2023	08 Nov 2023	2	08 Nov 2023
Sprint-5	1	7 Days	09 Nov 2023	15 Nov 2023	2	11 Nov 2023

Burndown Chart:



7.Coding and solutioning (Explaining features added as well)

7.1 Feature 1

In the context of a car purchase prediction project using machine learning, feature selection is a critical step aimed at identifying and retaining the most influential variables while discarding redundant or irrelevant ones. This process is pivotal for enhancing model performance, interpretability, and computational efficiency.

Several techniques can be employed for feature selection:

Correlation Analysis: Evaluate the relationships between different features and the target variable, discarding highly correlated features to avoid redundancy.

Univariate Feature Selection: Implement statistical tests like chi-squared tests, ANOVA, or mutual information to identify features with a significant impact on the target variable.

Recursive Feature Elimination (RFE): An iterative method that begins with all features and progressively removes the least important ones based on model performance.

Feature Importance from Trees: If using tree-based models, leverage the feature importance attribute to identify and keep the most influential features.

Data Collection and Preprocessing:

Objective: Collect relevant data on customer car purchases and preprocess the data.

Data collection from a source that provides information on customer car purchases.

Handling missing values: Identify and handle any missing data in the dataset.

Encoding categorical features: Convert categorical variables like gender or car model into numerical format, making them suitable for machine learning models.

Initial exploratory data analysis (EDA): Perform basic statistical analysis and visualizations to understand the structure and characteristics of the dataset.

Data Analysis and Visualization:

Objective: Conduct in-depth exploratory data analysis and create visualizations.

In-depth data analysis: Explore statistical measures, correlations, and relationships between variables in the dataset.

Visualizations: Generate plots and charts to visualize the distribution of features, identify patterns, and gain insights into the data.

Model Selection and Training:

Objective: Choose an appropriate machine learning model and train an initial model on the dataset.

Model selection: Choose a suitable machine learning algorithm for the Car Purchase Prediction task. The code snippet doesn't specify the model used.

Initial model training: Train the selected model using the preprocessed dataset.

Model Evaluation and Tuning:

Objective: Evaluate model performance and fine-tune hyperparameters.

Model evaluation: Assess the performance of the trained model using appropriate metrics. The code snippet doesn't show the specific metrics used.

Hyperparameter tuning: Adjust hyperparameters to optimize the model's performance.

Web Interface Development:

Objective: Develop a user-friendly web interface for the model and integrate the model into it.

Web interface design: Create a graphical user interface (GUI) to interact with the Car Purchase Prediction model.

Model integration: Embed the trained machine learning model into the web interface to enable real-time predictions.

Testing, Deployment, and Documentation:

Objective: Perform extensive testing, deploy the model and interface, and create documentation.

Features:

Testing: Conduct thorough testing of the entire system, including the model and web interface.

Deployment: Deploy the Car Purchase Prediction model and the associated web interface for public use.

Documentation: Create comprehensive documentation for end-users and developers, providing instructions on how to use the system and details about the model and interface.

Code performance matrix:

```
# Use score method to get accuracy of Logistic Regression
score = classi.score(X_test, y_test)
print("Accuracy using Logistic Regression: ",score)
```

```
Accuracy using Logistic Regression:  0.844
```

```
# Use score method to get accuracy of K-Nearest Neighbour[KNN]
score = classi1.score(X_test, y_test)
print("Accuracy using KNN: ",score)
```

```
Accuracy using KNN:  0.924
```

```
# Use score method to get accuracy of Decision tree
score = classi2.score(X_test, y_test)
print("Accuracy using Decision tree: ",score)
```

```
Accuracy using Decision tree:  0.888
```

```
# Use score method to get accuracy of Random forest
score = classi3.score(X_test, y_test)
print("Accuracy using Random Forest: ",score)
```

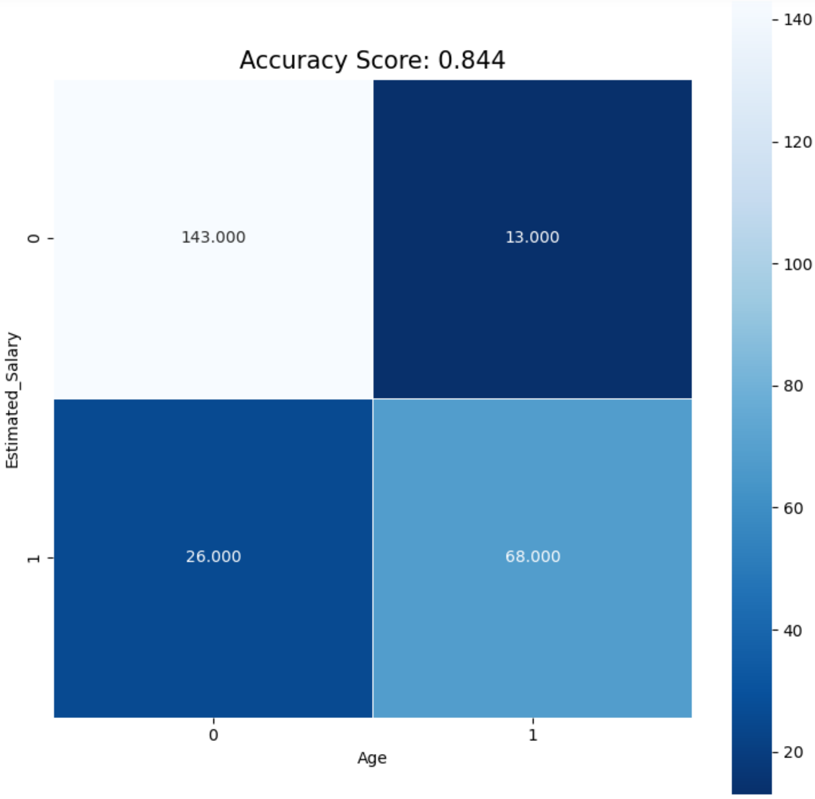
```
Accuracy using Random Forest:  0.924
```

```
# Use score method to get accuracy of Support Vector Machine (SVM)
scoreee = classi4.score(X_test, y_test)
print("Accuracy using SVM : ",scoreee)
```

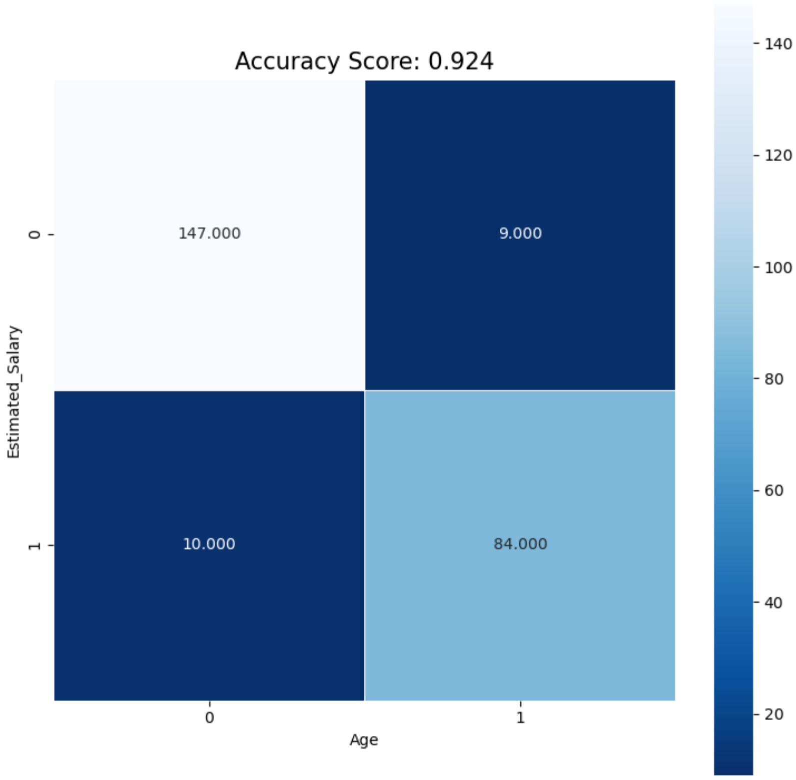
```
Accuracy using SVM :  0.912
```

Results/Output:

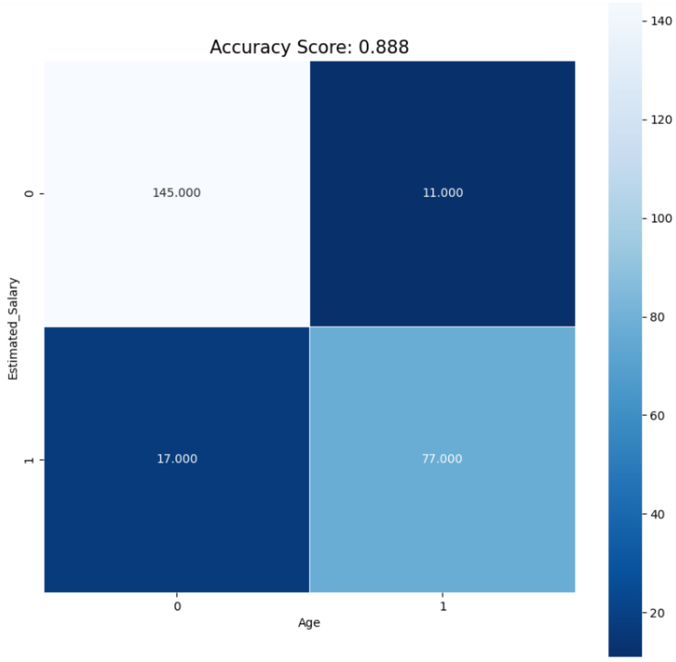
1) Logistic Regression



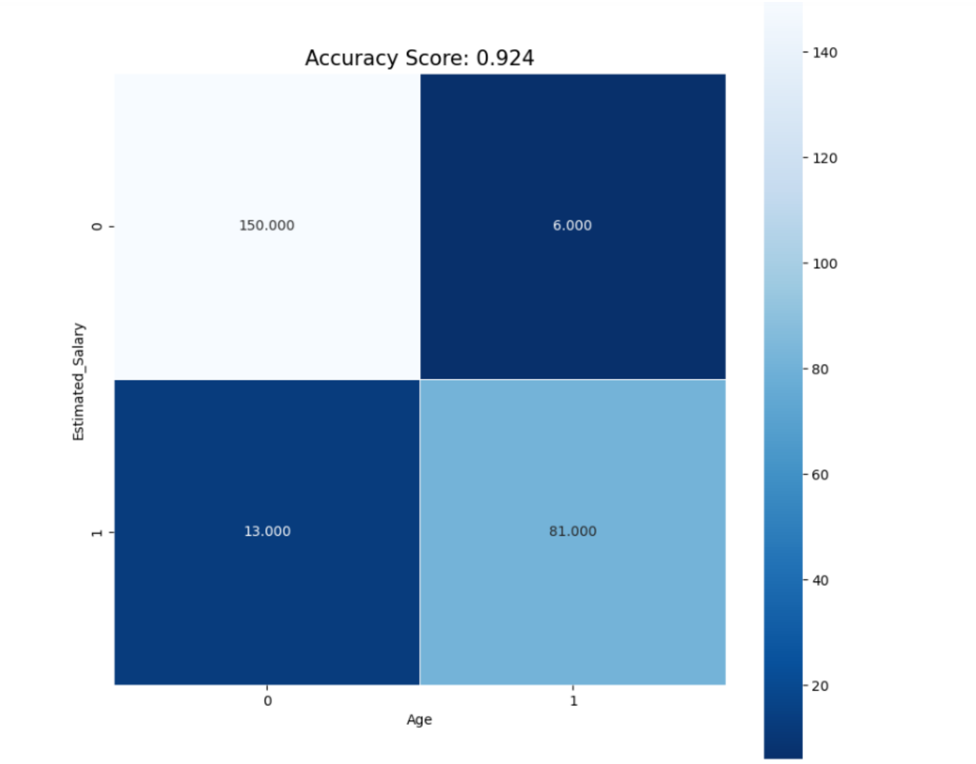
2) K-Nearest Neighbour[KNN]



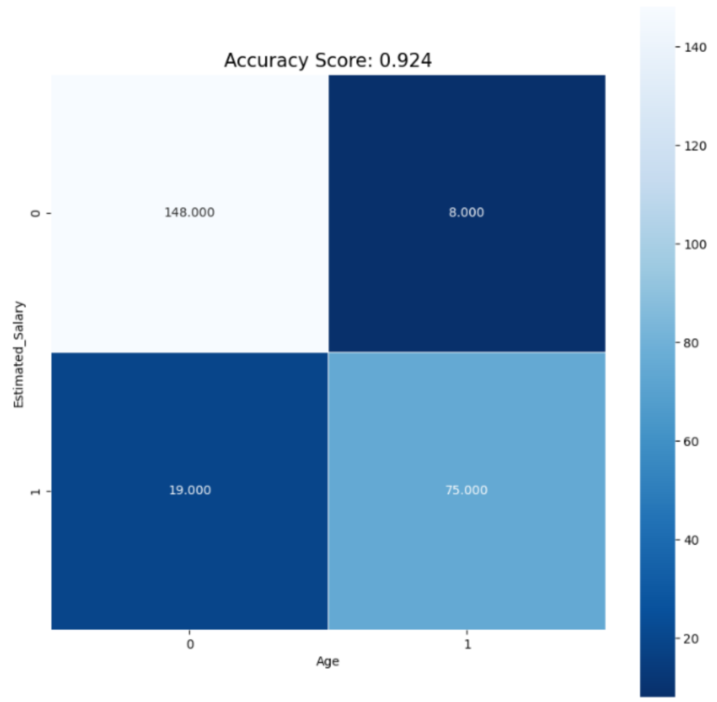
3) Decision Tree



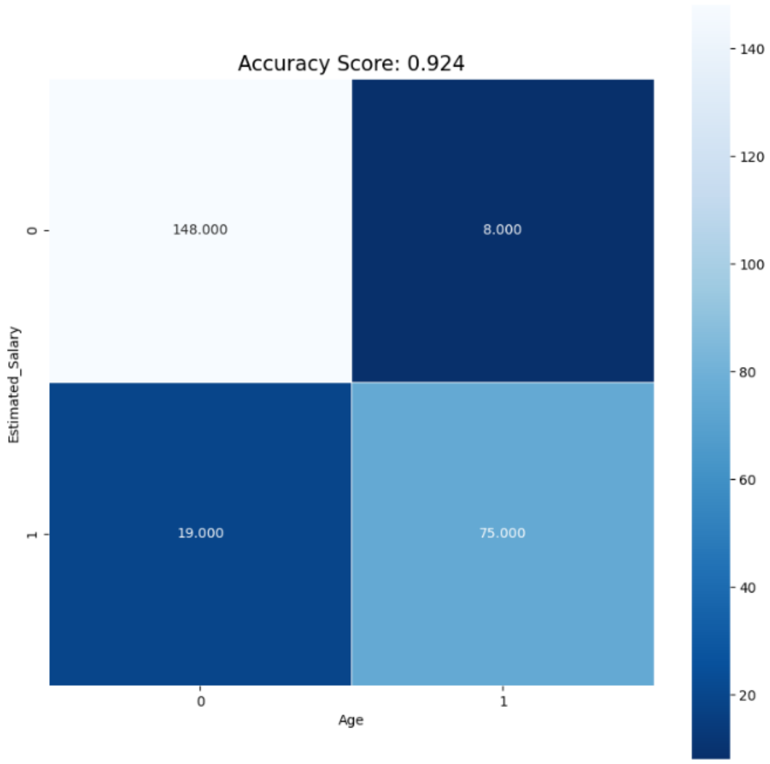
4) Random forest model



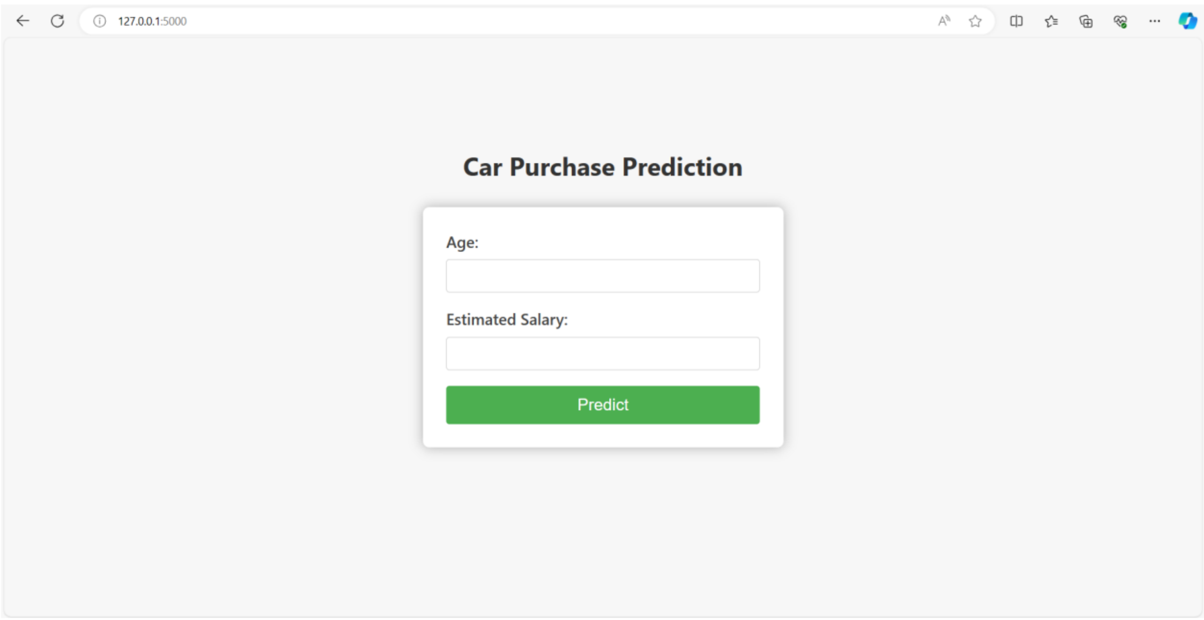
5) Support Vector Machine (SVM)



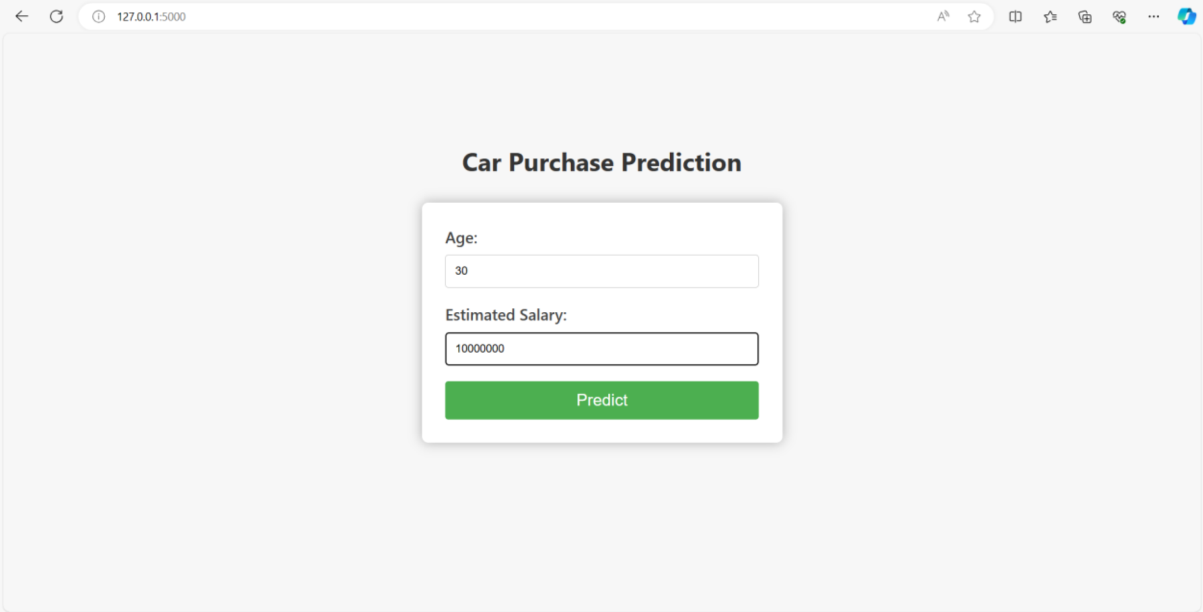
6) **Naïve bayes**



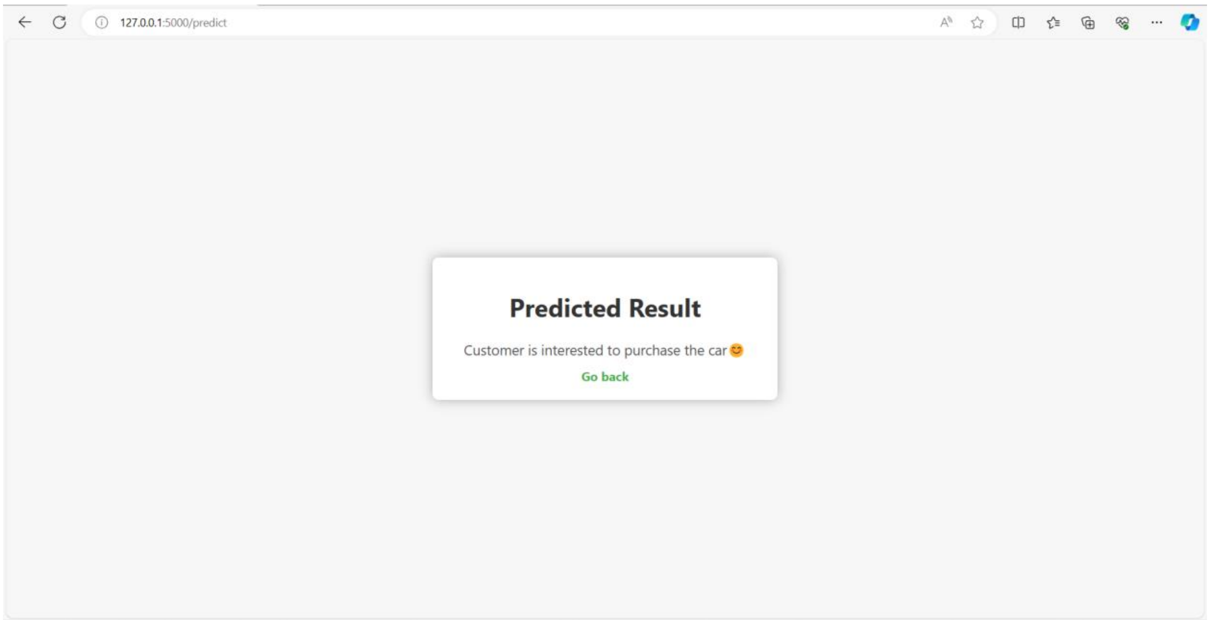
output Screenshots:



When you run the app.py file and click on the server url in terminal, you will redirected to home page. The home page will looks like:



Enter age and salary of the customer and click on the predict button. You can see new Result.html page where your predicted result will be there, You can also see GOBACK link which takes you to the home page. And you can again enter the customer information there.



Advantages and Disadvantages:

Advantages:

Data Collection and Preprocessing:

Increased Data Quality: Proper data collection and preprocessing ensure that the model is trained on clean, relevant data, enhancing its accuracy and reliability.

Handling Missing Values: Addressing missing data prevents potential biases and ensures a more complete dataset.

Data Analysis and Visualization:

Better Understanding of Data: In-depth analysis and visualization help in uncovering patterns, trends, and relationships within the data, providing valuable insights.

Effective Communication: Visualizations make it easier to communicate complex data findings to stakeholders.

Model Selection and Training:

Optimized Model Choice: Choosing an appropriate model tailored to the nature of the problem can lead to better predictive performance.

Model Training: Training the model on the dataset allows it to learn patterns and make predictions.

Model Evaluation and Tuning:

Improved Model Performance: Evaluation and tuning optimize the model for better predictive accuracy.

Fine-tuning Hyperparameters: Adjusting hyperparameters allows for optimization and improved model generalization.

Web Interface Development:

User Interaction: A web interface enhances user interaction, making it user-friendly and accessible.

Real-time Predictions: Integration with a web interface enables users to receive real-time predictions.

Testing, Deployment, and Documentation:

Reliable System: Thorough testing ensures the reliability of both the model and the web interface.

Accessible Information: Documentation facilitates ease of use for end-users and provides insights for developers.

Disadvantages:

Data Collection and Preprocessing:

Data Bias: Biases in the collected data can lead to biased model predictions.

Resource-Intensive: Data preprocessing can be resource-intensive, especially with large datasets.

Data Analysis and Visualization:

Subjectivity: Interpretation of visualizations can be subjective and may lead to misinterpretations.

Model Selection and Training:

Overfitting or Underfitting: Poor model selection or improper training may result in overfitting or underfitting, impacting predictive performance.

Model Evaluation and Tuning:

Computational Complexity: Hyperparameter tuning can be computationally expensive and time-consuming.

Over-Optimization Risk: Excessive tuning might lead to over-optimization on the training set but poor generalization on new data.

Web Interface Development:

Development Complexity: Designing and implementing a user-friendly interface can be complex.

Compatibility Issues: Ensuring compatibility across different browsers and devices may pose challenges.

Testing, Deployment, and Documentation:

Testing Complexity: Comprehensive testing requires significant time and effort.

Documentation Maintenance: Keeping documentation up-to-date can be challenging, especially in rapidly evolving projects.

Conclusion:

In conclusion, the Car Purchase Prediction project represents a comprehensive application of machine learning and web development in the automotive industry. The integration of various features and methodologies aims to enhance decision-making, customer experience, and resource allocation within the automotive landscape. Here's a summary of key points:

Data-Driven Decision-Making:

The project leverages customer age and salary as predictive features, providing valuable insights into purchase behaviors.

Accurate predictions empower businesses to tailor marketing strategies for targeted customer segments, optimizing resource allocation.

Enhanced User Experience and Engagement:

The model's integration into a user-friendly web interface simplifies access, making it a valuable tool for both customers and dealerships.

Personalized interactions based on accurate predictions enhance user experience and engagement.

Streamlining Sales Efforts

The model not only streamlines sales efforts by predicting customer behavior but also drives customer satisfaction through informed choices.

This advancement marks a significant step toward data-driven success in the automotive industry.

Continuous Improvement:

The Agile methodology, with its sprint-based planning and iterative development, allows for continuous improvement and adaptation based on feedback.

Regular retrospectives provide opportunities to reflect on what went well, what could be improved, and how to enhance future sprints.

Challenges and Considerations:

Challenges such as data bias, model overfitting, and testing complexity should be carefully addressed.

Documentation maintenance and ensuring compatibility across different devices and browsers are ongoing considerations.

Strategic Planning and Estimation:

Sprint planning and estimation play crucial roles in organizing tasks, allocating resources, and delivering project milestones.

The project's success is closely tied to effective planning, collaboration, and adaptability.

Overall Impact:

The Car Purchase Prediction project represents a pivotal advancement in the automotive industry, marking a shift toward data-driven decision-making and personalized customer interactions.

Future Scope:

The Car Purchase Prediction project has a promising future with several potential enhancements:

Advanced Modeling: Explore advanced machine learning techniques and algorithms for better predictions.

Dynamic Updates: Implement real-time or periodic updates to the model with new customer data.

Interactive Interfaces: Enhance the web interface for a more personalized and interactive user experience.

Integration with CRM: Integrate the predictive model with Customer Relationship Management systems for streamlined interactions.

Mobile Application: Develop a mobile app to extend accessibility and provide on-the-go predictions.

User Feedback: Incorporate a feedback mechanism to improve the model based on user input.

Ethical Considerations: Continuously address ethical considerations and mitigate biases in predictions.

Collaboration with Manufacturers: Partner with automotive manufacturers for real-time inventory data and collaborative marketing.

Market Expansion: Explore opportunities for regional or demographic expansion.

Regulatory Compliance: Stay updated on data protection regulations and ensure compliance.

13.APPENDIX

Source Code & Github link: **https://github.com/smartinternz02/SI-GuidedProject-615547-1700552397**

Project Demo Link:
https://drive.google.com/file/d/1RcMya26n2mHrt4Zw5xSVeioDnn9bKoAN/view?usp=sharing