Travel Insurance Prediction using Machine Learning

TEAM ID –PNT2022TMID592722

VEMULA SAI AKHIL (21BAI1864)

SHAIK AYAZ AHMED (21BRS1304)

PADIGIREDDY SIVA KARTHIK REDDY (21BAI1850)

GUDIVADA JAYANTH (21BAI1853)

PROJECT REPORT

1.INTRODUCTION

1.1 Project Overview:

In an age where travel has become an integral aspect of our lives, the significance of travel insurance cannot be overstated in safeguarding individuals against unforeseen circumstances. However, predicting whether an individual will choose to acquire travel insurance while planning a trip is a complex challenge influenced by various factors. The project titled "Travel Insurance Prediction Using Machine Learning" aims to tackle this challenge by utilizing data-driven insights and machine learning. Through the analysis of historical data and patterns, the project seeks to develop a predictive model capable of anticipating whether a traveler is inclined to purchase travel insurance. This endeavor ultimately aims to enhance the offering and marketing of travel insurance.

The motivation behind the "Travel Insurance Prediction Using Machine Learning" project lies in the desire to improve the overall experience for travelers and increase the efficiency of the travel insurance industry. The primary objective is to create a predictive model that can accurately foresee the likelihood of an individual purchasing travel insurance. This, in turn, assists both insurance providers and travel agencies in tailoring their services and refining their marketing strategies. By delving into and leveraging historical data, the project aims to unravel the intricate relationship between travelers and travel insurance, enabling more informed decisions for all stakeholders involved.

The critical elements of the project encompass:

- **Data Collection:** Thoroughly gathering historical data on travelers, encompassing their personal details, travel plans, and previous insurance choices.
- **Data Preprocessing:** Rigorously cleaning, transforming, and structuring the data to prepare it for analysis and model training.
- **Feature Selection and Engineering:** Identifying the most pertinent features and crafting new variables to augment the predictive capabilities of the model.
- **Model Development:** Implementing and refining machine learning algorithms that can adeptly anticipate travel insurance purchase behavior.
- **Model Evaluation:** Scrutinizing the model's performance through the use of industry-standard metrics such as accuracy, precision, recall, and F1-score.
- **Insights and Recommendations:** Furnishing actionable insights derived from the model's predictions to insurance providers and travel agencies, empowering them to enhance their marketing strategies and tailor insurance offerings.

1.2 Purpose

The fundamental objective of the "Travel Insurance Prediction Using Machine Learning" project is to leverage the capabilities of data and machine learning to enrich the traveler's experience while concurrently improving the operational efficiency of the travel insurance industry. Through the achievement of precise predictions regarding insurance purchase behavior, the project seeks to:Enable travel agencies to tailor insurance offerings to better suit the preferences and needs of their customers.

- **Facilitate Customized Offerings:** Enable travel agencies to tailor insurance offerings more effectively to align with the preferences and needs of their customers.
- **Optimize Marketing Efforts:** Empower insurance providers to optimize their marketing strategies and allocate resources efficiently by targeting individuals with a higher likelihood of purchasing insurance.
- **Enhance Risk Assessment:** Improve the risk assessment capabilities of insurance companies by gaining a clearer understanding of the profiles of insured travelers.
- **Boost Financial Viability:** Contribute to the financial viability and profitability of the insurance industry by aligning offerings with customer expectations, thereby increasing the adoption of travel insurance.

2. LITERATURE SURVEY

2.1 Existing problem

The landscape of travel insurance has undergone significant transformations in recent years, primarily propelled by the rise of digital platforms and advancements in data analytics. However, a core challenge persists in the industry, revolving around the unpredictable nature of travelers' decisions to purchase insurance. The factors influencing these decisions are diverse, including risk perception, past travel experiences, demographics, and the specifics of travel plans. Accurately predicting this behavior is a crucial issue with implications for both travelers and the insurance industry.

Historically, insurance providers and travel agencies have relied on demographic data and generic marketing strategies to promote travel insurance. While these methods have achieved some success, they often lack the precision needed to effectively target potential customers. Consequently, there is a growing demand for a more data-driven and tailored approach. This is where machine learning becomes pivotal, as it has the capacity to utilize historical data to make informed 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 categorizationbased 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/37_1

- 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 at hand involves crafting a robust and precise machine learning model with the capability to predict whether an individual traveler will opt for travel insurance in their upcoming journey. This prediction hinges on a comprehensive analysis of historical data, encompassing the traveler's demographics, travel plans, and previous insurance choices. The ultimate objective is to furnish actionable insights for insurance providers and travel agencies, empowering them to tailor their offerings and marketing strategies to align more closely with customer preferences and needs. By addressing this challenge, the project strives to elevate the efficiency and efficacy of the travel insurance industry, concurrently enhancing the overall experience for travelers.

In essence, the project seeks to narrow the divide between traveler expectations and the offerings of the insurance industry. This is achieved through the utilization of data-driven predictions, aiming to better synchronize the two realms. The articulated problem statement forms the bedrock for the development of the predictive model and subsequent insights, shaping the path toward providing valuable recommendations to industry stakeholders.

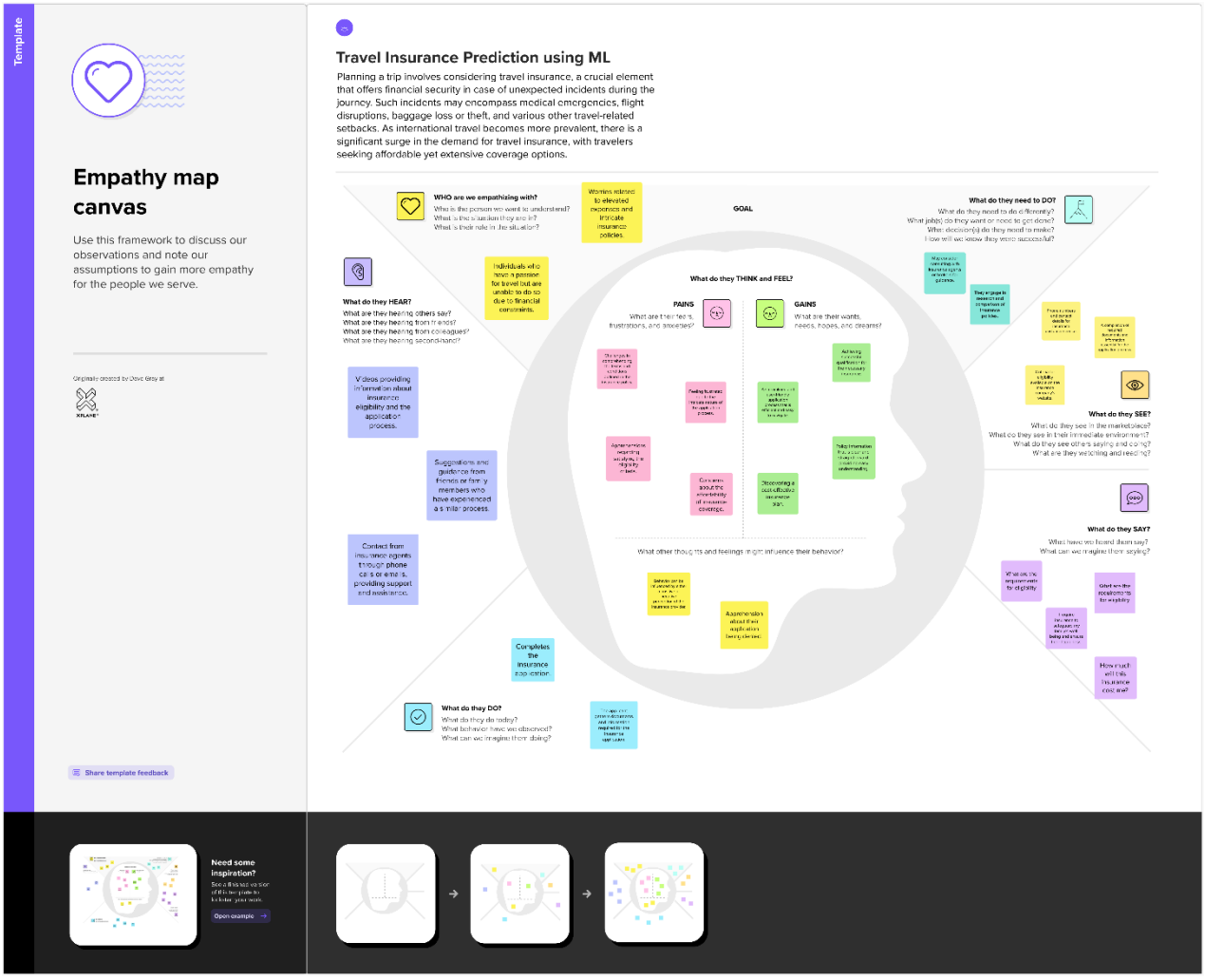
3.IDEATION & PROPOSED SOLUTION

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/intern4910/m/intern4910/1700461361168/140799238fbd5355ac410a326e148c1ad975eddb?sender=u4bf7b24b70cf26e2ae1d7818



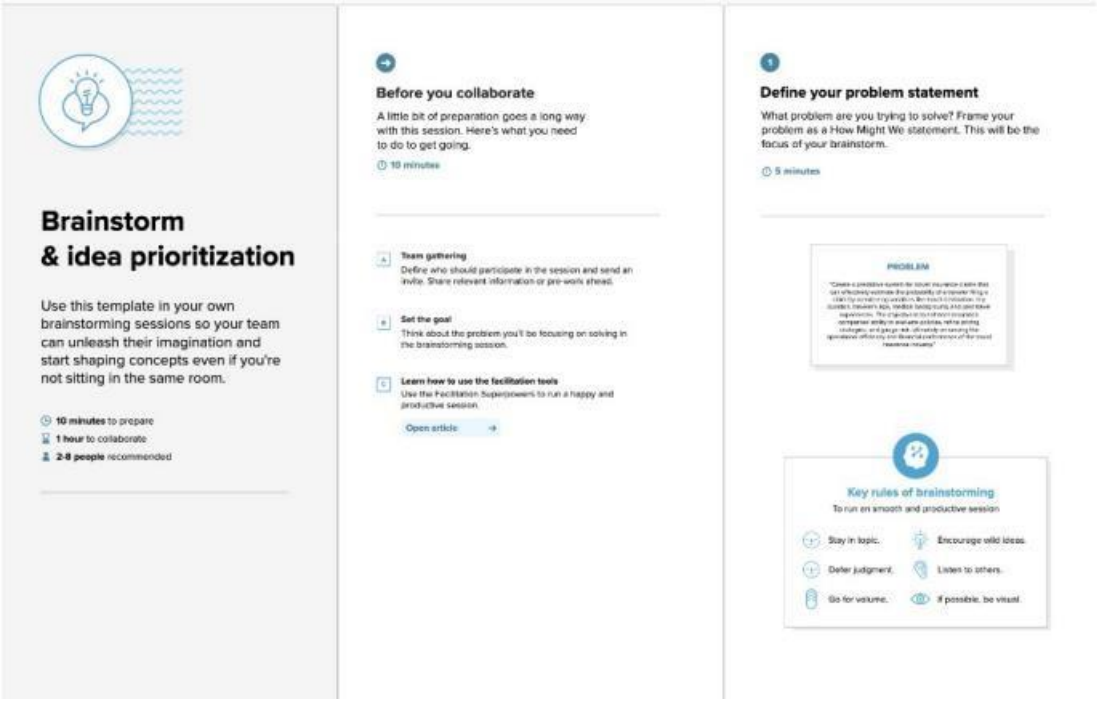
3.2 Ideation & Brainstorming

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.

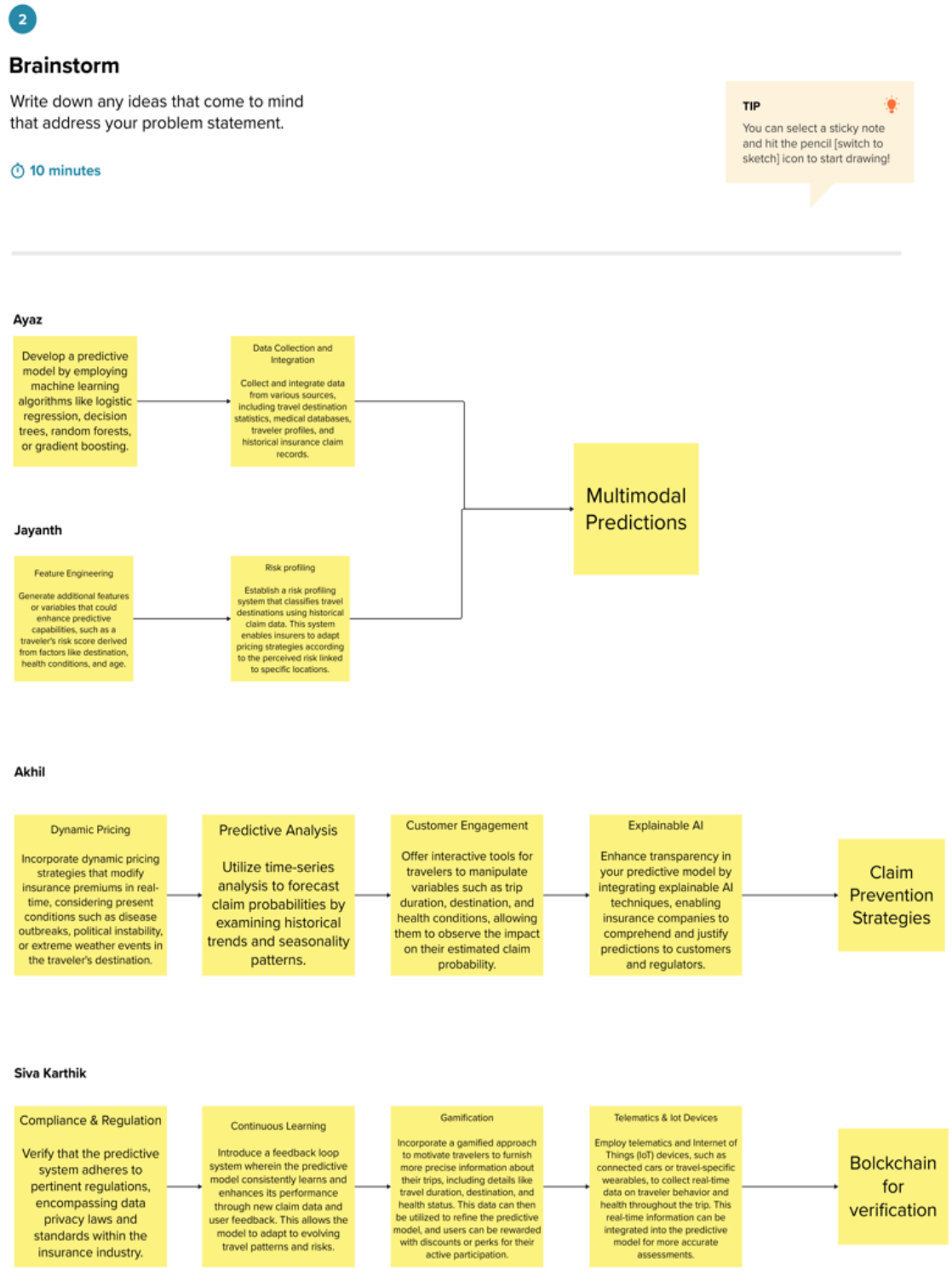
Link:

https://app.mural.co/t/intern4910/m/intern4910/1700465821605/3fcba4dbacaacc53e5480f61058e4d6338d90267?sender=u4bf7b24b70cf26e2ae1d7818

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



Step-2: Brainstorm, Idea Listing and Grouping



3

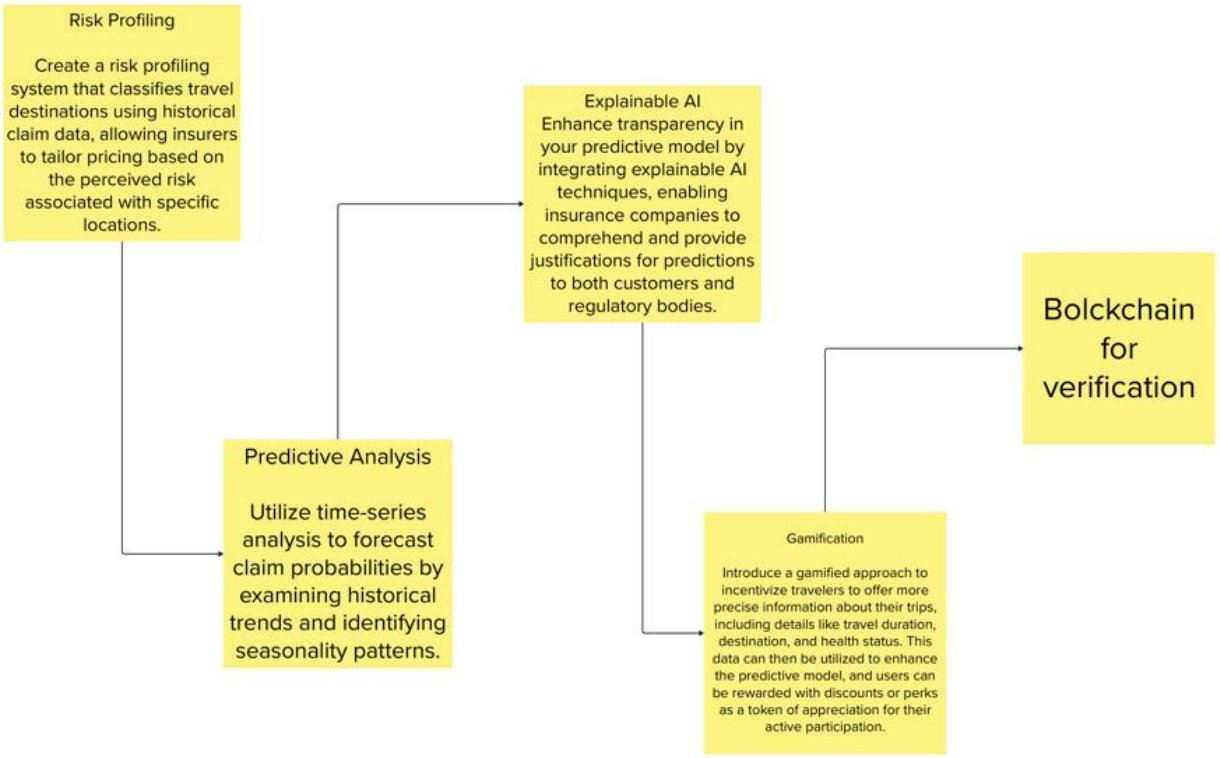
Group ideas

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.



Step-3: Idea Prioritization

4

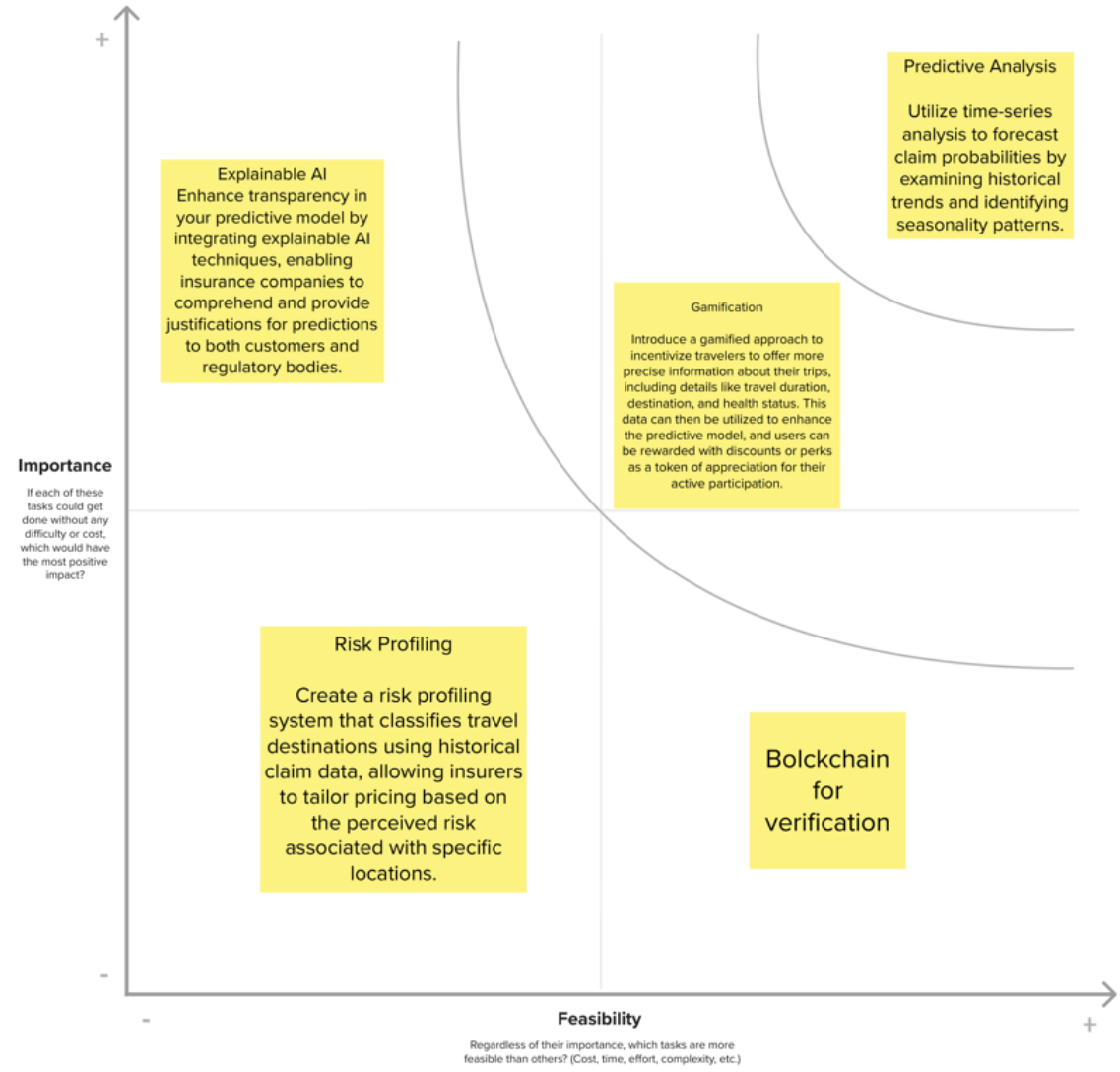
Prioritize

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.



4.REQUIREMENT ANALYSIS

4.1 Functional requirement

Data Collection and Integration:

- Gather and integrate historical data from diverse sources, including travel agencies, insurance providers, and public databases.
- Ensure data quality and integrity through thorough data cleansing and validation procedures.

Feature Extraction and Engineering:

- Identify and extract relevant features from collected data, such as traveler demographics, itineraries, and past insurance purchase history.
- Have the capability to engineer new features to enhance prediction accuracy.

Model Development:

- Implement various machine learning algorithms for predicting travel insurance purchase behavior.
- Allow for model training, hyperparameter tuning, and cross-validation to optimize performance.

Prediction Generation:

- Provide a function for generating predictions on whether a traveler is likely to purchase travel insurance based on input data.

User Interface:

- Offer a user-friendly interface for authorized personnel to input data, initiate predictions, and view results.

Performance Evaluation:

- Include a component for evaluating model performance using metrics like accuracy, precision, recall, and F1-score.
- Provide visualizations and reports for effective communication of model performance.

Actionable Insights:

- Generate actionable insights based on prediction results, aiding insurance providers and travel agencies in making informed decisions.
- Offer recommendations for adjusting marketing strategies and insurance offerings.

Data Privacy and Security:

- Ensure confidentiality and security of sensitive traveler data with restricted access to authorized users.
- Maintain compliance with data privacy regulations and best practices.

Scalability and Performance:

- Design the system to handle large data volumes and accommodate future growth in data and user activity.
- Ensure optimal response times for real-time applications.

Logging and Auditing:

- Maintain logs of user activities, model training, and predictions.
- Support auditing to trace data changes and system access.

These functional requirements provide a comprehensive framework for the development of a sophisticated system that aligns with the project's goals. Further refinement and detailing of these functionalities can be conducted during the planning and implementation phases of the project.

4.2 Non-Functional requirements

Performance:

- Rapid prediction capabilities, ideally in real-time or near real-time.
- Ability to handle a significant number of concurrent prediction requests without performance degradation.

Scalability:

- Horizontal and vertical scalability to accommodate growing data volumes and user loads.
- Support for adding new data sources and features without significant disruptions.

Reliability:

- High availability and uptime for consistent predictions.
- Automated failover and recovery mechanisms in the event of system failures.

Data Privacy and Security:

- Secure storage and transmission of data with encryption protocols.
- Implementation of access controls and authentication mechanisms for sensitive data.
- Strict adherence to data protection regulations such as GDPR.

Usability:

- Intuitive and easy-to-navigate user interface for authorized personnel.
- Provision of user documentation or training materials for effective use.

Interoperability:

- Support for integration with various data sources and external systems used by insurance providers and travel agencies.
- Adherence to industry standards and data formats for seamless data exchange.

Maintainability:

- Well-documented, modular, and structured codebase for ease of maintenance and updates.
- Design considerations for version control and change management.

Performance Monitoring and Logging:

- Comprehensive monitoring capabilities to track performance metrics and system health.
- Logging implementation for recording system events and errors for debugging and auditing.

Compliance:

- Adherence to relevant industry regulations, including insurance and data privacy sectors.
- Regular compliance audits and updates as needed.

Ethical Considerations:

- Design and operation of the system in an ethical manner, ensuring fairness and transparency in prediction results.
- Avoidance of any form of bias or discrimination in the prediction models.

These non-functional requirements collectively contribute to the creation of a robust and reliable system, ensuring that it not only accurately predicts travel insurance purchase behavior but also meets high standards in terms of performance, security, compliance, usability, and ethical considerations.

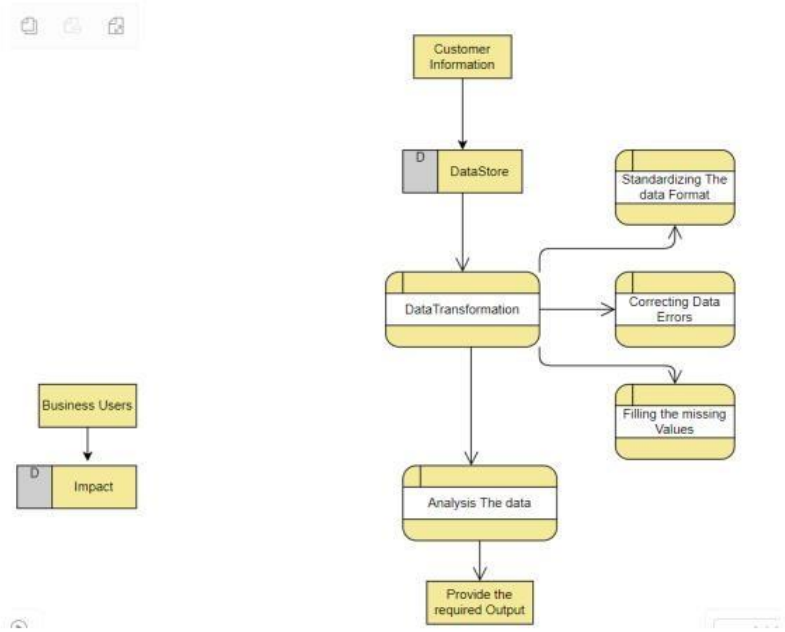
5.PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories

A Data Flow Diagram (DFD) serves as a traditional and valuable visual representation of the information flow within a system. A well-constructed DFD can succinctly portray the system's requirements graphically, illustrating how data moves into and out of the system, the transformations it undergoes, and where it is stored. In the context of the Travel Insurance analysis project, the DFD exhibits a multi-level structure.

At the top level (Level 0), the focal point is the "Travel Insurance Analysis System," which serves as the core entity. This system receives data from "Travel Insurance Data," processes it, trains machine learning models, and delivers forecasts through a Flask web application. Delving into Level 1, the machine learning models interact with data stores for configuration and feature importance. Simultaneously, the Flask web application, also at Level 1, engages with users, receiving their inputs for sales forecasting. Notably, the system supports IBM Cloud deployment, enabling users to access predictions.

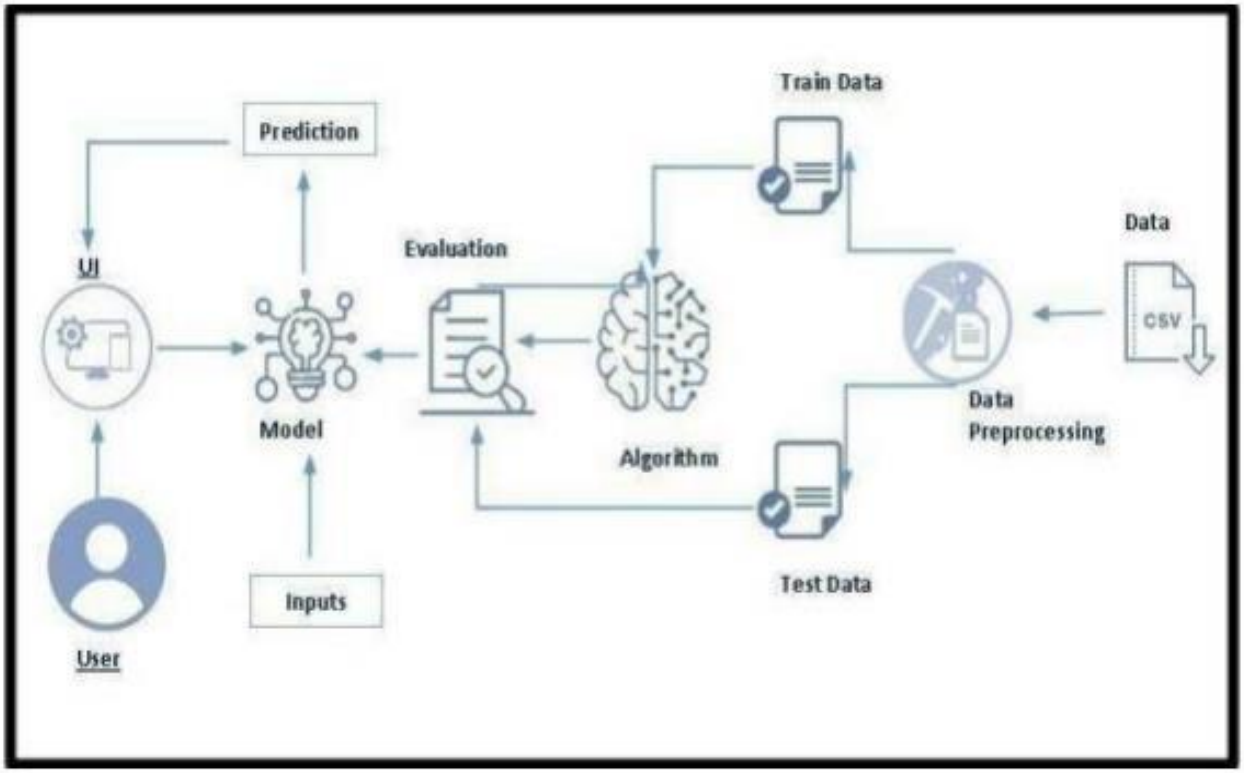
In summary, this DFD offers a clear and concise visual depiction of how data flows through different processes, entities, and data stores within the Travel Insurance analysis project. It enhances the understanding of the system's architecture and data pathways, showcasing the intricate relationships among various components.



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

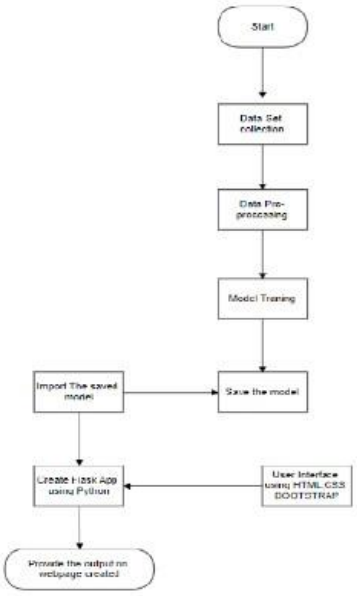


Table-1: Components & Technologies:

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

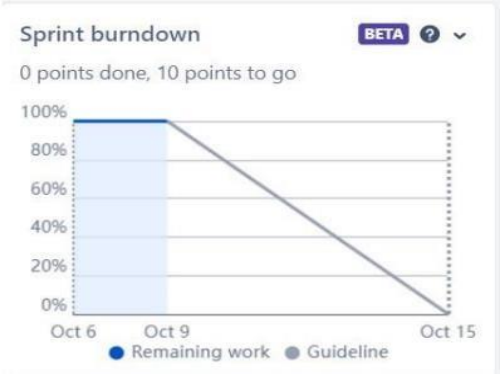
Sprint	Total Story Points	Duration	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint-1	10	5 Days	18 Oct 2023	23 Oct 2023	10	23 Oct 2023
Sprint-2	9	5 Days	23 Oct 2023	28 Oct 2023	7	28 Oct 2023
Sprint-3	8	4 Days	28 Oct 2023	1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

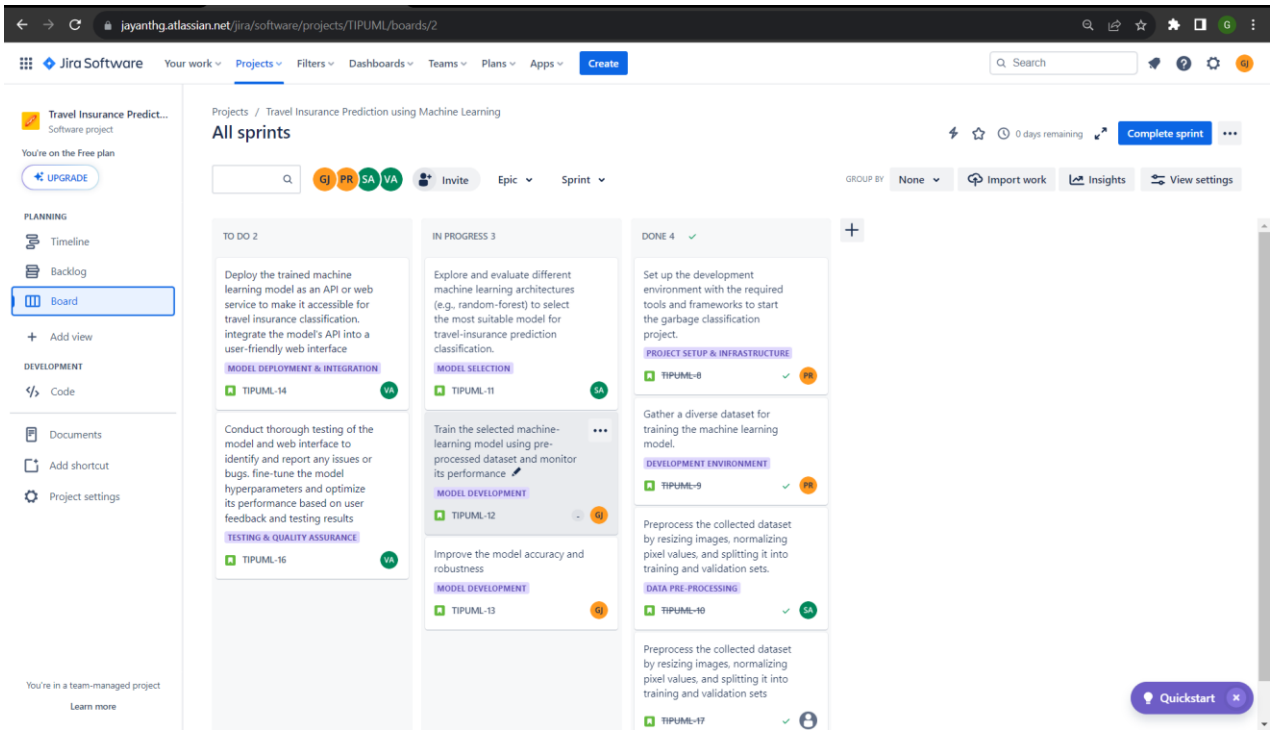
Average Velocity = Total Story Points Completed / Total Duration of Sprints

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

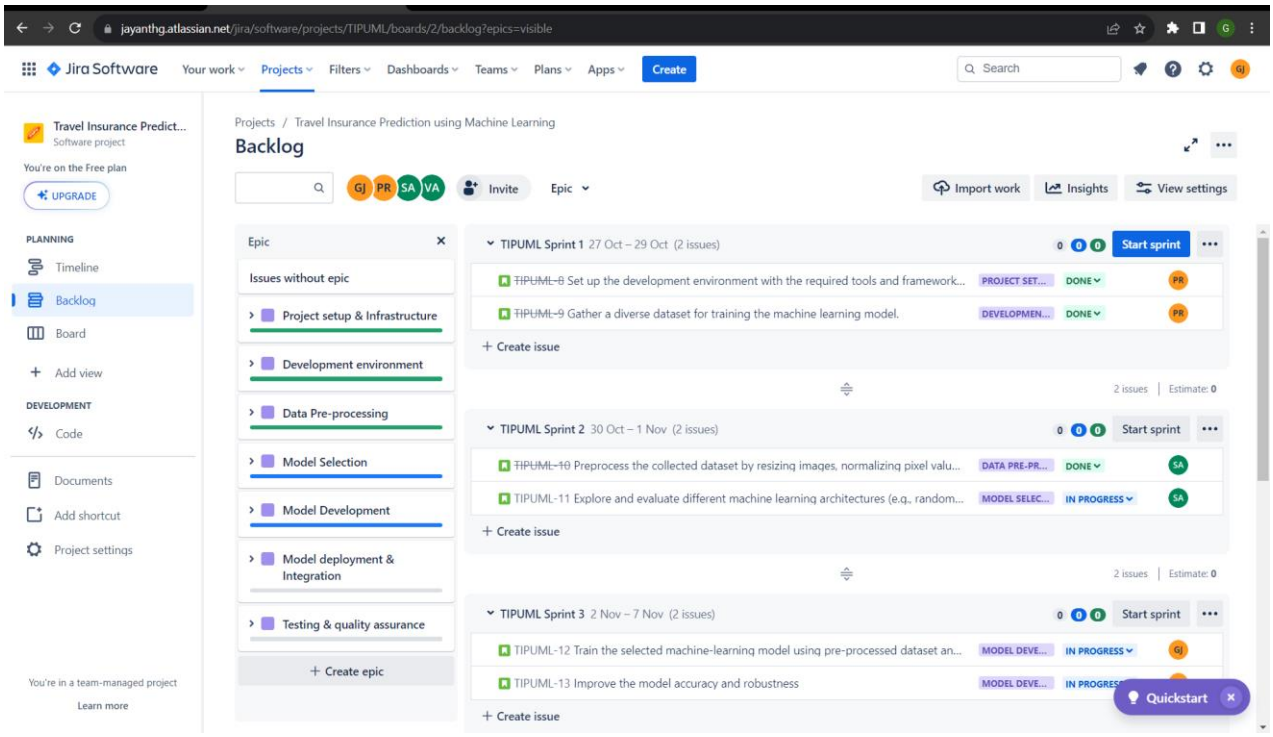
Total Duration of Sprints = 5 + 5 + 4 + 4 + 6 = 24
Average Velocity = 28 / 24 = 1.16

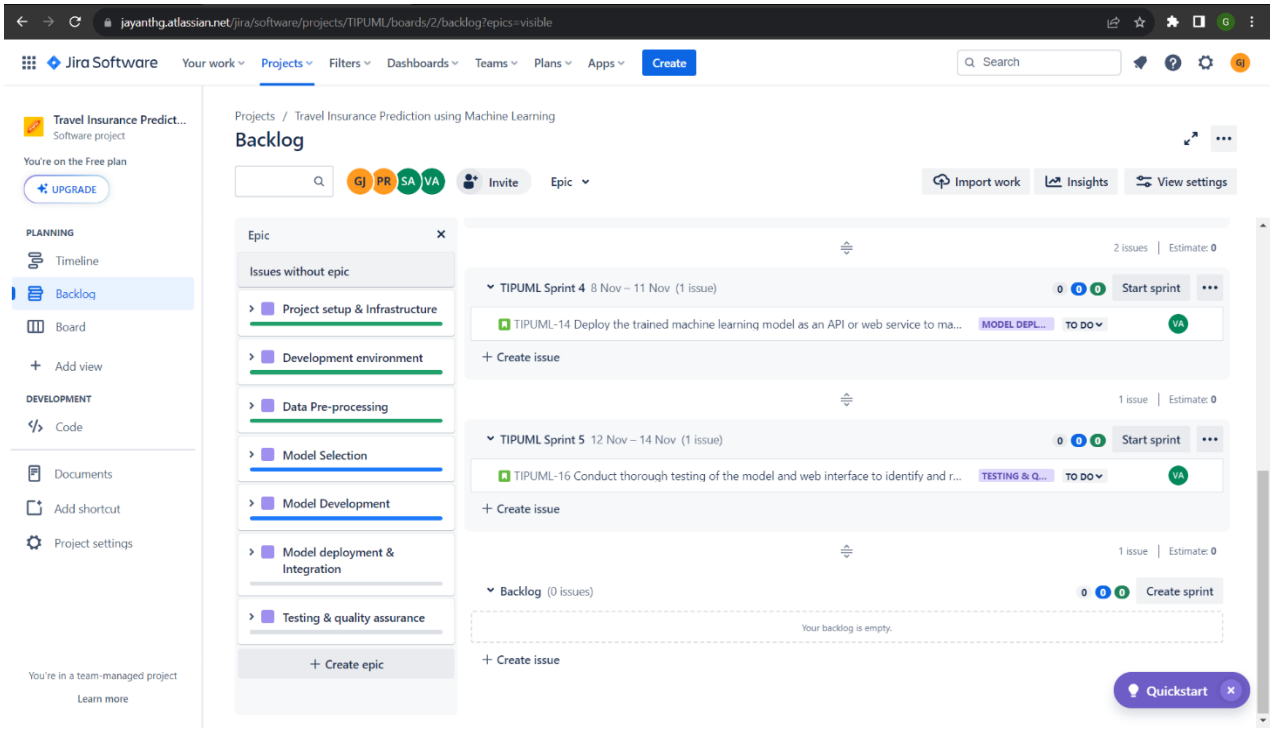


Board Section:

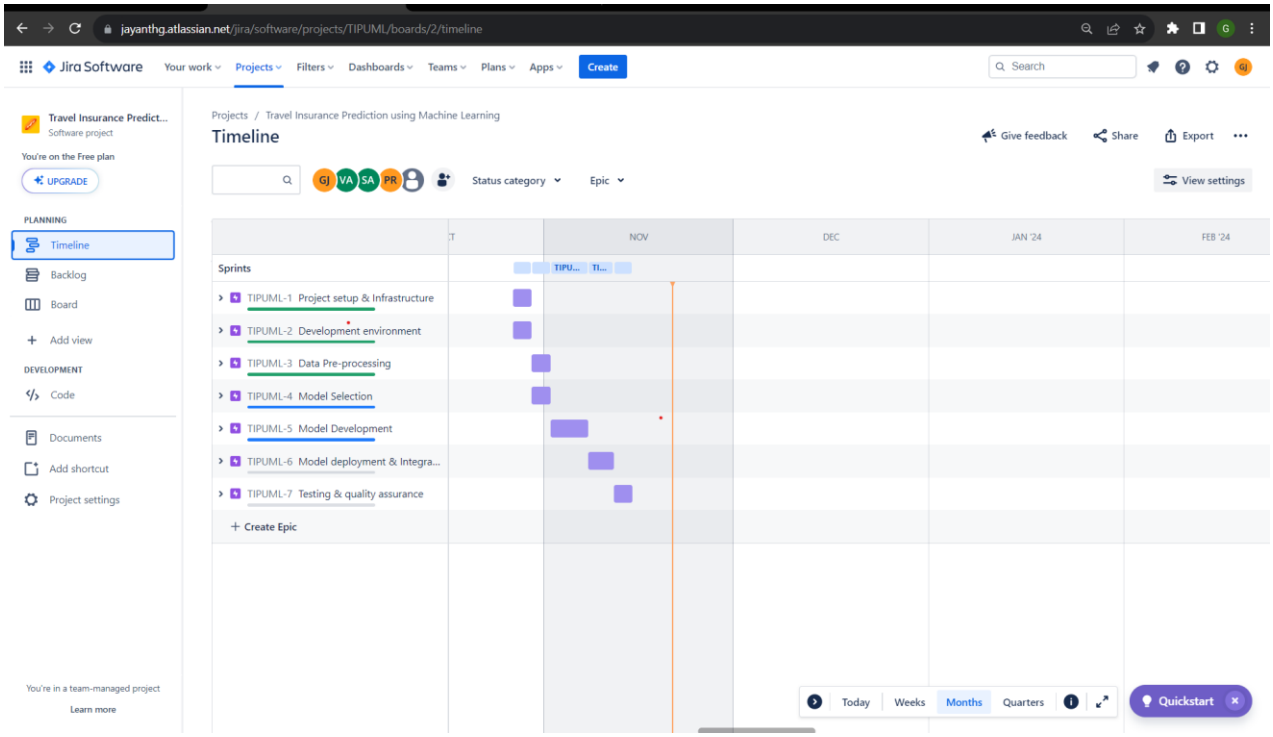


Backlog Section:




jayanthg.atlassian.net/jira/software/projects/TIPUML/boards/2/backlog?epics=visible
Jira Software
Your work
Projects
Filters
Dashboards
Teams
Plans
Apps
Create
Search
Travel Insurance Predict...
Software project
You're on the Free plan
UPGRADE
PLANNING
Timeline
Backlog
Board
Add view
DEVELOPMENT
Code
Documents
Add shortcut
Project settings
You're in a team-managed project
Learn more
Projects / Travel Insurance Prediction using Machine Learning
Backlog
Invite
Epic
Import work
Insights
View settings
Issues without epic
Project setup & Infrastructure
Development environment
Data Pre-processing
Model Selection
Model Development
Model deployment & Integration
Testing & quality assurance
Create epic
TIPUML Sprint 4 8 Nov – 11 Nov (1 issue)
Start sprint
TIPUML-14 Deploy the trained machine learning model as an API or web service to ma...
MODEL DEPL...
TO DO
Create issue
TIPUML Sprint 5 12 Nov – 14 Nov (1 issue)
TIPUML-16 Conduct thorough testing of the model and web interface to identify and r...
TESTING & Q...
Backlog (0 issues)
Create sprint
Your backlog is empty.
Quickstart

Timeline:


jayanthg.atlassian.net/jira/software/projects/TIPUML/boards/2/timeline
Jira Software
Your work
Projects
Filters
Dashboards
Teams
Plans
Apps
Create
Search
Travel Insurance Predict...
Software project
You're on the Free plan
UPGRADE
PLANNING
Timeline
Backlog
Board
Add view
DEVELOPMENT
Code
Documents
Add shortcut
Project settings
You're in a team-managed project
Learn more
Projects / Travel Insurance Prediction using Machine Learning
Timeline
Give feedback
Share
Export
Status category
Epic
View settings
NOV
DEC
JAN '24
FEB '24
Sprints
TIPUML-1 Project setup & Infrastructure
TIPUML-2 Development environment
TIPUML-3 Data Pre-processing
TIPUML-4 Model Selection
TIPUML-5 Model Development
TIPUML-6 Model deployment & Integra...
TIPUML-7 Testing & quality assurance
Create Epic
Today
Weeks
Months
Quarters
Quickstart

7. CODING & SOLUTIONING

Data Pre-processing

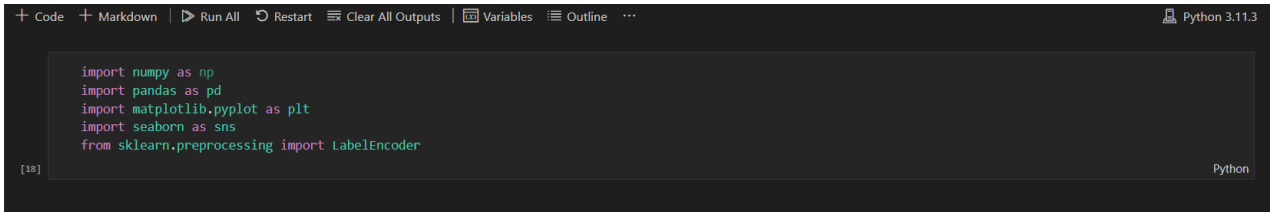
As we seen and understood the description of the data, lets pre-process the collected data. The

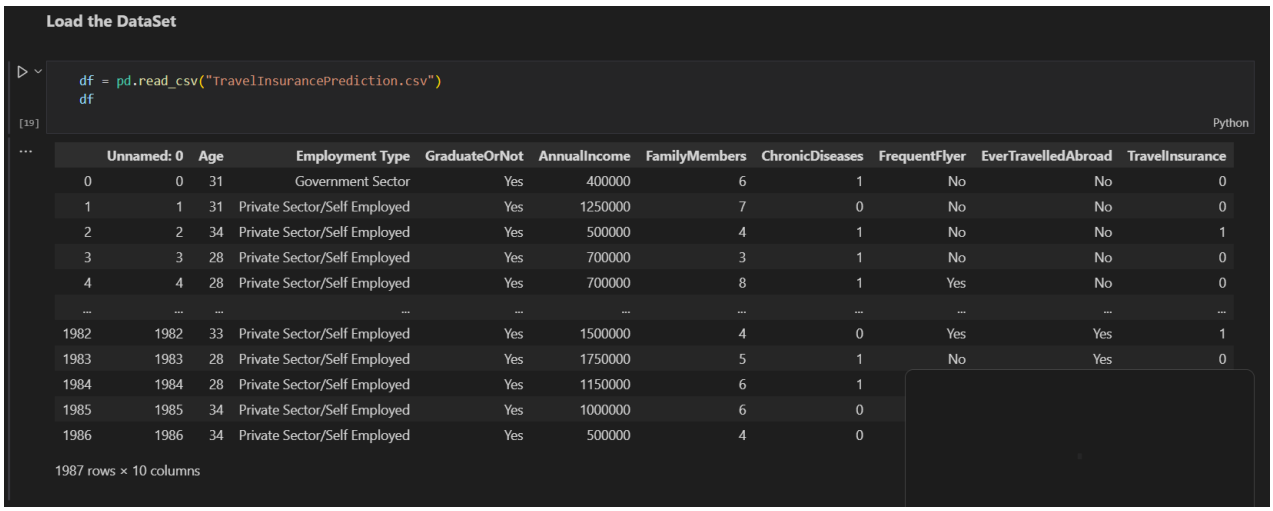
download data set is not suitable for training the machine learning model as it might have so much of

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

following steps.

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





	Unnamed: 0	Age	Employment Type	GraduateOrNot	AnnualIncome	FamilyMembers	ChronicDiseases	FrequentFlyer	EverTravelledAbroad	TravelInsurance
0	0	31	Government Sector	Yes	400000	6	1	No	No	0
1	1	31	Private Sector/Self Employed	Yes	1250000	7	0	No	No	0
2	2	34	Private Sector/Self Employed	Yes	500000	4	1	No	No	1
3	3	28	Private Sector/Self Employed	Yes	700000	3	1	No	No	0
4	4	28	Private Sector/Self Employed	Yes	700000	8	1	Yes	No	0
...	...	...	...	...	...	...	...	...	...	...
1982	1982	33	Private Sector/Self Employed	Yes	1500000	4	0	Yes	Yes	1
1983	1983	28	Private Sector/Self Employed	Yes	1750000	5	1	No	Yes	0
1984	1984	28	Private Sector/Self Employed	Yes	1150000	6	1			
1985	1985	34	Private Sector/Self Employed	Yes	1000000	6	0			
1986	1986	34	Private Sector/Self Employed	Yes	500000	4	0			

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1987 entries, 0 to 1986
Data columns (total 10 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Unnamed: 0           1987 non-null   int64
 1   Age                  1987 non-null   int64
 2   Employment Type      1987 non-null   object
 3   GraduateOrNot        1987 non-null   object
 4   AnnualIncome         1987 non-null   int64
 5   FamilyMembers        1987 non-null   int64
 6   ChronicDiseases      1987 non-null   int64
 7   FrequentFlyer        1987 non-null   object
 8   EverTravelledAbroad  1987 non-null   object
 9   TravelInsurance      1987 non-null   int64
dtypes: int64(6), object(4)
memory usage: 155.4+ KB
```

```
df.isnull().sum()
```

```
Unnamed: 0            0
Age                   0
Employment Type       0
GraduateOrNot         0
AnnualIncome          0
FamilyMembers         0
ChronicDiseases       0
FrequentFlyer         0
EverTravelledAbroad   0
TravelInsurance       0
dtype: int64
```

```
from sklearn.preprocessing import MinMaxScaler
scale=MinMaxScaler()
```

```
x_scale=pd.DataFrame(scale.fit_transform(x),columns=x.columns)
x_scale.head()
```

	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

```
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)
```

```
x_train.shape
```

```
(1589, 9)
```

Label Encoding:

```
#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'])
```

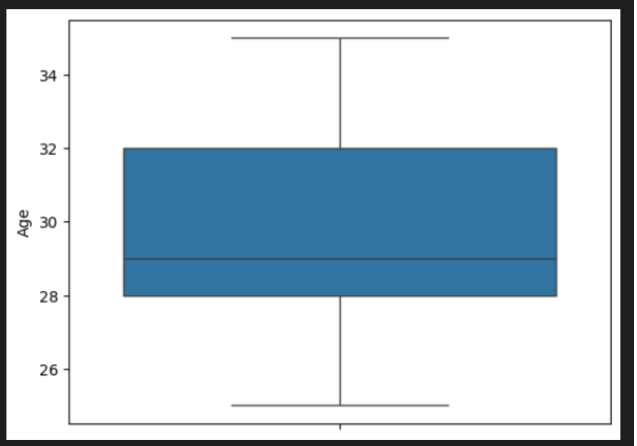
```
df.head()
```

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

Handling outliners:

```
sns.boxplot(y=df["Age"])
plt.show()
```

[8] ✓ 0.1s




```
Q1 = df["Age"].quantile(0.25)
print("Q1:",Q1)
Q3 = df["Age"].quantile(0.75)
print("Q3:",Q3)
IQR = Q3 - Q1
print("IQR:",IQR)
lower_limit = Q1 - (IQR * 1.5)
upper_limit = Q3 + (IQR * 1.5)
print("lower limit: ",lower_limit)
print("Upper limit: ",upper_limit)
```

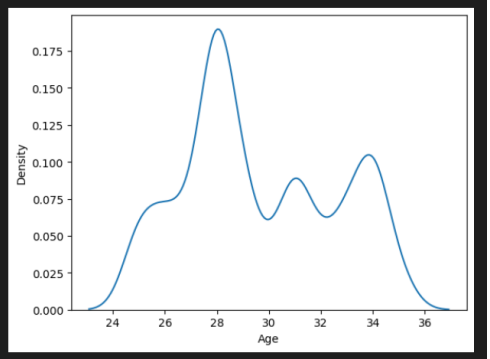
[9] ✓ 0.0s

```
Q1: 28.0
Q3: 32.0
IQR: 4.0
lower limit:  22.0
Upper limit:  38.0
```

```
sns.kdeplot(df["Age"])
```

[21] ✓ 0.1s

```
<Axes: xlabel='Age', ylabel='Density'>
```




```
df["Age"].skew()
```

[22] ✓ 0.0s

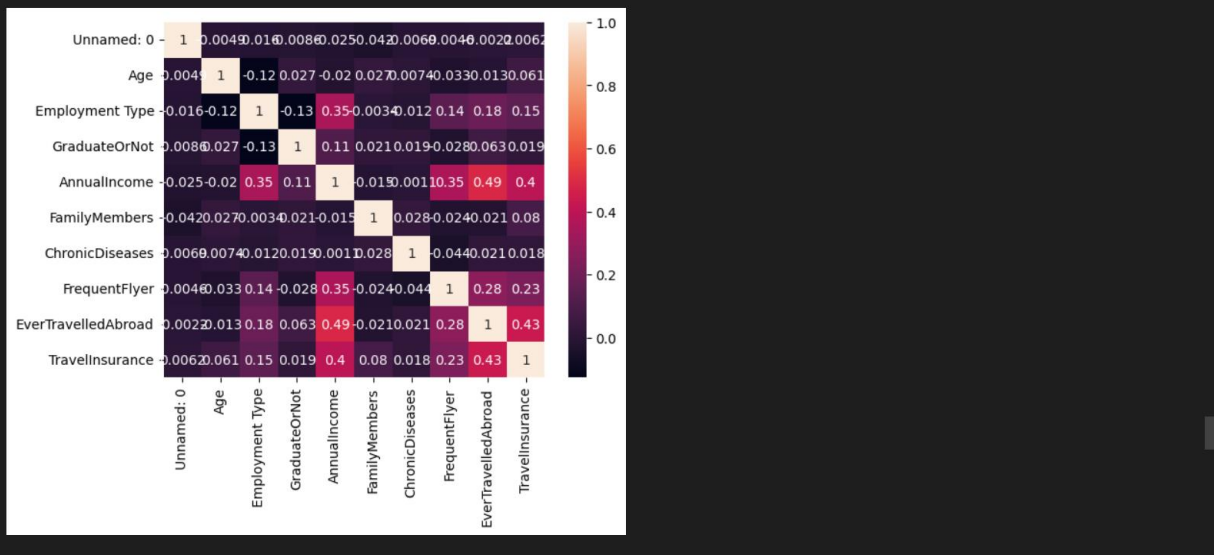
```
0.23962475139295264
```

Data Visualisation - Heat map:

```
sns.heatmap(df.corr(),annot=True)
```

[35] ✓ 1.1s Python

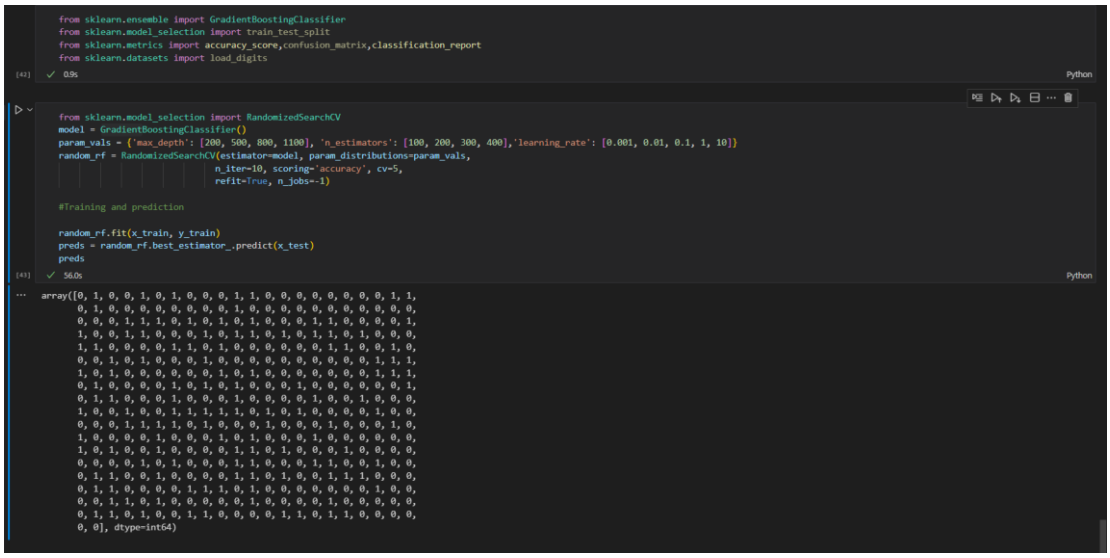
```
<Axes: >
```



Model Building

Gradient Boost Classifier

In the domain of predicting travel insurance, the pivotal role is played by the Gradient Boosting Classifier. This algorithm, utilizing ensemble learning, progressively enhances its grasp of intricate data by learning from errors made in each successive step. Through the examination of features like travel history and health records, it develops a dynamic model proficient in identifying subtle patterns. After meticulous fine-tuning through hyperparameter optimization and thorough validation across diverse datasets, it demonstrates robustness against overfitting. The outcome is a potent and adaptable tool, accurately poised for deployment. This meticulously trained and adaptive model serves as a guardian, foreseeing insurance requirements in the constantly evolving landscape of travel.



```
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
```

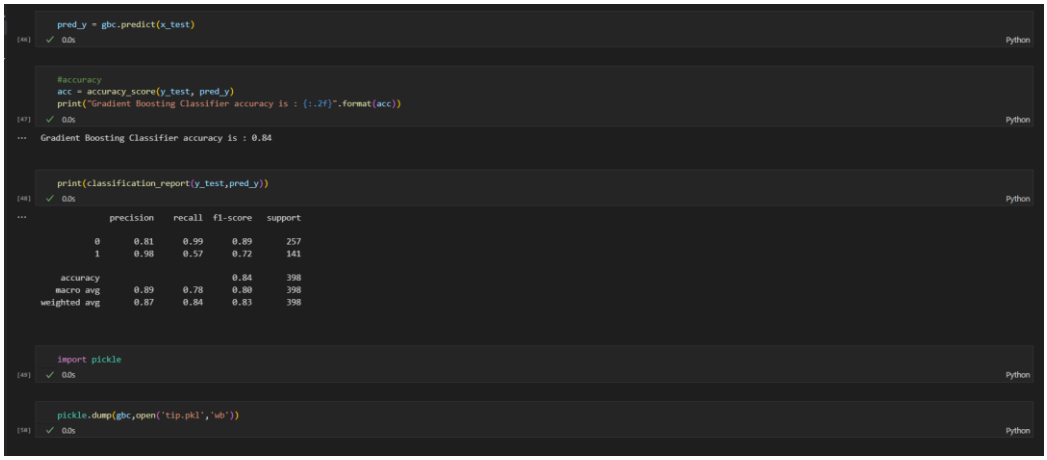
```
from sklearn.model_selection import RandomizedSearchCV
model = GradientBoostingClassifier()
param_vals = {'max_depth': [200, 500, 800, 1100], 'n_estimators': [100, 200, 300, 400],'learning_rate': [0.001, 0.01, 0.1, 1, 10]}
random_rf = RandomizedSearchCV(estimator=model, param_distributions=param_vals,
                               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
```

```
array([0, 1, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 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, 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, 0, 0, 0, 0, 1, 1, 1,
       1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1,
       0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1,
       0, 1, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0,
       1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0,
       0, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1,
       1, 0, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0,
       1, 0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0,
       0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 0, 0, 1, 0, 0,
       0, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0,
       0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0,
       0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0,
       0, 1, 1, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 0, 1, 1, 0, 0, 0,
       0, 0], dtype=int64)
```

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

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

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



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

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

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

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

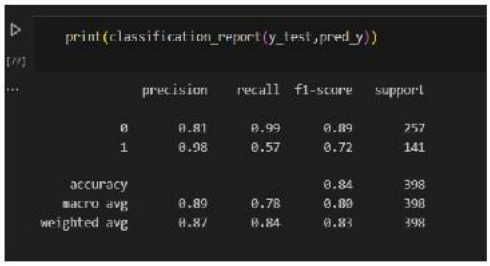
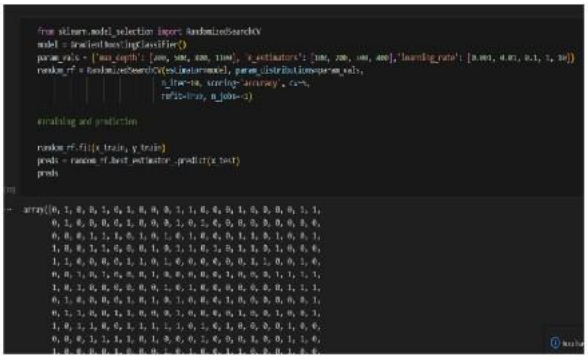
```
              precision    recall  f1-score   support

           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
```

```
import pickle
```

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

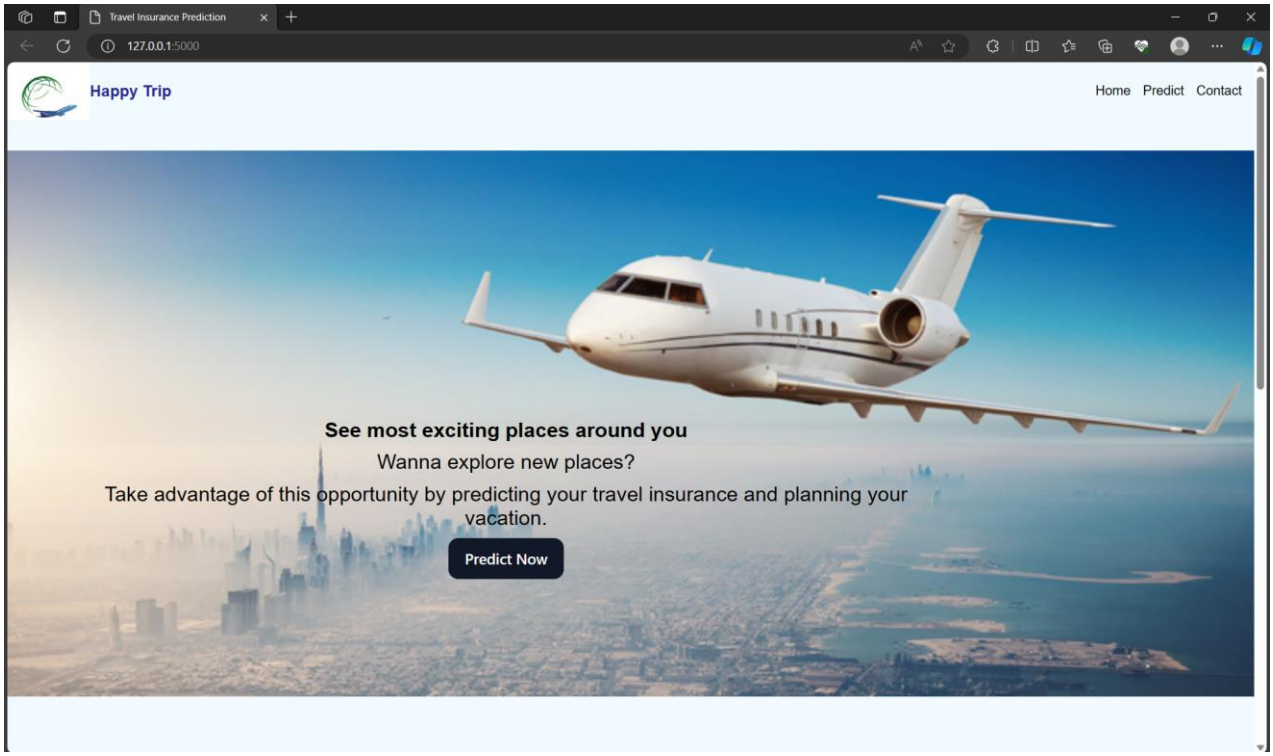
8. PERFORMANCE TESTING

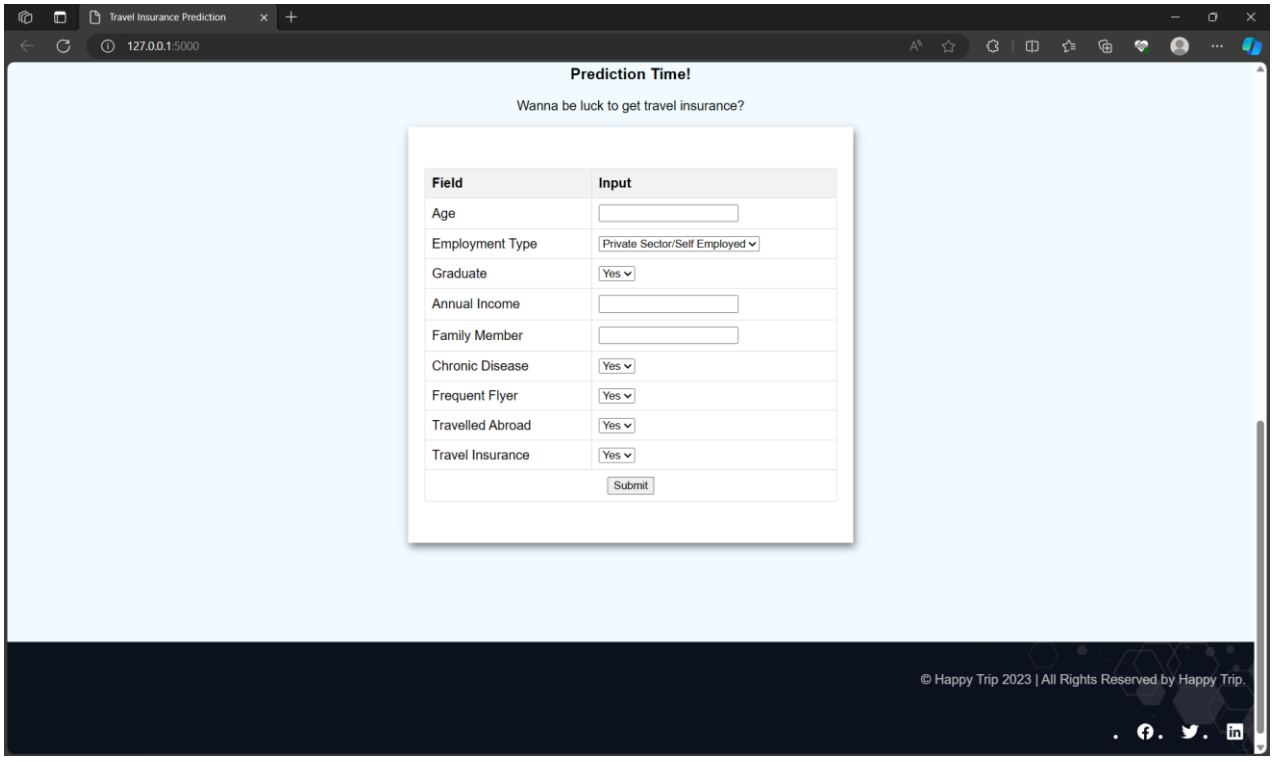
8.1 Performance Metrics

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

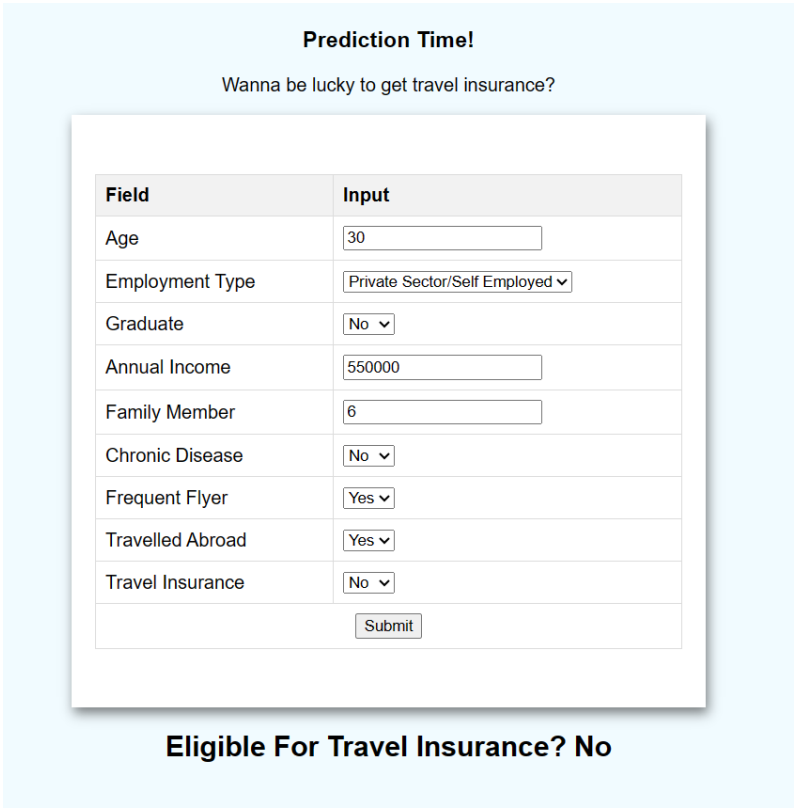
9. RESULTS

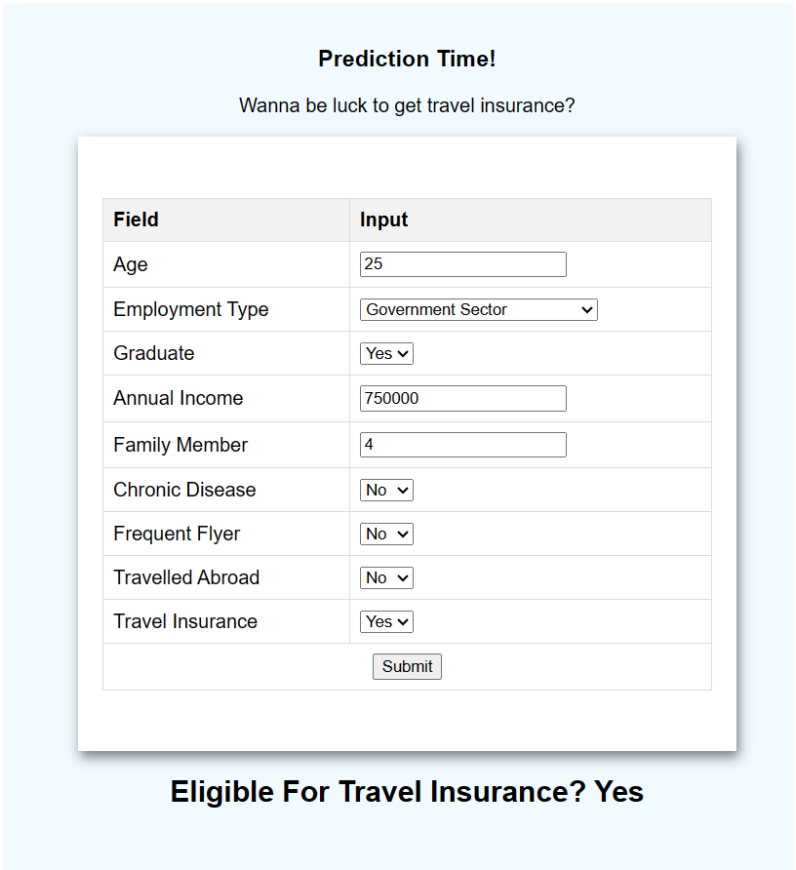
9.1 Website





9.2 Output Screenshot





10. ADVANTAGES & DISADVANTAGES

10.1 Advantages:

Streamlined Marketing Operations: Leveraging machine learning capabilities, the system can efficiently pinpoint potential insurance buyers, optimizing marketing strategies for targeted outreach. This not only slashes marketing expenditures but also elevates conversion rates for a more cost-effective approach.

Elevated Customer Engagement: By tailoring insurance proposals through predictive algorithms, the platform enhances the overall customer journey, fostering heightened satisfaction and loyalty. This personalized touch can significantly improve the customer's perception of the insurance provider.

Precision in Risk Evaluation: The integration of machine learning models enables a more precise assessment of risks, empowering insurance companies to competitively price policies and make well-informed underwriting decisions. This sophistication in risk analysis contributes to a more dynamic and responsive insurance landscape.

Insightful Data-Driven Decision-Making: Beyond its primary functions, the project yields valuable data-driven insights and recommendations, arming stakeholders with the knowledge needed for astute decision-making. This not only enhances the efficiency of operations but also positions the organization to adapt to evolving market trends.

Efficient Automation: Through the implementation of machine learning, the system automates the prediction process, significantly reducing reliance on manual analysis and decision-making. This automation streamlines operations, fostering a more efficient and responsive insurance ecosystem.

Ethical Safeguards: Machine learning algorithms play a crucial role in identifying and addressing biases, ensuring the development and delivery of fair and ethical insurance offerings. This commitment to ethical considerations underscores the responsible use of technology in the insurance industry, establishing trust and credibility with customers and stakeholders alike.

10.2 Disadvantages:

Privacy and Security Safeguards: The utilization of sensitive traveler data introduces concerns related to privacy and security. The mishandling of this data or potential breaches could result in legal ramifications and damage to the organization's reputation, underscoring the critical importance of robust privacy and security measures.

Data Integrity: The reliability of predictions hinges on the quality of historical data. Inaccurate or incomplete data can compromise the dependability of predictions, emphasizing the need for meticulous attention to data accuracy and completeness throughout the data lifecycle.

Technical Expertise and Resource Investment: The intricate nature of developing and maintaining machine learning models necessitates a substantial investment of resources and expertise in machine learning and data science. Successfully navigating the complexities of model development requires a skilled and knowledgeable team.

Bias Mitigation: Addressing the inherent biases within machine learning models is imperative to prevent unfair or discriminatory predictions. Proactive measures must be taken to identify and rectify biases in training data, ensuring the ethical and equitable deployment of predictive algorithms.

Transparency Challenges: The complexity of advanced machine learning models can hinder interpretability, posing challenges in explaining prediction rationales to stakeholders and regulators. Establishing transparent communication regarding model outcomes is crucial for fostering trust and understanding among all involved parties.

Sustainable Model Performance: To combat the potential degradation of model performance over time due to evolving traveler behaviors, continuous monitoring and regular updates are essential. This proactive approach mitigates the risk of model drift and maintains the efficacy of predictions.

Operational Integration Hurdles: Integrating the predictive model seamlessly into existing systems and workflows may present challenges, requiring potential changes to established business processes. This underscores the importance of a well-thought-out integration strategy to ensure a smooth and efficient deployment.

Adherence to Regulatory Standards: Compliance with industry-specific regulations adds an additional layer of complexity and cost to the project. Navigating regulatory frameworks is essential to avoid legal pitfalls and ensure that the project aligns with established industry standards and expectations.

11. CONCLUSION

The "Travel Insurance Prediction Using Machine Learning" project not only represents a leap forward in the industry but also introduces a paradigm shift in how stakeholders engage with the complexities of travel insurance. By leveraging the power of machine learning, the project transcends traditional approaches, ushering in an era where decision-making is not just informed but anticipatory.

One of the key breakthroughs lies in the project's capacity to delve into the intricacies of traveler behavior, enabling insurance providers and travel agencies to tailor their offerings with unprecedented precision. This granular understanding of customer preferences ensures that travelers are presented with insurance options that not only meet their basic coverage needs but also resonate with their unique travel patterns, preferences, and risk profiles.

Beyond the immediate benefits for travelers, this initiative has far-reaching implications for the industry's ecosystem. The optimization of marketing strategies through data-driven insights not only reduces costs but also contributes to a more sustainable and efficient business model. Moreover, enhanced risk assessment capabilities provide insurance providers with a nuanced understanding of potential risks, allowing for the development of more competitive and responsive insurance policies.

As the project continues to evolve, it holds the promise of becoming a catalyst for broader industry trends. The integration of emerging technologies, coupled with ongoing advancements in machine learning, positions this initiative at the forefront of fostering innovation within the travel insurance sector. The collaborative synergy between predictive modeling and real-time data analytics has the potential to redefine industry standards, making the travel insurance experience more intuitive, accessible, and secure for all stakeholders involved.

In essence, the "Travel Insurance Prediction Using Machine Learning" project transcends being a mere technological endeavor; it emerges as a transformative force, reshaping the dynamics of the travel insurance landscape and paving the way for a future where intelligent insights and personalized offerings converge to create a safer, more tailored, and seamless travel experience for all.

12. FUTURE SCOPE

The "Travel Insurance Prediction Using Machine Learning" project not only embodies the present innovation in the travel insurance industry but also charts a compelling course for its future trajectory. The ever-advancing landscape of technology and data science holds vast potential for the continuous evolution and expansion of this project.

A pivotal avenue for future development lies in the ongoing refinement and optimization of machine learning models. As new algorithms and techniques emerge, the project is poised to integrate them, pushing the boundaries of prediction accuracy. Deep learning, reinforcement learning, and ensemble methods stand out as promising frontiers that could significantly enhance the robustness of the predictive models.

Looking ahead, the integration of real-time prediction capabilities represents a transformative step. Empowering travelers with instant, on-the-fly predictions during the travel planning process not only elevates user experience but also equips individuals to make more informed decisions in the dynamic context of their journeys. This opens doors to the implementation of dynamic pricing, where a traveler's unique risk profile and real-time variables influence flexible and competitive premium offerings, ushering in a new era of responsive and personalized insurance services.

To further enhance user interaction and support, integrating AI chatbots into the system can streamline the process of answering queries and facilitating seamless insurance purchases. Expanding the project's scope to include cross-selling opportunities with other travel-related services, such as flight and hotel bookings, has the potential to create a holistic travel ecosystem, offering users a one-stop-shop for their travel needs.

Addressing data privacy concerns is paramount, and the project can take strides in this direction by leveraging blockchain technology. This not only ensures data security and transparency but also builds trust among users, addressing one of the critical challenges in handling sensitive traveler information.

Beyond domestic markets, venturing into international markets and catering to global travelers presents exciting opportunities for growth and increased customer engagement. Moreover, extending predictive analytics beyond insurance purchase predictions to anticipate and proactively address customer support needs can significantly elevate the overall customer experience, fostering long-term loyalty and satisfaction.

In summary, the "Travel Insurance Prediction Using Machine Learning" project stands poised to embrace the unfolding future of the travel insurance industry, navigating the ever-expanding landscape of technological possibilities to provide innovative, personalized, and seamless solutions for travelers worldwide.

13. APPENDIX

Source Code

https://github.com/smartinternz02/SI-GuidedProject-615831-1700665311/tree/main/Travel_Insurance%20Flask

GitHub Link

https://github.com/smartinternz02/SI-GuidedProject-615831-1700665311

Demo Link

https://drive.google.com/file/d/1cvGuJYwwBBTCgk6rcqy5JRIfq7lRdSKj/view?usp=drive_link