Project Report (TTeam-593136)

1. INTRODUCTION

1.1 Project Overview

This unique research predicts car purchases using a machine learning model and crucial client data such as age, income, and prior purchase behaviours. The model's high predicted accuracy is effortlessly incorporated into a user-friendly interface thanks to powerful algorithms and rigorous data pretreatment. This interface enables prospective car purchasers to enter their demographic information and receive specific purchase likelihoods, aiding them in their decision-making process. The idea has far-reaching ramifications for the automotive industry, in addition to individual users. It transforms marketing strategy by providing insights for targeted approaches. Dealerships can target specific demographics of customers, maximizing resource allocation and making data-driven decisions. Overall, this ground-breaking application of machine learning improves client experiences, allows for more informed choices, and enables automobile companies to easily modify their approaches based on smart data analysis. The goal of this project is to use machine learning strategically to reinvent the car-buying experience. The initiative tries to estimate car purchase likelihoods with exceptional precision by leveraging client data such as age, income, and prior purchase trends. The generated model is smoothly incorporated into a user-friendly interface, utilizing powerful algorithms and rigorous data pretreatment. This interface is a tool for prospective car purchasers, giving them with accurate forecasts based on their demographics and enabling them to make informed decisions.

1.2 Purpose

The goal of this project is to use machine learning strategically to reinvent the car-buying experience. The initiative tries to estimate car purchase likelihoods with exceptional precision by leveraging client data such as age, income, and prior purchase trends. The generated model is smoothly incorporated into a user-friendly interface, utilizing powerful algorithms and rigorous data pretreatment. This interface is a tool for prospective car purchasers, giving them with accurate forecasts based on their demographics and enabling them to make informed decisions.

2. LITERATURE SURVEY

2.1 Existing problem

The existing difficulty in the automobile business is a lack of accurate and individualized prediction models for car purchases. Traditional methods frequently overlook critical customer data, resulting in less accurate estimates. This mismatch limits the industry's capacity to successfully adjust marketing campaigns and maximize resource allocation.

2.2 References

- o Supervised learning: https://www.javatpoint.com/supervised-machine-learning
- o Unsupervised learning: https://www.javatpoint.com/unsupervised-machine-learning
- o Decision Tree: https://www.analyticsvidhya.com/blog/2022/03/decision-tree-machine-learningusing-p ython/
- o Random Forest: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/
- o Evaluation metrics: https://www.analyticsvidhya.com/blog/2019/08/11-important-model-evaluationerror-metrics/

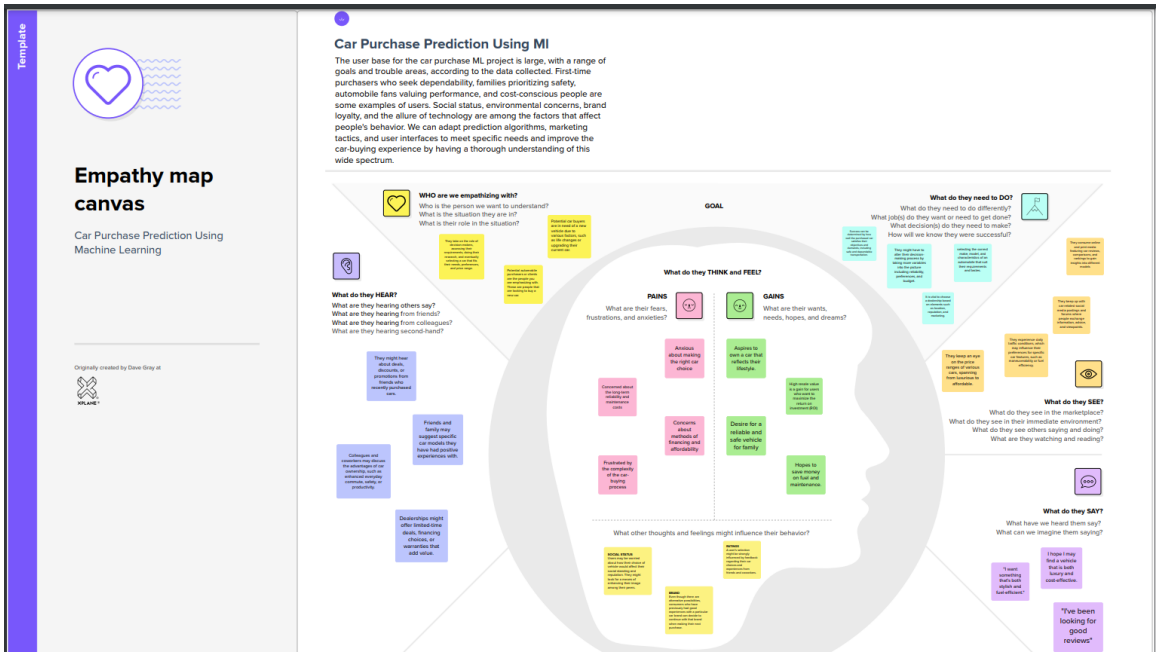
2.3 Problem statement definition

The project's goal is to solve the existing problem by creating an innovative machine learning system that properly forecasts car purchases based on extensive consumer data. Improving predictive accuracy, integrating complex algorithms, and developing a user-friendly interface are all part of the problem statement in order to empower customers and revolutionize marketing strategies in the automotive industry.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

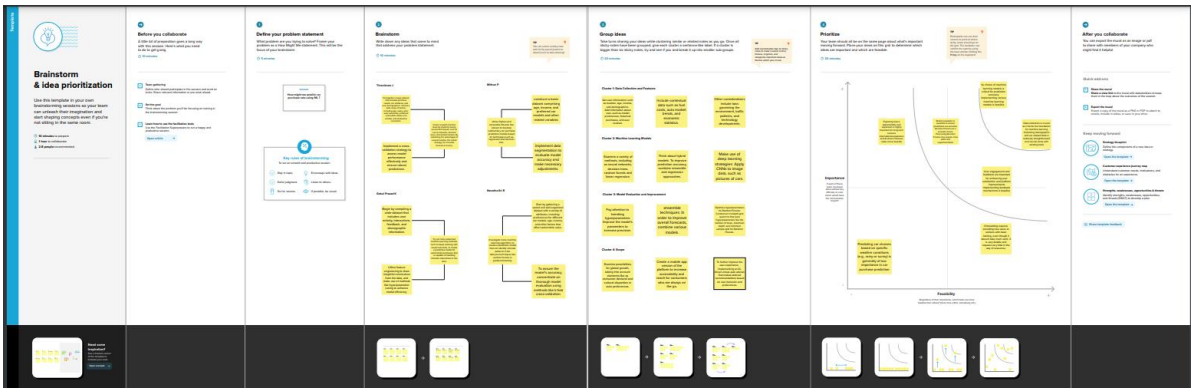
An empathy map is a basic, easy-to-understand picture that captures information about a user's actions and attitudes. It is a valuable tool for assisting teams in better understanding their users. Understanding the true problem and the person experiencing it is necessary for developing an effective solution. The map-making exercise lets participants consider things from the user's point of view, as well as his or her goals and obstacles.



GITHUB LINK
https://github.com/smartinternz02/SI-GuidedProject-601748-1697636519/blob/main/Empathy%20Map.pdf

3.2 Ideation & Brainstorming

Brainstorming creates a free and open environment in which everyone in a team is encouraged to participate in the creative thought process that leads to problem solving.
Prioritizing volume over value, out-of-the-box ideas are welcomed and developed upon, and all participants are encouraged to cooperate, assisting one another in developing a wealth of creative solutions. Use this template in your own brainstorming sessions to let your team's imagination run wild and start developing notions even if you're not all in the same room.



GITHUB LINK

https://github.com/smartinternz02/SI-GuidedProject-601748-1697636519/blob/main/Brainstorm%20and%20prioritize%20ideas%20CANVAS%20.pdf

4. REQUIREMENT ANALYSIS

4.1 Functional Prerequisites:

The project's functional requirements include the creation of a robust machine learning model capable of reliably predicting car purchases. This includes the following:

4.1.1 Data Collection: Collect detailed client information such as age, income, and previous purchasing trends.

4.1.2 Feature Engineering: To improve prediction accuracy, use advanced algorithms for thorough data pretreatment and feature engineering.

4.1.3 Model Integration: Integrate the machine learning model into a user-friendly interface, allowing users to enter demographic information.

4.1.4 Prognosis Output: Based on the input data, provide users with precise purchase likelihoods, allowing for informed decision-making.

4.1.5 Real-time Updates: Ensure that the model can react to changing data patterns and provide predictions in real time.

4.2 Non-Functional requirements

Non-functional requirements concentrate on the project's performance, usability, and scalability:

4.2.1 Accuracy: The model must predict car purchases with a high degree of accuracy.

4.2.2 User Interface: For a seamless user experience, the interface should be intuitive, user-friendly, and visually appealing.

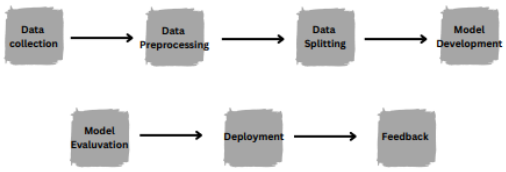
4.2.3 Scalability: The solution must be scalable in order to accept an increasing number of user data and respond to industry developments.

4.2.4 Security: Put in place strong security measures to safeguard sensitive consumer data and preserve privacy.

4.2.5 Performance: The system should have low latency, allowing users to make quick and accurate predictions.

5. PROJECT DESIGN

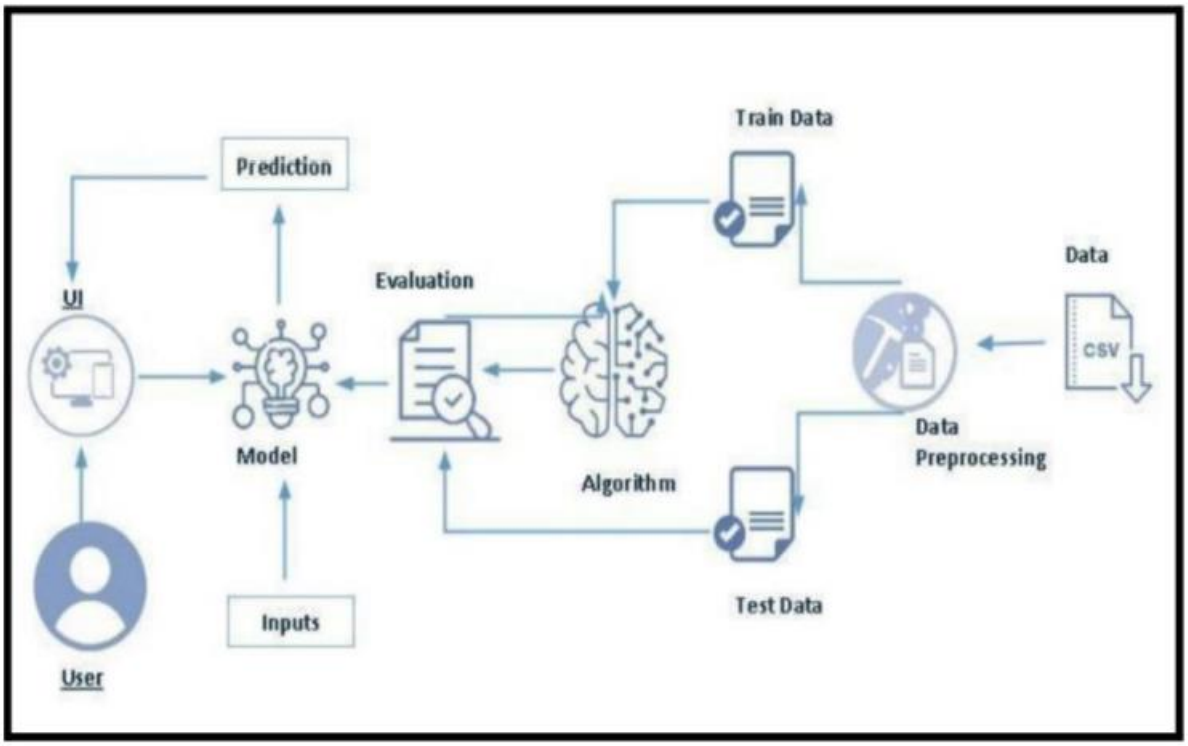
5.1Data Flow Diagrams & User Stories



User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Data Scientist	Data Collection Epic	USN-1	As a data scientist, I want to collect diverse data to train the model effectively	- Data collected from at least 1000 individuals. - Data includes age, gender, annual income, and car purchase information.	high	1
Data Engineer	Data Preprocessing Epic	USN-2	As a data engineer, I want to preprocess the raw data for better model performance.	- Missing values handled appropriately. - Categorical variables (like gender) encoded numerically. - Numerical feature scaling applied.	high	3
ML Developer	Data Splitting Epic	USN-3	As a machine learning developer, I want to split the data into training and testing sets.	- Data split into 80% training and 20% testing sets.	medium	2
ML Developer	Model Development Epic	USN-4	As a machine learning developer, I want to choose and train a suitable classification algorithm.	- Select a classification algorithm (e.g., Logistic Regression, Decision Trees). - Train the model on the training set.	high	1
ML Developer	Model Evaluation Epic	USN-5	As a machine learning developer, I want to evaluate the model's performance.	- Calculate accuracy, precision, recall, and F1 score. - Ensure model meets predefined performance thresholds.	low	1
DevOps Engineer	Deployment Epic	USN-6	As a DevOps engineer, I want to deploy the trained model for predictions on new data.	- Model deployed to a production environment. - Able to make predictions on new data.	high	3
Business Analyst	Feedback Loop Epic	USN-7	As a business analyst, I want a feedback loop to continuously improve the model.	- Collect feedback on model predictions. - Implement updates based on feedback.	medium	2

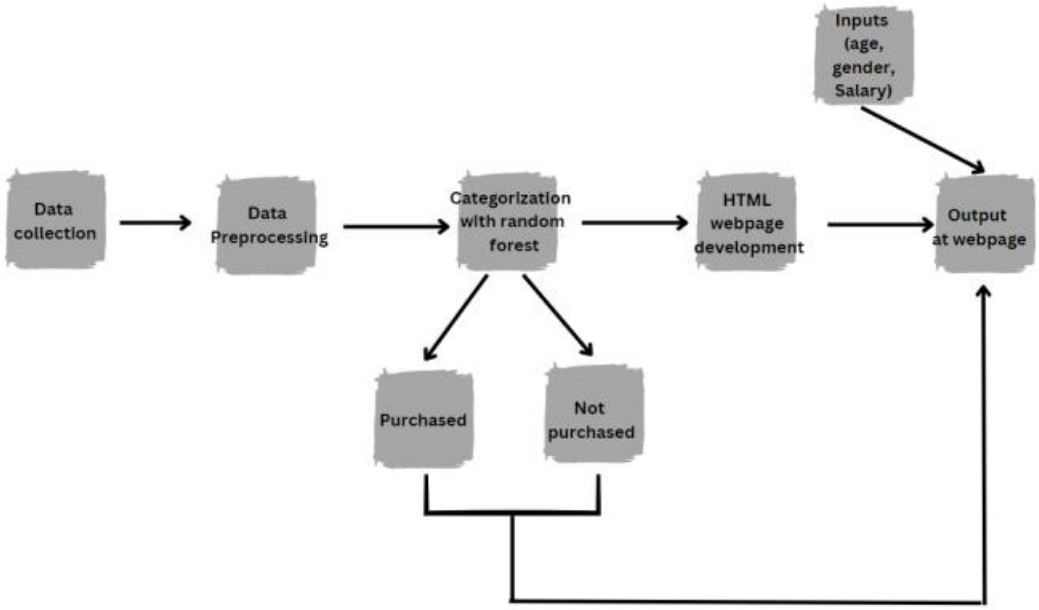
GITHUB LINK
https://github.com/smartinternz02/SI-GuidedProject-601748-1697636519/blob/main/Data%20Flow%20Diagrams%20and%20User%20Stories.pdf

5.2 Solution Architecture



6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture



6.2 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Story Points	Priority	Team Members
Sprint-1	Data collection epic	USN-1	As a data scientist, I want to collect diverse data to train the model effectively	Data collected from at least 1000 individuals. - Data includes age,gender, annual income, and car purchase information	1	High	Gokul Prasad
Sprint-1	Data processing epic	USN-2	As a data engineer, I want to preprocess the raw data for better model performance	Missing values handled appropriately. - Categorical variables (like gender)encoded numerically. - Numerical feature scaling applied	2	High	Mithun
Sprint-2	Data splitting epic	USN-3	As a machine learning developer, I want to split the data into training and testing sets.	Data split into 80% training and20% testing sets.	2	Medium	Threvikram
Sprint-2	Model development epic	USN-4	As a machine learning developer, I want to choose and train a suitable classification algorithm	Select a classification algorithm (e.g.,Logistic Regression, Decision Trees). - Train the model on the training set.	1	High	Nevatha
Sprint-3	Model Evaluation epic	USN-5	As a machine learning developer, I want to evaluate the model's performance	- Calculate accuracy, precision, recall, andF1 score. - Ensure model meets predefined performance thresholds	1	Low	Mithun
Sprint-2	Deployment epic	USN-6	As a DevOps engineer, I want to deploy the trained model for predictions on new data..	- Model deployed to a production environment. - Able to make predictions on new data.	3	High	Gokul prasad
Sprint-4	Feedback loop epic	USN-7	As a business analyst, I want a feedback loop to continuously improve the model.	- Collect feedback on model predictions. - Implement updates based on feedback	2	medium	Threvikram

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	3	2 Days	28 Oct 2023	30 Oct 2023		
Sprint-2	5	5 Days	1 Nov 2023	6 Nov 2023		
Sprint-3	10	2 Days	7 Nov 2023	8 Nov 2023		
Sprint-4	1	2 Day	9 Nov 2023	10 Nov 2023		

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

7.1 Feature 1

Our model is made using random forest which is a combination of multiple Decision trees, making the result more accurate. We have also added another feature, named “Age_Salaray_Interact” which is product of age and Salary for any given observation. This feature is used in our model building, hence during the input phase of our HTML webpage, the “Age_Salary_Interact” is a hidden input as it is not something the user is required to give input.

HTML page code

app.py > about

```python
@app.route("/about")
def about():
    return render_template("about.html")

@app.route("/carprediction")
def carprediction():
    return render_template("carprediction.html")

@app.route("/predict", methods=["POST"])
def predict():
    gender_str = str(request.form.get("gender"))
    if gender_str.lower() == "male":
        gender = 0
    else:
        gender = 1
    age = int(request.form.get("age"))
    annual_salary = int(request.form.get("annualSalary"))
    age_salary_interact = age * annual_salary
    prediction = rf_model.predict([[gender, age, annual_salary, age_salary_interact]])
    return render_template("carprediction.html", prediction=prediction)

if __name__ == "__main__":
    open_browser()
    app.run(host='127.0.0.1', port=5000, debug=True)
```

New feature added

```python
from sklearn.model_selection import train_test_split

# since age and annual salary show relatively high correlation to purchased, we can club them to form an interatction
df['Age_Salary_Interact'] = df['Age'] * df['AnnualSalary']

# Define the features and target variable
X = df[['Gender', 'Age', 'AnnualSalary', 'Age_Salary_Interact']]
y = df['Purchased']

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)
```

```python
X.head()
```

	Gender	Age	AnnualSalary	Age_Salary_Interact
0	1	35	20000	700000
1	1	40	43500	1740000
2	1	49	74000	3626000
3	1	40	107500	4300000
4	1	25	79000	1975000

7.2 Feature 2

Added a code to make sure the server shuts down with a press of a button in the terminal

```
43
44  def detect_q_key(event):
45      if event.event_type == keyboard.KEY_DOWN and event.name == 'q':
46          print("Terminating the server...")
47          keyboard.unhook_all()  # Unhook all keyboard events
48          exit()  # Terminate the serverq
49
50  # Hook the "q" key event
51  keyboard.hook(detect_q_key)
```

8. PERFORMANCE TESTING

8.1 Performance Metrics

Accuracy

```
# accuracy scores
print("Train Set Accuracy:", train_accuracy)
print("Test Set Accuracy:", test_accuracy)
```

```
Train Set Accuracy: 0.9975
Test Set Accuracy: 0.94
```

Confusion Matrix

```
from sklearn.metrics import confusion_matrix
pd.crosstab(y_test,y_pred)
```

col_0	0	1
Purchased		
0	115	6
1	6	73

Classification Report

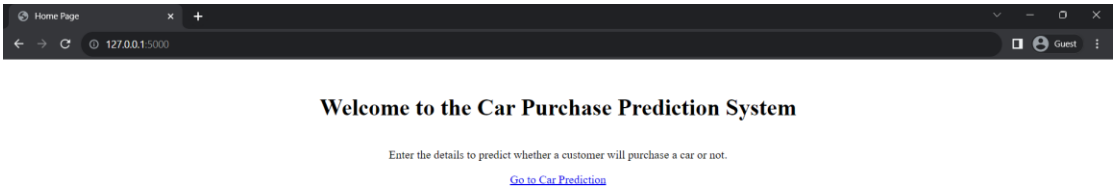
```
#Display a classification report
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.95      0.95      0.95       121
           1       0.92      0.92      0.92        79

    accuracy                           0.94       200
   macro avg       0.94      0.94      0.94       200
weighted avg       0.94      0.94      0.94       200
```

9. RESULTS

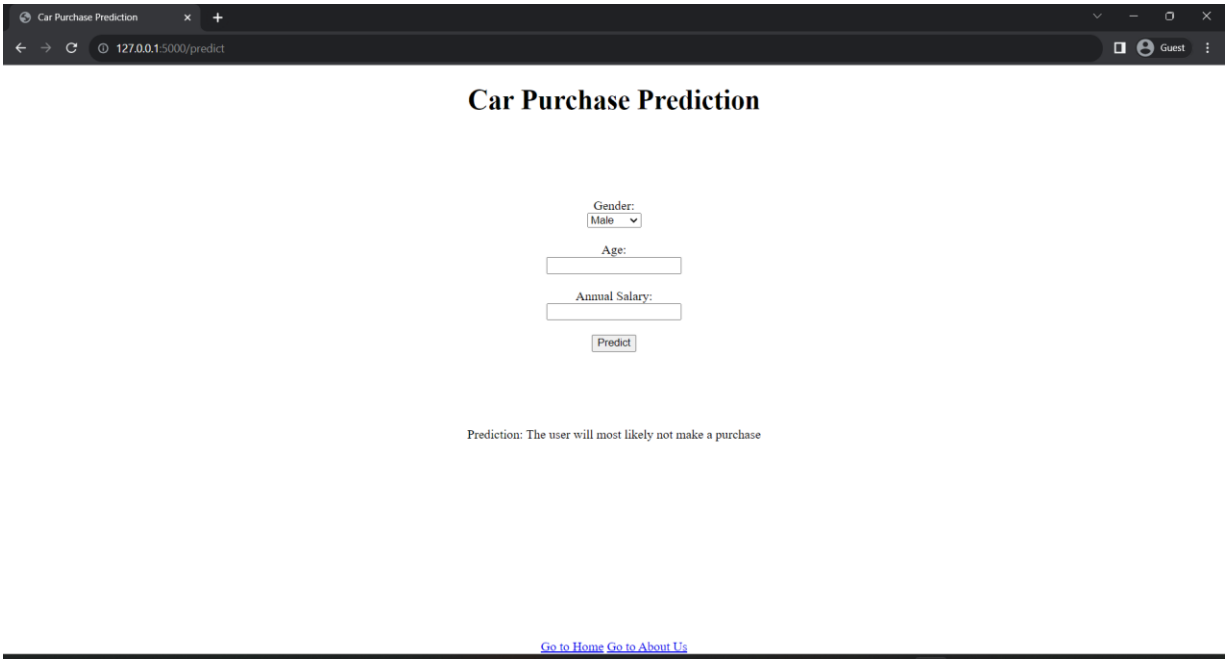
9.1 Output Screenshots

Home Page

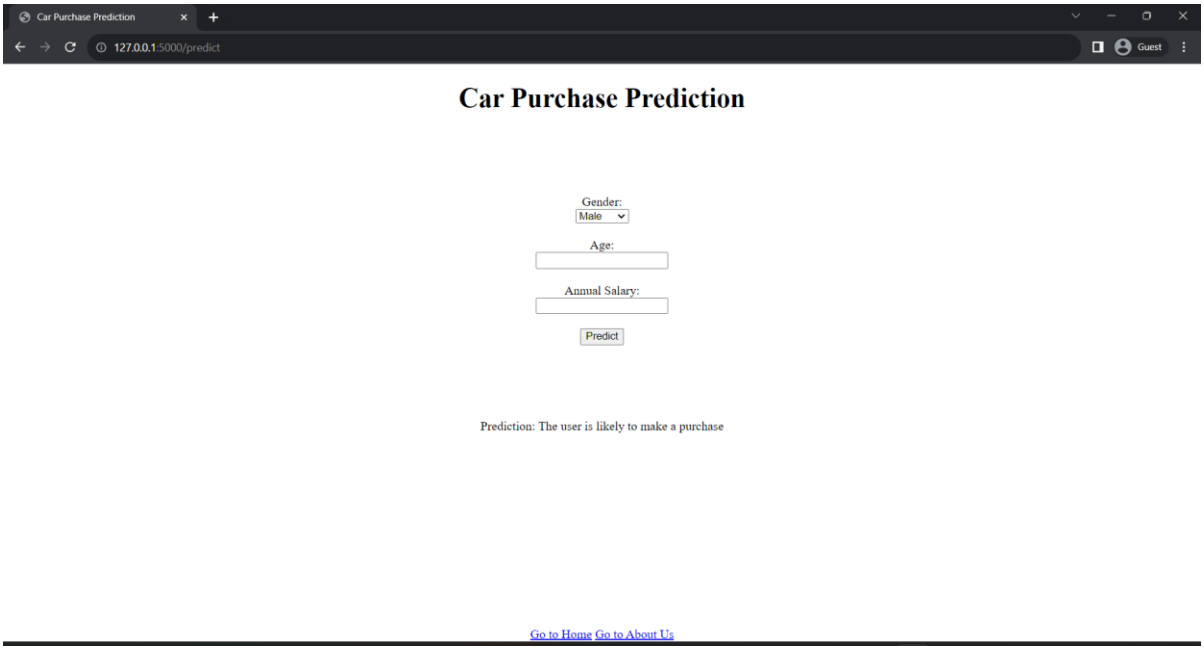


Prediction Page

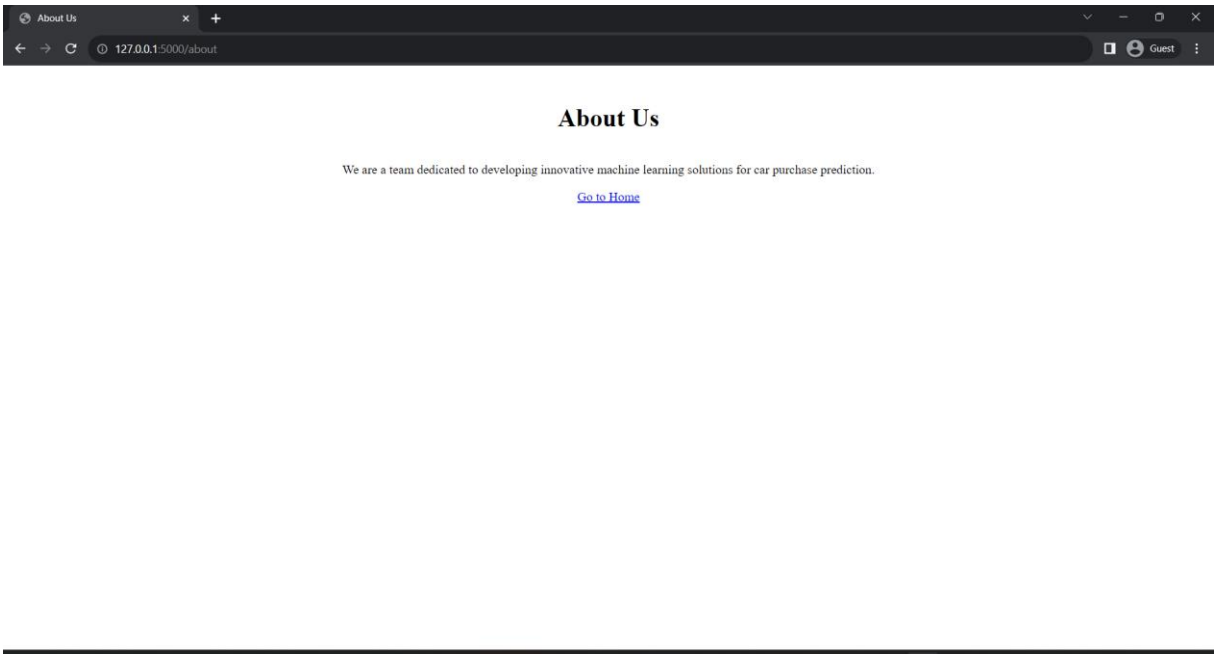
Output for (Male, Age 10, Salary 100)



Output for (Female, Age 45, Salary 98000)



About Us Page



10. ADVANTAGES & DISADVANTAGES

Advantages:

10.1 Precision: The model makes very accurate predictions, allowing potential car buyers to make better decisions.

10.2 Tailored Marketing: Allows the automotive sector to create customized marketing plans while maximizing resource allocation.

10.3 User Empowerment: Customers are empowered by the user-friendly interface, which provides exact purchase likelihoods based on their demographics.

Disadvantages:

10.4 Data Dependency: The accuracy of the model is largely dependent on the availability and quality of consumer data, which may restrict its efficacy in certain cases.

10.5 Over-Reliance on Historical Data: The emphasis on historical purchase patterns may fail to account for rapid shifts in customer behavior, resulting in mistakes on occasion.

11. CONCLUSION

Finally, the study provides a ground-breaking machine learning technique that enhances the accuracy of predicting car purchases dramatically. The concept empowers users, alters marketing techniques, and promotes informed decision-making by eliminating existing limits in the automotive sector. Despite some data quality dependencies, the benefits of precision and personalized marketing techniques exceed the drawbacks. This project establishes the groundwork for a data-driven revolution in the automotive industry.

12. FUTURE SCOPE

The project's future scope includes ongoing development of the machine learning model to respond to changing consumer patterns. Integration with emerging technologies, such as AI-powered chatbots for individualized consumer encounters, opens up new possibilities. Furthermore, expanding the model's skills to predict market trends and offer inventory management tactics could boost its impact on the automobile industry even further. Continuous research and development will assure the solution's continued relevance and efficacy in an ever-changing market scenario.

13. APPENDIX

Source Code

https://github.com/smartinternz02/SI-GuidedProject-601748-1697636519/blob/main/Project%20Development%20phase/Project%20files/car_purchase.ipynb

GitHub

https://github.com/smartinternz02/SI-GuidedProject-601748-1697636519

Project Demo Link

https://drive.google.com/file/d/1HuuGZ7WsMDiwGIX_UwF4na7Hpt4x0WFo/view?usp=drive_link