Travel Insurance Prediction using Machine Learning

TEAM ID –PNT2022TMID592946

VRUSHIKA SUNILKUMAR MODI (21BAI1901)

ARCHANA PALLAPOTHU (21BAI1871)

VARANASI VENKATA HERAMBHA PRANAV (21BAI1596)

PROJECT REPORT

1.INTRODUCTION

1.1 Project Overview:

In an era where travel has become an integral part of our lives, travel insurance plays a vital role in safeguarding travelers against unforeseen circumstances. However, predicting whether an individual will opt for travel insurance when planning a trip is a multifaceted challenge influenced by a myriad of factors. The "Travel Insurance Prediction Using Machine Learning" project seeks to address this challenge by harnessing the power of data-driven insights and machine learning. By analyzing historical data and patterns, the project aims to develop a predictive model that can anticipate whether a traveler is likely to purchase travel insurance, ultimately enhancing the way travel insurance is offered and marketed.

The "Travel Insurance Prediction Using Machine Learning" project is motivated by the need to improve the traveler's experience and the efficiency of the travel insurance industry. Our objective is to create a predictive model that can accurately forecast the likelihood of an individual purchasing travel insurance, thereby assisting both insurance providers and travel agencies in tailoring their services and marketing strategies. By understanding and leveraging historical data, we can unravel the intricate relationship between travelers and travel insurance, facilitating more informed decisions for all stakeholders involved.

The project's key components include:

- Data Collection: Gathering comprehensive historical data on travelers, including their personal details, travel plans, and past insurance choices.
- Data Preprocessing: Cleaning, transforming, and structuring the data for analysis and model training.
- Feature Selection and Engineering: Identifying the most relevant features and creating new variables that enhance the model's predictive capabilities.
- Model Development: Implementing and fine-tuning machine learning algorithms that can effectively forecast travel insurance purchase behavior.
- Model Evaluation: Assessing the model's performance using industry-standard metrics such as accuracy, precision, recall, and F1-score.
- Insights and Recommendations: Providing actionable insights derived from the model's predictions to insurance providers and travel agencies, enabling them to refine their marketing strategies and customize insurance offerings.

1.2 Purpose

The core purpose of the "Travel Insurance Prediction Using Machine Learning" project is to harness the power of data and machine learning to enhance the traveler's experience while simultaneously increasing the operational efficiency of the travel insurance industry. By achieving accurate predictions of insurance purchase behavior, the project aims to:

- Enable travel agencies to tailor insurance offerings to better suit the preferences and needs of their customers.
- Empower insurance providers to optimize their marketing efforts and resources by targeting individuals who are more likely to purchase insurance.
- Enhance the risk assessment capabilities of insurance companies by having a clearer understanding of the profiles of insured travelers.
- Contribute to the financial viability and profitability of the insurance industry by aligning offerings with customer expectations, thus increasing the adoption of travel insurance.

2. LITERATURE SURVEY

2.1 Existing problem

The landscape of travel insurance has witnessed significant changes in recent years, primarily due to the emergence of digital platforms and advancements in data analytics. However, a fundamental challenge that persists in the industry is the unpredictability of traveler's insurance purchase decisions. Travelers' choices regarding whether to buy insurance are influenced by a multitude of variables, such as their risk perception, past travel experiences, demographics, and the nature of their travel plans. The existing problem revolves around the need to accurately predict this behavior, as it has implications for both travelers and the insurance industry.

Traditionally, insurance providers and travel agencies have relied on demographic data and generic marketing strategies to promote travel insurance. While these approaches have yielded some success, they often lack the precision required to effectively target potential customers. As such, the need for a more data-driven and tailored approach has emerged. This is where machine learning comes into play, as it has the potential to leverage historical data to make predictions about travelers' insurance preferences.

2.2 References

- Machine Learning Prediction of Consumer Travel Insurance Purchase Behavior, Maksuda Akter Rubi (2021)
  https://ieeexplore.ieee.org/abstract/document/9984470/references#references

- Insurance Risk Prediction Using Machine Learning, Rahul Sahai (April 2023)
  https://link.springer.com/chapter/10.1007/978-981-99-0741-0_30

- Application of machine learning and data visualization techniques for decision support in the insurance sector, Seema Rawat (November 2021)
  https://www.sciencedirect.com/science/article/pii/S2667096821000057

- PREDICTING THE INSURANCE CLAIM BY EACH USER USING MACHINE LEARNING ALGORITHMS, Seshu Kumar (2022)
  https://jesne.org/index.php/JESNE/article/view/2

- Predicting the Willingness and Purchase of Travel Insurance During the COVID-19 Pandemic, Muhammad Khalilur Rahman (July 2022)
  https://www.frontiersin.org/articles/10.3389/fpubh.2022.907005/full

- Travel Insurance Prediction, PAUL HOLDSWORTH (2021)
  https://www.kaggle.com/code/paulh2718/travel-insurance-prediction

- By contrasting decision trees with logistic regression, a novel categorization-based cost prediction method for travel insurance may be developed under supervision, G. Satya Mounika Kalyani , Deepa. N (2019)
  https://sifisheriessciences.com/journal/index.php/journal/article/view/371

- Prediction of insurance fraud detection using machine learning algorithms, Laiqa Rukhsar; Waqas Haider Bangyal; Kashif Nisar; Sana Nisar (Jan 2022)
  https://search.informit.org/doi/abs/10.3316/informit.263147785515876

2.3 Problem Statement Definition

"The challenge is to develop a robust and accurate machine learning model capable of predicting whether an individual traveller will purchase travel insurance for their upcoming trip. This prediction should be based on historical data encompassing the traveller's demographics, travel plans, and past insurance decisions. The goal is to provide actionable insights for insurance providers and travel agencies, enabling them to personalize their offerings and marketing strategies to better cater to customer preferences and needs. By solving this problem, we aim to enhance the efficiency and effectiveness of the travel insurance industry while improving the traveller's experience."

In essence, the project aims to bridge the gap between traveller expectations and insurance industry offerings by leveraging data-driven predictions to better align the two. This problem statement serves as the foundation for the development of the predictive model and the subsequent insights and recommendations that will be provided to industry stakeholders.

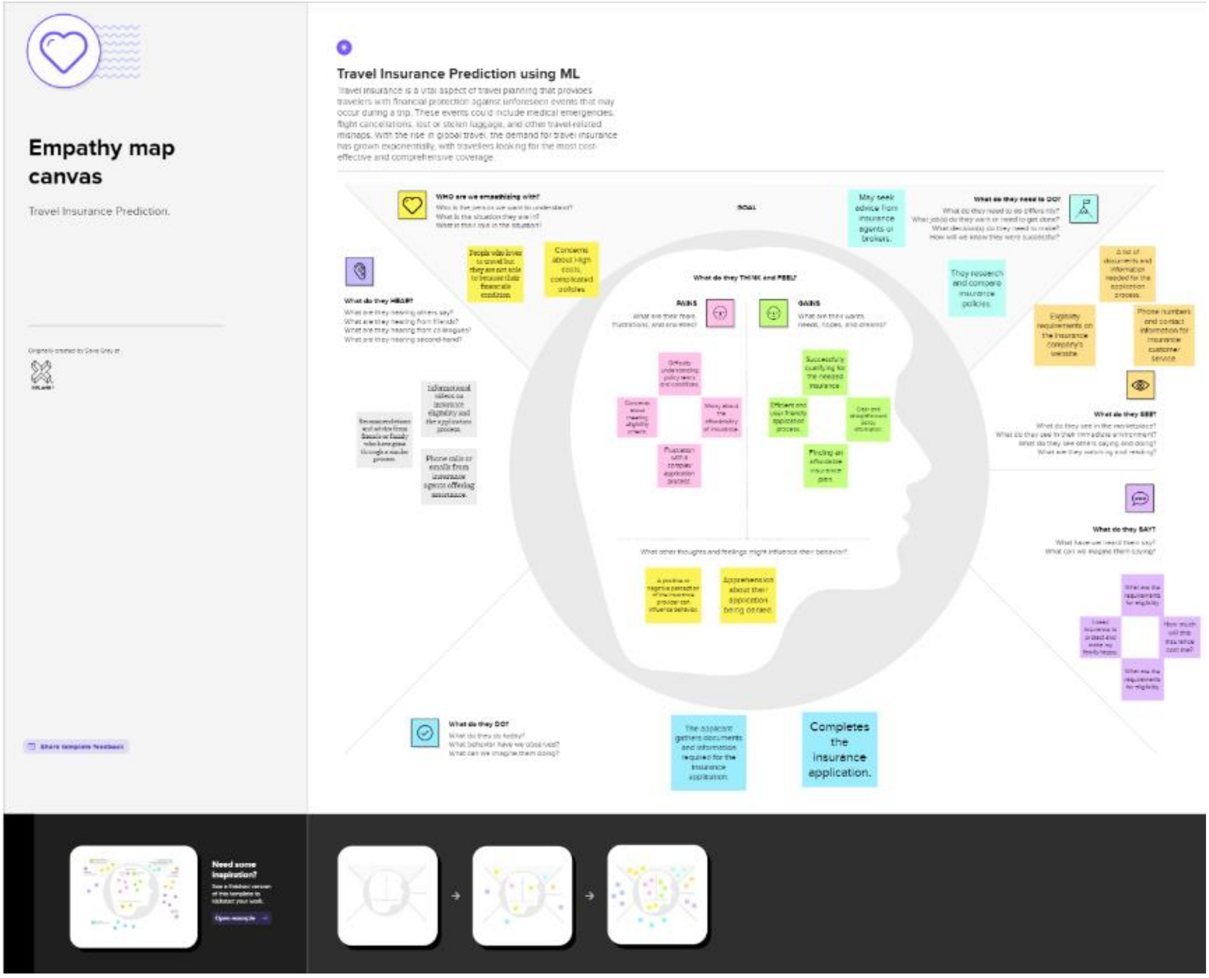
3.IDEATION & PROPOSED SOLUTION

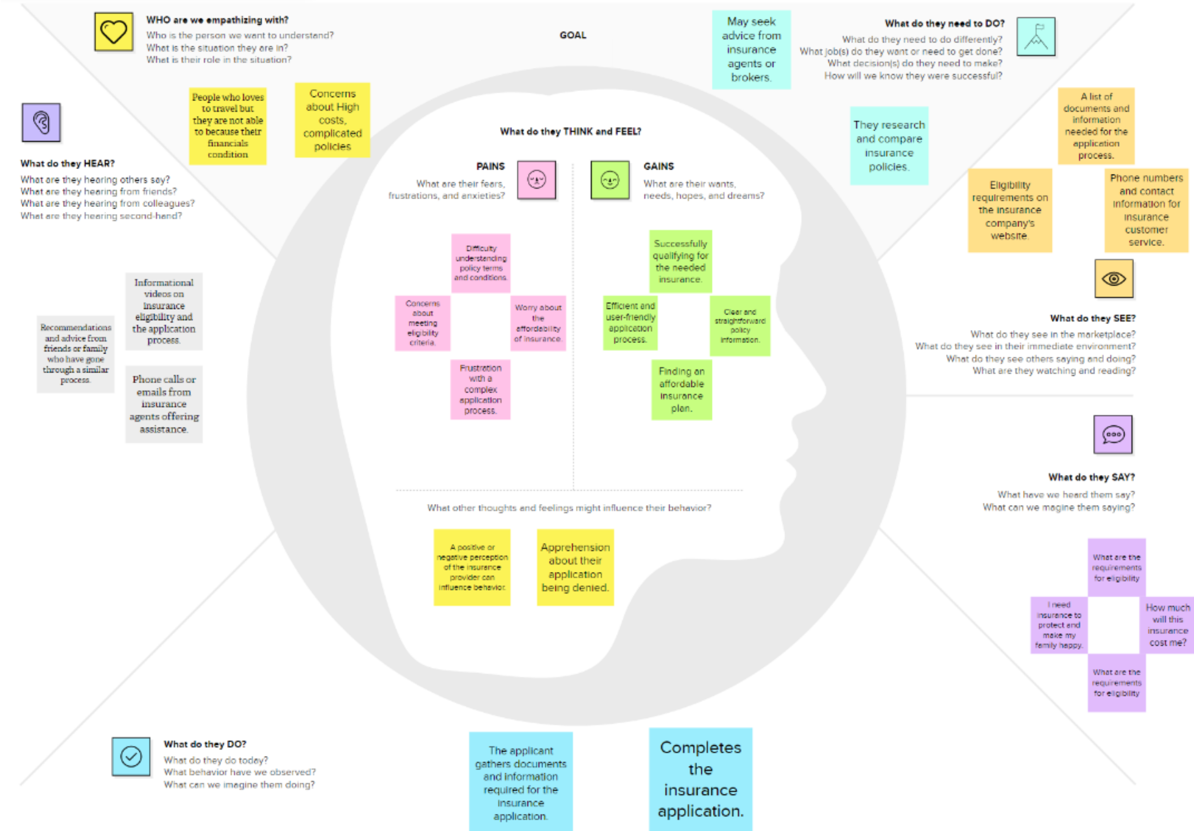
3.1Empathy Map Canvas

An empathy map is a straightforward and easily comprehensible visual tool that summarizes information about a user's actions and perspectives.
It serves as a valuable instrument for teams to gain a deeper understanding of their users.
To devise an effective solution, it is crucial to grasp the root of the problem and the individual experiencing it.
The process of creating the map encourages participants to consider matters from the user's vantage point, taking into account their objectives and difficulties.

Access link:

https://app.mural.co/t/travelinsuranceprediction9826/m/travelinsuranceprediction9826/1697474246693/9ab825f800a0b874f8f29d30ac3fe1c79d71c6f7?sender=ub6fbace4defee7a892633663


Empathy map canvas
Travel Insurance Prediction.
Share template feedback
Travel Insurance Prediction using ML
Travel insurance is a vital aspect of travel planning that provides travelers with financial protection against unforeseen events that may occur during a trip. These events could include medical emergencies, flight cancellations, lost or stolen luggage, and other travel-related mishaps. With the rise in global travel, the demand for travel insurance has grown exponentially, with travellers looking for the most cost-effective and comprehensive coverage.
WHO are we empathizing with?
GOAL
May seek advice from insurance agents or brokers.
What do they need to DO?
People who loves to travel but they are not able to because their financial condition
Concerns about high costs, complicated policies
What do they HEAR?
What are they hearing others say?
What are they hearing from friends?
What are they hearing from colleagues?
What are they hearing second-hand?
What do they THINK and FEEL?
PAINS
GAINS
They research and compare insurance policies.
Eligibility requirements on the insurance company's website
Phone numbers and contact information for insurance customer service
Informational videos on insurance eligibility and the application process.
Phone calls or emails from insurance agents offering assurance.
What do they SEE?
What do they SAY?
What other thoughts and feelings might influence their behavior?
Apprehension about their application being denied.
What do they DO?
What do they do today?
What behavior have we observed?
What can we imagine them doing?
The applicant gathers documents and information required for the insurance application.
Completes the insurance application.
Need some inspiration?



3.2 Ideation & Brainstorming

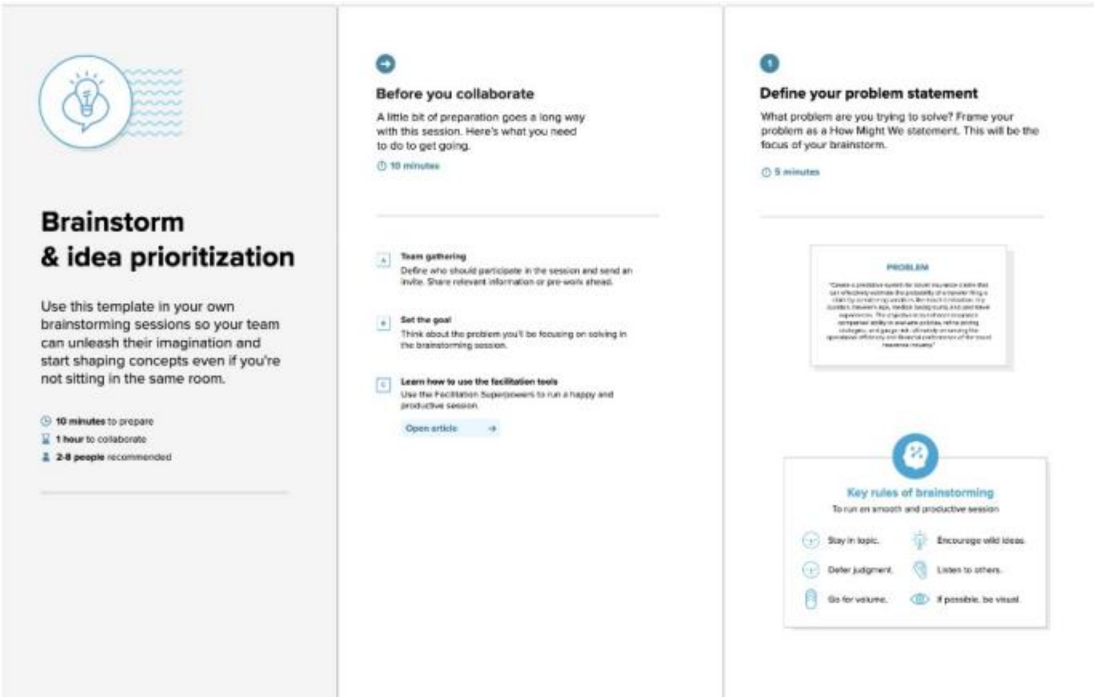
Brainstorming fosters an environment of openness and inclusivity, inviting all team members to engage in the imaginative thought process that paves the way for problem resolution.

By emphasizing quantity over quality, unconventional ideas are embraced and elaborated upon, while collaboration is encouraged among participants, enabling them to collectively generate a wealth of innovative solutions.

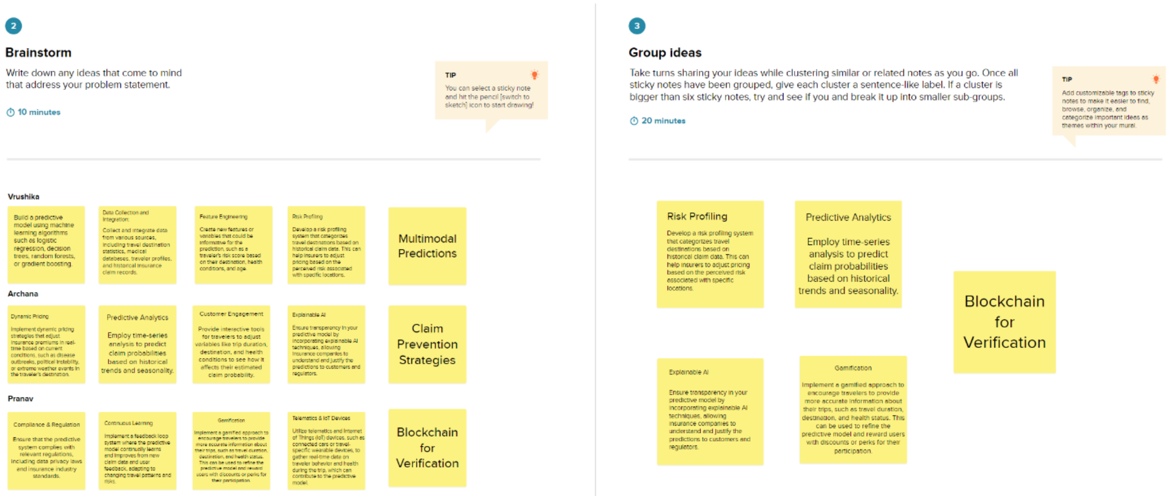
Link:

https://app.mural.co/t/archanapallapothu7560/m/archanapallapothu7560/1697649986160/8118f6a4a5e77293bf8d8fd68c6362672f391134?sender=uf551e4198b1165d998529392

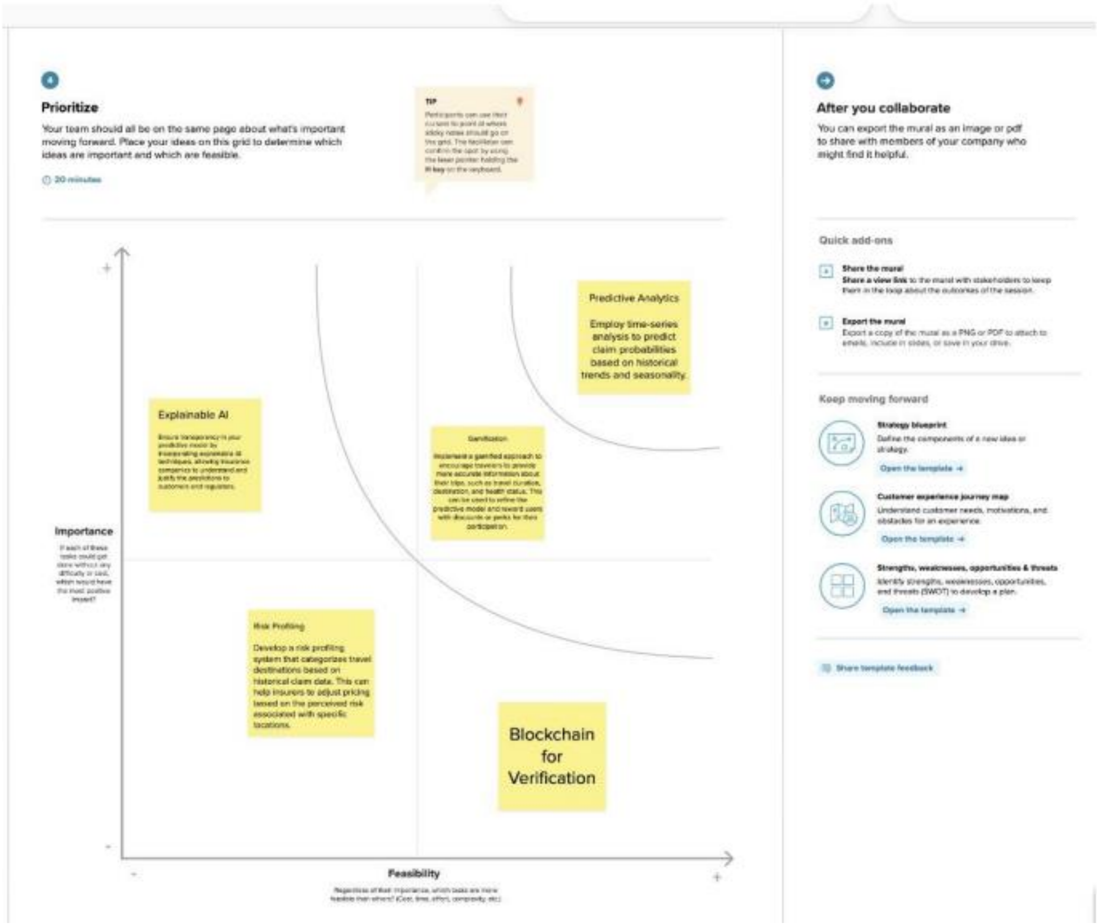
Step-1: Team Gathering, Collaboration and Select the Problem Statement



Step-2: Brainstorm, Idea Listing and Grouping



Step-3: Idea Prioritization



4.REQUIREMENT ANALYSIS

4.1 Functional requirement

Data Collection and Integration:

The system shall be able to gather and integrate historical data from various sources, including travel agencies, insurance providers, and relevant public databases.

The system should ensure the quality and integrity of the collected data through data cleansing and validation procedures.

Feature Extraction and Engineering:

The system shall identify and extract relevant features from the collected data, such as traveller demographics, travel itineraries, and previous insurance purchase history.

It should also have the capability to engineer new features to enhance prediction accuracy.

Model Development:

The system must implement various machine learning algorithms for predicting travel insurance purchase behaviour.

It should allow for model training, hyperparameter tuning, and cross-validation to optimize model performance.

Prediction Generation:

The system should provide a function for generating predictions on whether a traveller is likely to purchase travel insurance based on the provided input data.

User Interface:

The system should offer a user-friendly interface, accessible to authorized personnel, to input data, initiate predictions, and view the results.

Performance Evaluation:

The system must include a component for evaluating model performance through appropriate metrics such as accuracy, precision, recall, and F1-score.

It should provide visualizations and reports to convey the model's performance effectively.

Actionable Insights:

The system should be capable of generating actionable insights based on prediction results, allowing insurance providers and travel agencies to make informed decisions.

It should offer recommendations for adjusting marketing strategies and insurance offerings.

Data Privacy and Security:

The system must ensure the confidentiality and security of the sensitive traveller data it handles. Access should be restricted to authorized users only.

Compliance with data privacy regulations and best practices should be maintained.

Scalability and Performance:

The system should be designed to handle large volumes of data and accommodate future growth in data and user activity.

It should have optimal response times to meet the needs of real-time applications.

Logging and Auditing:

The system should maintain logs of user activities, model training, and predictions.

It should support auditing to trace data changes and system access.

These functional requirements lay the foundation for the development of a system that can predict travel insurance purchase behaviour accurately and provide actionable insights for stakeholders in the travel and insurance industries. The specific functionalities can be further refined and detailed during the project's planning and implementation phases.

4.2 Non-Functional requirements

Performance:

The system should be able to make predictions rapidly, ideally in real-time or near real-time, to support dynamic decision-making.

It must be capable of handling a significant number of prediction requests concurrently without degradation in performance.

Scalability:

The system should be designed to scale horizontally and vertically to accommodate growing data volumes and user loads.

It must support the addition of new data sources and features without significant disruptions.

Reliability:

The system must have high availability and uptime to ensure that predictions can be made consistently.

It should include automated failover and recovery mechanisms in the event of system failures.

Data Privacy and Security:

Data should be stored and transmitted securely, with encryption protocols in place.

Access controls and authentication mechanisms should be implemented to safeguard sensitive data.

Compliance with data protection regulations (e.g., GDPR) should be strictly adhered to.

Usability:

The system's user interface should be intuitive and easy to navigate for authorized personnel.

The system should provide user documentation or training materials for effective use.

Interoperability:

The system should support integration with various data sources and external systems used by insurance providers and travel agencies.

It should adhere to industry standards and data formats to facilitate seamless data exchange.

Maintainability:

The system's codebase should be well-documented, modular, and structured for ease of maintenance and updates.

It should be designed with version control and change management in mind.

Performance Monitoring and Logging:

The system must include comprehensive monitoring capabilities to track performance metrics and system health.

Logging should be implemented to record system events and errors for debugging and auditing.

Compliance:

The system should comply with relevant industry regulations, such as those in the insurance and data privacy sectors.

Regular compliance audits and updates should be conducted as needed.

Ethical Considerations:

The system should be designed and operated in an ethical manner, ensuring fairness and transparency in prediction results.

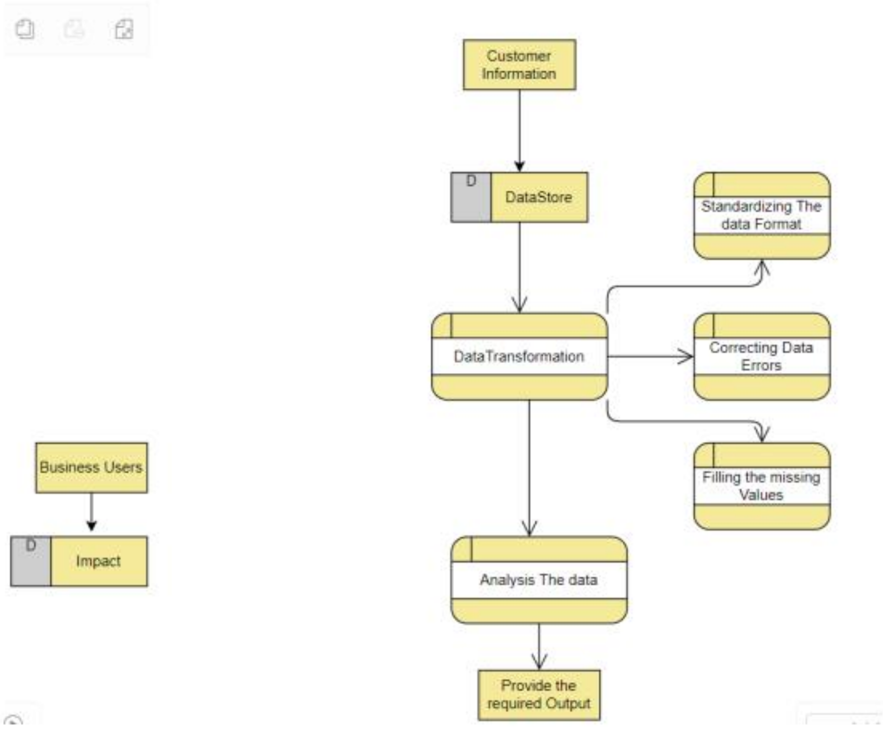
It should avoid any form of bias or discrimination in the prediction models.

These non-functional requirements are essential for creating a robust and reliable system that not only predicts travel insurance purchase behaviour accurately but also meets the highest standards of performance, security, and compliance while ensuring ethical considerations are taken into account.

5.PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories

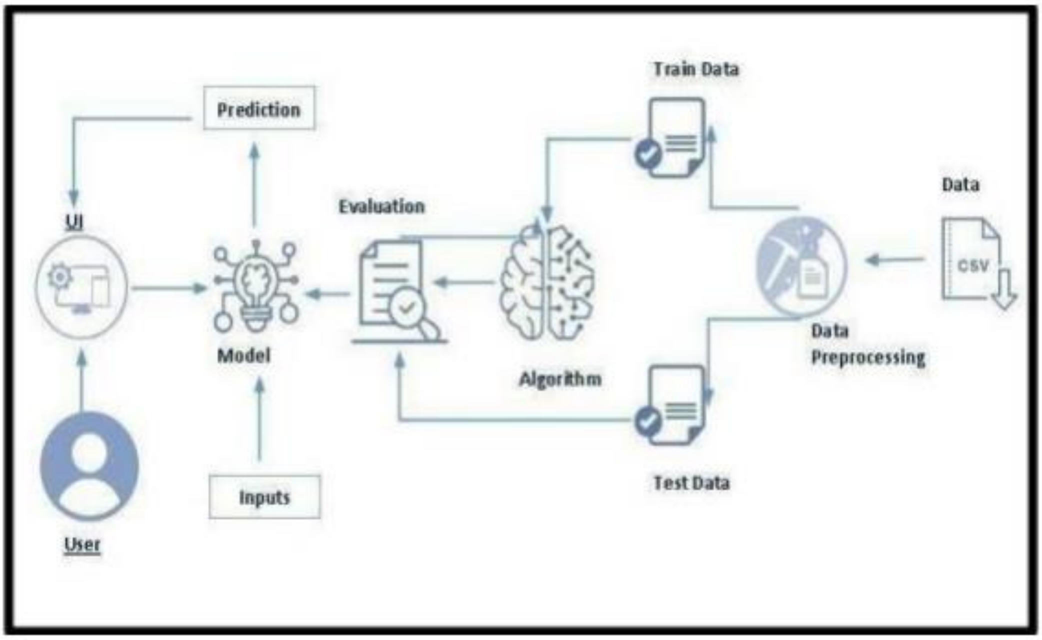
A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored. The data flow diagram (DFD) for the Travel Insurance analysis project illustrates a multi-level structure. At the top level (Level 0), the "Trave Insurance Analysis System" acts as the core, receiving data from "Travel Insurance Data," processing it, training machine learning models, and providing forecasts through a Flask web application. The machine learning models, at Level 1, interact with data stores for configuration and feature importance. The Flask web application, also at Level 1, communicates with users and passes their inputs for sales forecasting. Additionally, the system supports IBM Cloud deployment, allowing users to retrieve predictions. This DFD provides a concise visual representation of how data flows through various processes, entities, and data stores within the project, facilitating the understanding of the system's architecture and data pathways.



User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Data Analyst	Data Preparation	USN-1	As a data Analyst, I will collect the history of travel insurance	The System should be able to store the data in structured format	High	Sprint-1
		USN-2	As a data analyst, I need to preprocess the collected data, including handling missing values and outliers.	The system should successfully clean and preprocess the data, resulting in a high-quality dataset for analysis	High	Sprint-1
	Sales Forecasting	USN-3	As a data analyst, I want to apply machine learning algorithms like Random Forest, Decision Tree, XgBoost, and ARIMA to forecast future sales	The system should train and test these algorithms, providing accurate sales forecasts.	High	Sprint-2
	Deployment and Integration	USN-4	As a data analyst, I want to integrate the analysis and forecasting models into a Flask web application.	The system should create a user-friendly web interface for stakeholders to access the analysis and forecasts.	Medium	Sprint-3
		USN-5	As a data analyst, I need to deploy the Flask application on IBM Cloud for easy access and scalability.	The system should deploy the Flask application on the IBM Cloud platform, ensuring it is accessible to authorized users	High	Sprint-4

5.2 Solution Architecture



6.PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture

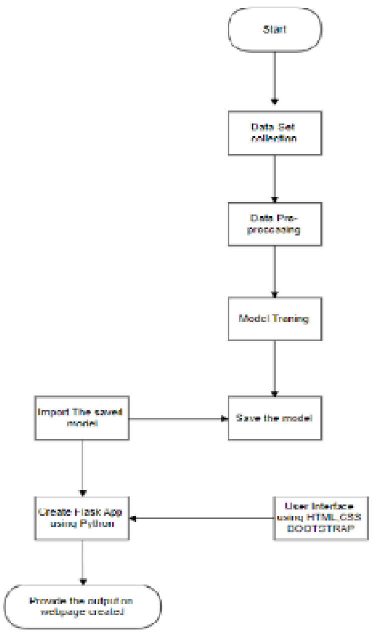


Table-1: Components & Technologies:

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

Table-2: Application Characteristics:

S. No	Characteristics	Description	Technology
1.	Open-Source Frameworks	Utilize open-source frameworks for development, machine learning, and data analysis.	Python's Flask, Scikit-Learn
2.	Security Implementations	Implement security measures to protect data and user interactions within the application.	SSL/TLS, Encryption, Authentication.
3.	Scalable Architecture	Design the architecture to be scalable, allowing the application to handle growing data and user loads.	Cloud Services (e.g., AWS Auto Scaling), Load Balancing
4.	Availability	Ensure high availability of the application, minimizing downtime and disruptions	Redundancy, Failover, Monitoring and Alerting
5.	Performance	Optimize application performance for responsiveness and efficient use of resources	Caching, Database Indexing, Efficient Algorithms

6.2 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint-1	Project setup & Infrastructure	USN-1	Set up the development environment with the required tools and frameworks to start the garbage classification project.	5	High	Vrushika
Sprint-1	development environment	USN-2	Gather a diverse dataset for training the machine learning model.	5	High	Archana
Sprint-2	Data Pre-processing	USN-3	Preprocess the collected dataset by resizing images, normalizing pixel values, and splitting it into training and validation sets.	5	High	Pranav
Sprint-2	Model Selection	USN-4	Explore and evaluate different machine learning architectures (e.g., random-forest) to select the most suitable model for travel-insurance prediction classification.	4	High	Archana
Sprint-3	Model Development	USN-5	Train the selected machine-learning model using pre-processed dataset and monitor its performance	5	High	Vrushika
Sprint-3		USN-6	Improve the model accuracy and robustness	3	Medium	Archana
Sprint-4	model deployment & Integration	USN-7	Deploy the trained machine learning model as an API or web service to make it accessible for travel insurance classification. integrate the model's API into a user-friendly web interface for users to receive travel insurance classification results based on the user input.	1	Medium	Vrushika
Sprint-5	Testing & quality assurance	USN-8	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.	1	Medium	Pranav

6.3 Sprint Delivery Schedule

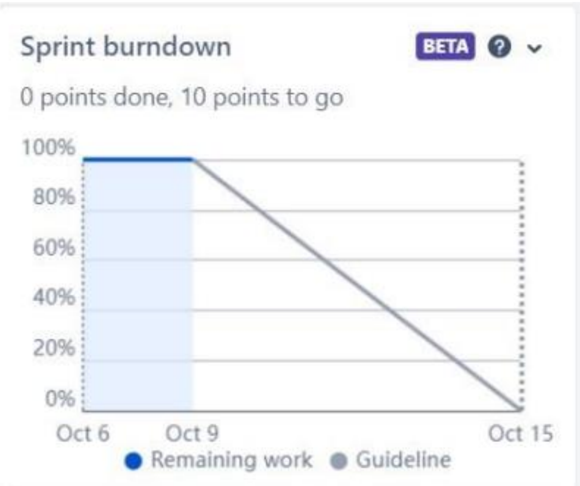
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	10	5 Days	18 Oct 2023	23 Oct 2023	10	23 Oct 2023
Sprint-2	9	5 Days	23 Oct 2023	28 Oct 2023	7	28 Oct 2023
Sprint-3	8	4 Days	28 Oct 2023	1 Nov 2023	5	
Sprint-4	5	4 Days	1 Nov 2023	4 Nov 2023	3	
Sprint-5	5	6 Days	4 Nov 2023	9 Nov 2023	2	

Velocity:

Average Velocity = Total Story Points Completed / Total Duration of Sprints

Total Story Points Completed = 10 + 7 + 5 + 3+ 3 = 28

Total Duration of Sprints = 5 + 5 + 4 + 4 + 6 = 24
Average Velocity = 28 / 24 = 1.16



7. CODING & SOLUTIONING

Data Pre-processing

As we seen and understood the description of the data, lets pre-process the collected data. The

download data set is not suitable for training the machine learning model as it might have so much of

randomness so, the dataset has to be cleaned properly in order to fetch good results. This activity

includes the following steps.

- Handling missing values
- Handling categorical data
- Handling outliers
- Splitting dataset into training and test set

In [41]:
```
from sklearn.preprocessing import MinMaxScaler
scale=MinMaxScaler()
```

In [42]:
```
x_scale=pd.DataFrame(scale.fit_transform(x),columns=x.columns)
x_scale.head()
```

Out[42]:

	Unnamed: 0	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad
0	0.000000	0.6	0.0	1.0	0.066667	0.571429	1.0	0.0	0.0
1	0.000504	0.6	1.0	1.0	0.633333	0.714286	0.0	0.0	0.0
2	0.001007	0.9	1.0	1.0	0.133333	0.285714	1.0	0.0	0.0
3	0.001511	0.3	1.0	1.0	0.266667	0.142857	1.0	0.0	0.0
4	0.002014	0.3	1.0	1.0	0.266667	0.857143	1.0	1.0	0.0

In [43]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x_scale,y,test_size=0.2,random_state=42)
```

In [44]:
```
x_train.shape
```

Out[44]: (1589, 9)

Label Encoding:

In [30]:
```
#Converting Categorical Data to Numerical Data using Encoding
le = LabelEncoder()
df['Employment Type'] = le.fit_transform(df['Employment Type'])
df['GraduateOrNot'] = le.fit_transform(df['GraduateOrNot'])
df['EverTravelledAbroad'] = le.fit_transform(df['EverTravelledAbroad'])
df['FrequentFlyer'] = le.fit_transform(df['FrequentFlyer'])
df['TravelInsurance'] = le.fit_transform(df['TravelInsurance'])
```

In [31]:
```
df.head()
```

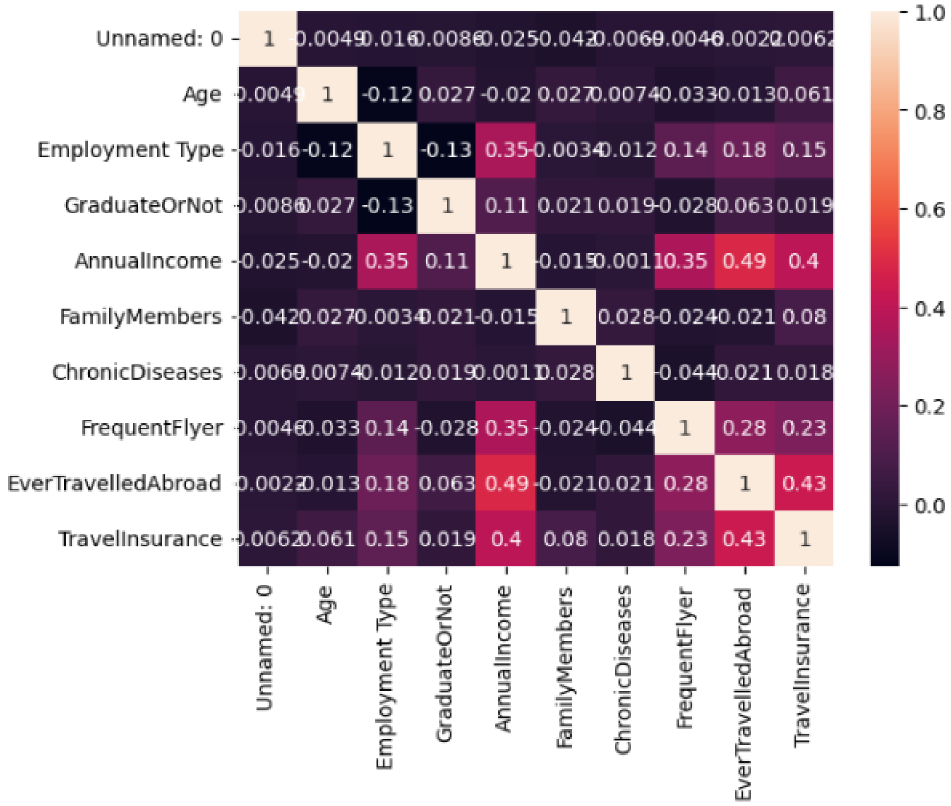
Out[31]:

	Unnamed: 0	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad
0	0	31	0	1	400000	6	1	0	0
1	1	31	1	1	1250000	7	0	0	0
2	2	34	1	1	500000	4	1	0	0
3	3	28	1	1	700000	3	1	0	0
4	4	28	1	1	700000	8	1	1	0

Data Visualisation-Heat map:

```
In [34]: sns.heatmap(df.corr(),annot=True)

Out[34]: <Axes: >
```



Model Building

Gradient Boost Classifier

In the realm of travel insurance prediction, the Gradient Boosting Classifier emerges as the linchpin. Leveraging ensemble learning, this algorithm sequentially refines its understanding of complex data, learning from its mistakes at each step. By analysing features like travel history and health records, it constructs a dynamic model adept at discerning nuanced patterns. Fine-tuned through hyperparameter optimization and validated against diverse datasets, it stands resilient against overfitting. The result is a formidable tool, accurate and versatile, ready for deployment. This model, a product of meticulous training and adaptation, becomes the guardian predicting insurance needs in the ever-evolving landscape of travel.

In [57]:

```python
from sklearn.ensemble import GradientBoostingClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,confusion_matrix,classification_report
from sklearn.datasets import load_digits
```

In [72]:

```python
from sklearn.model_selection import RandomizedSearchCV
model = GradientBoostingClassifier()
param_vals = {'max_depth': [200, 500, 800, 1100], 'n_estimators': [100, 200, 300, 400],'learning_rate': [0.001, 0.01, 0.1,
random_rf = RandomizedSearchCV(estimator=model, param_distributions=param_vals,
                               n_iter=10, scoring='accuracy', cv=5,
                               refit=True, n_jobs=-1)

#Training and prediction

random_rf.fit(x_train, y_train)
preds = random_rf.best_estimator_.predict(x_test)
preds
```

Out[72]:

```
array([0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1,
       0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1,
       1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 0, 0,
       1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1, 0,
       0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 1,
       1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 1,
       0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1,
       0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1,
       1, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0,
       0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 0,
       1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0,
       1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0,
       0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 1, 0, 0, 0, 1, 1, 0, 0, 1, 0, 0,
       0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 0,
       0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0,
       0, 0, 1, 1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 0,
       0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0, 1,
       0, 0])
```

In [73]:

```python
gbc = GradientBoostingClassifier(n_estimators=300,learning_rate=0.05,random_state=100,max_features=5 )
```

In [74]:

```python
gbc.fit(x_train, y_train)
```

Out[74]:

```
GradientBoostingClassifier(learning_rate=0.05, max_features=5, n_estimators=300,
                           random_state=100)
```

In [75]:

```python
pred_y = gbc.predict(x_test)
```

In [76]:

```python
#accuracy
acc = accuracy_score(y_test, pred_y)
print("Gradient Boosting Classifier accuracy is : {:.2f}".format(acc))
```

```
Gradient Boosting Classifier accuracy is : 0.84
```

In [77]:

```python
print(classification_report(y_test,pred_y))
```

```
              precision    recall  f1-score   support

           0       0.81      0.99      0.89       257
           1       0.98      0.57      0.72       141

    accuracy                           0.84       398
   macro avg       0.89      0.78      0.80       398
weighted avg       0.87      0.84      0.83       398
```

In [78]:

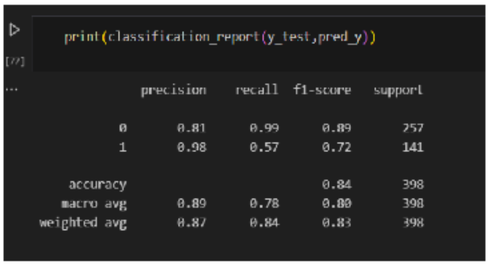
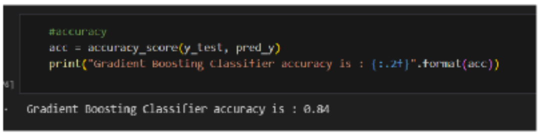
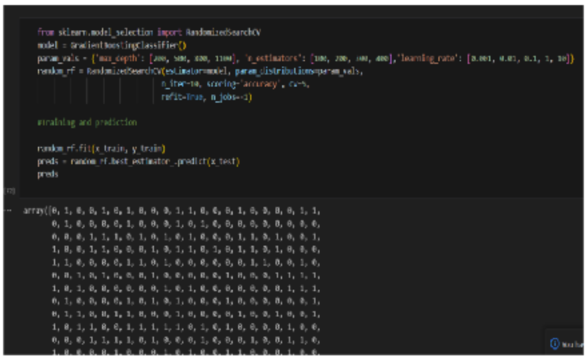
```python
import pickle
```

In [79]:

```python
pickle.dump(gbc,open('tip.pkl','wb'))
```

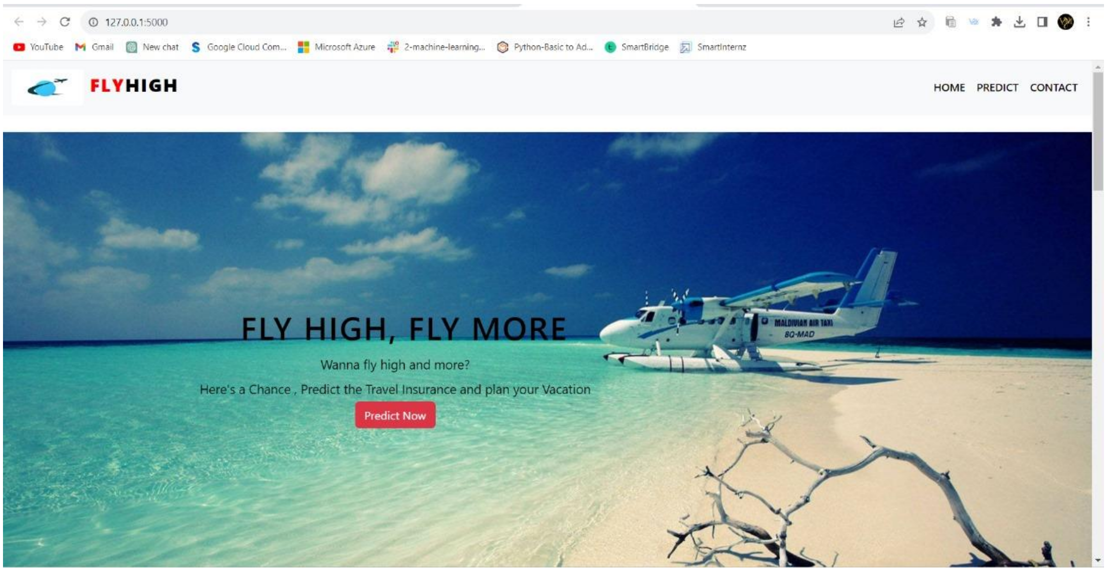
8. PERFORMANCE TESTING

8.1 Performance Metrics

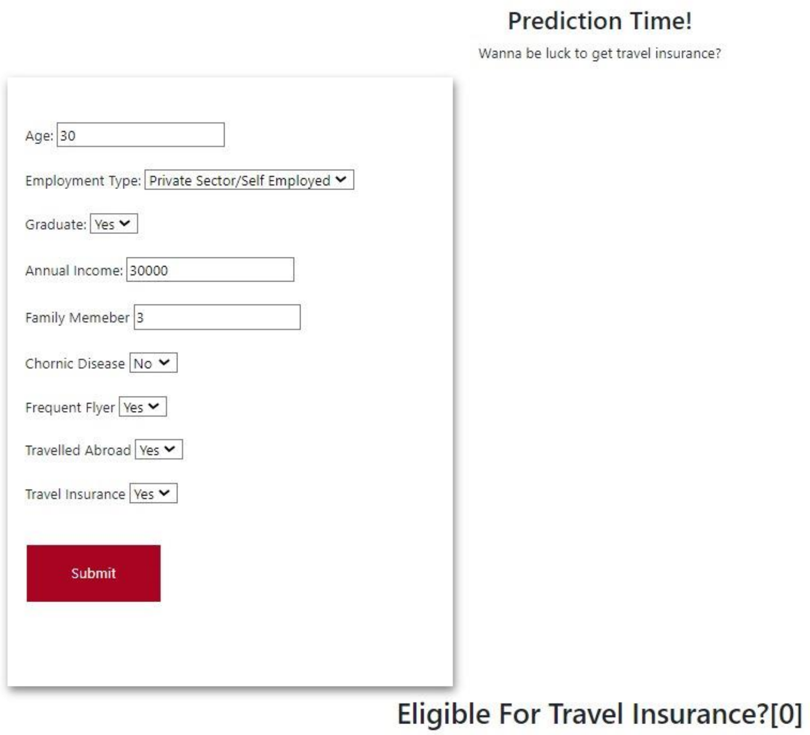
S.No.	Parameter	Values	Screenshot
1.	Metrics	**Classification Model:** Confusion Matrix -, Accuracy Score- & Classification Report -	 
2.	Tune the Model	Hyperparameter Tuning - Validation Method -	

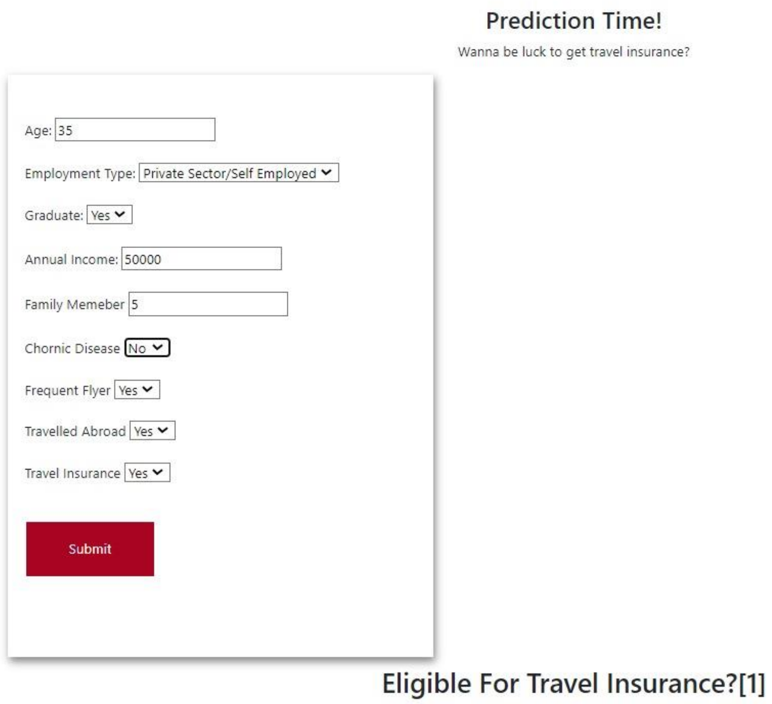
9. RESULTS

9.1 Website



9.2 Output Screenshot





10. ADVANTAGES & DISADVANTAGES

10.1 Advantages:

Improved Marketing Efficiency: Machine learning can identify travelers more likely to purchase insurance, allowing for targeted marketing efforts. This reduces marketing costs and increases conversion rates.

Enhanced Customer Experience: Personalized insurance offerings based on predictions can improve the customer experience, increasing customer satisfaction and loyalty.

Risk Assessment: Machine learning models can assess risk more accurately, allowing insurance providers to price policies more competitively and make informed underwriting decisions.

Data-Driven Insights: The project can provide valuable data-driven insights and recommendations to stakeholders, leading to more informed decision-making.

Automation: The system can automate the prediction process, reducing the need for manual analysis and decision-making.

Ethical Considerations: Machine learning can help identify and mitigate biases, ensuring fair and ethical insurance offerings.

10.2 Disadvantages:

Data Privacy Concerns: Handling sensitive traveler data poses privacy and security risks. Mishandling or data breaches can result in legal and reputational issues.

Data Quality: Accurate predictions rely on high-quality historical data. Incomplete or inaccurate data can lead to unreliable predictions.

Model Complexity: Developing and maintaining machine learning models can be complex and resource-intensive. It requires expertise in machine learning and data science.

Bias and Fairness: Machine learning models can inherit biases from training data, potentially leading to unfair or discriminatory predictions.

Interpretability: Complex machine learning models may lack interpretability, making it challenging to explain the basis for predictions to stakeholders and regulators.

Model Drift: Over time, the model's performance may degrade due to changing traveler behaviors. Continuous monitoring and model updates are necessary to mitigate drift.

Integration Challenges: Integrating the predictive model with existing systems and workflows may be challenging and require changes in business processes.

Regulatory Compliance: The project must comply with industry-specific regulations, which can add complexity and cost to the development process.

11. CONCLUSION

The "Travel Insurance Prediction Using Machine Learning" project represents a significant step forward in the travel insurance industry, transforming the way insurance is offered, marketed, and experienced. By harnessing the power of machine learning, it empowers insurance providers and travel agencies to make data-driven decisions, leading to more personalized, efficient, and customer-centric offerings.

The project's core value lies in its ability to predict travel insurance purchase behaviour accurately, thereby aligning traveller expectations with insurance offerings. This benefits travellers by providing insurance options that suit their needs and preferences. Simultaneously, insurance providers and travel agencies can optimize their marketing strategies, increase conversion rates, and enhance their risk assessment capabilities.

In the ever-evolving landscape of the travel insurance industry, this project lays the foundation for a future where data-driven insights and predictive modelling drive decision-making, making travel safer and more convenient for all. As the project continues to evolve and expand its scope, it promises to remain at the forefront of innovation and transformation within the travel insurance sector.

12. FUTURE SCOPE

The "Travel Insurance Prediction Using Machine Learning" project presents a compelling vision for the future of the travel insurance industry. As technology and data science continue to advance, the project's potential for growth and development is substantial. One of the primary areas of future expansion is in the refinement and optimization of machine learning models. As new algorithms and techniques emerge, the project can incorporate them to enhance the accuracy of predictions. Deep learning, reinforcement learning, and ensemble methods offer promising avenues for improving model performance.

Real-time prediction is another promising direction. By providing travellers with instant, on-the-fly predictions during the travel planning process, the system can not only enhance user experience but also empower travellers to make more informed decisions in real-time.

Dynamic pricing, based on a traveller's unique risk profile and real-time variables, can become a reality, allowing insurance providers to offer flexible and competitive premiums. AI chatbots can be integrated to interact with travellers, answering queries and assisting them in understanding and purchasing insurance seamlessly.

Additionally, the project can explore cross-selling opportunities by integrating with other travel-related services, such as flight and hotel bookings, to create a holistic travel ecosystem. By leveraging blockchain technology for data security and transparency, the project can address data privacy concerns and build trust among users.

Expanding the project's reach to international markets and catering to global travellers can also open up new avenues for growth and customer engagement. Furthermore, predictive analytics can be extended beyond insurance purchase predictions to anticipate and proactively address customer support needs, enhancing the overall customer experience.

13. APPENDIX

Source Code

https://github.com/smartinternz02/SI-GuidedProject-590456-1697474954/tree/main/Travel_Insurance%20Flask

GitHub Link

https://github.com/smartinternz02/SI-GuidedProject-590456-1697474954/tree/main

Demo Link

https://drive.google.com/file/d/1MefVhcx92kIWSaYbfT1mzLEXl60hIE7L/view?usp=sharing