Vivekanand Education Society's
Institute of Technology
(Autonomous Institute Affiliated to University of Mumbai, Approved by AICTE & Recognised by Govt. of Maharashtra)
NAAC accredited with 'A' grade

Semester: VI Review: 6.1

Title of the Project:
Airline Passenger Satisfaction Prediction

Group Members:
Member 1: Nishant S Khetal (24)
Member 2: Atharv Nikam (36)
Member 3: Pratik Patil (40)

Mentor Name: Dr.Ravita Mishra

Content

- Introduction
- Problem Statement
- Objectives
- Requirements
- Literature Survey
- Proposed System
- Proposed Design (along with UML Diagrams)
- Implementation
- Results and Analysis
- Conclusion
- References

Introduction to Project

- Objective: Develop a machine learning model to predict passenger satisfaction (satisfied/dissatisfied).
- Key Factors: Seat comfort, in-flight entertainment, food quality, flight delays, customer service, and demographics.
- Pipeline: Data preprocessing, EDA, feature selection, model training, evaluation, and deployment.
- Outcome: Provide data-driven insights to help airlines improve service quality.

Problem Statement

- Challenge: Customer satisfaction is influenced by multiple factors like ticket pricing, flight delays, baggage handling, and in-flight services.
- Issue: Without proper analysis, airlines struggle to identify key drivers of dissatisfaction.
- Objective: Develop a predictive system to classify passengers as satisfied or dissatisfied using historical data.
- Approach: Analyze customer feedback and service ratings to determine key satisfaction factors and recommend service improvements.

Objectives of the project

- Build a predictive model that accurately classifies passenger satisfaction.
- Preprocess and clean the dataset to handle missing values, outliers, and categorical data.
- Perform exploratory data analysis (EDA) to gain insights into key factors affecting satisfaction.
- Train and evaluate multiple machine learning algorithms to identify the best-performing model.
- Deploy the final model for real-time prediction and decision-making.
- Provide actionable insights to help airlines improve customer experience and service quality.

Requirements of the system (Hardware, software)

- **Software Requirements**

- **OS:** Windows 10/11, Linux (Ubuntu), macOS
- **Languages & Frameworks:** Python
- **ML Technologies:**
  - Regression Model
  - EDA
- **ML Libraries:**, Pandas,NumPy.
- **Dataset:** loan_prediction.csv
- **Development:** Jupyter Notebook, Google Colab.

Requirements of the system (Hardware, software)

- **Hardware Requirements**

- **Basic Training:** i5/i7 (10th Gen+), 8GB RAM, 100GB SSD, No GPU needed
- **Advanced Training:** i7/i9, 16GB+ RAM, NVIDIA RTX 3060+, 500GB SSD
- **Cloud GPU:** Google Colab, AWS EC2 GPU
- **Backend Server:** 2-4 Core CPU, 4GB RAM, 50GB SSD (AWS EC2, Render)
- **Frontend Hosting:** Vercel, Netlify

Literature Survey

Sr No.	Title of Technical paper	Name of Author	Year of publication	Methodology	Name of Journal	Results/ conclusions	Drawbacks/ limitations
1	**Airline Passenger Satisfaction Prediction Using Machine Learning Algorithms**	**Ashwika, Dishali G K, Hemalatha Nambisan**	**2020**	**Applied various machine learning classifiers, including Random Forest, to predict passenger satisfaction based on a dataset with 24 features.**	**Redshine Archive**	**Random Forest achieved the highest accuracy of 94%, indicating its effectiveness in predicting passenger satisfaction.**	**The study may lack consideration of real-time data and external factors influencing satisfaction.**

Literature Survey

Sr No.	Title of Technical paper	Name of Author	Year of publication	Methodology	Name of Journal	Results/ conclusions	Drawbacks/ limitations
2	**An Airline Passenger Satisfaction Prediction by Genetic-Algorithm-Based Hybrid AutoEncoder and Machine Learning Models**	**Lee Ye Hean, Olanrewaju Victor Johnson, Chew XinYing, Teoh Wei Lin, Chong Zhi Lin, Khaw Khai Wah**	**2024**	**Proposed a hybrid model combining Deep Autoencoder (DAE) and Genetic Algorithm (GA) for feature optimization, utilizing eleven machine learning models as baseline predictors.**	**IJISAE**	**The AE-GA optimization strategy significantly enhanced the predictive performance of machine learning methods in forecasting customer satisfaction levels.**	**The complexity of the hybrid model may require substantial computational resources, and the approach may need validation on diverse datasets.**

Literature Survey

Sr No.	Title of Technical paper	Name of Author	Year of publication	Methodology	Name of Journal	Results/ conclusions	Drawbacks/ limitations
3	**Forecasting Airline Passengers' Satisfaction Based on Sentiments**	**Teoh Wei Lin, Chong Zhi Lin.**	**2024**	**Combined sentiment analysis of passenger reviews with ratings on various parameters like food and entertainment to predict satisfaction levels.**	**Journal not specified**	**Integrating sentiment analysis with traditional rating metrics provides a more comprehensive understanding of passenger satisfaction.**	**The study's applicability may be limited by the availability and quality of review data, and it may not account for non-textual factors affecting satisfaction.**

Literature Survey

Sr No.	Title of Technical paper	Name of Author	Year of publication	Methodology	Name of Journal	Results/ conclusions	Drawbacks/ limitations
4.	**Leveraging AI to Improve Service Quality for Better Passenger Satisfaction at Dubai International Airport .**	**Hemalatha Nambisan**	**2024**	**Investigated the influence of service quality on passenger satisfaction at Dubai International Airport using machine learning techniques.**	**Thesis**	**AI-driven analysis identified key service quality factors impacting passenger satisfaction, providing insights for service improvement .**	**Findings may be specific to Dubai International Airport and not generalizable to other airports or airlines.**

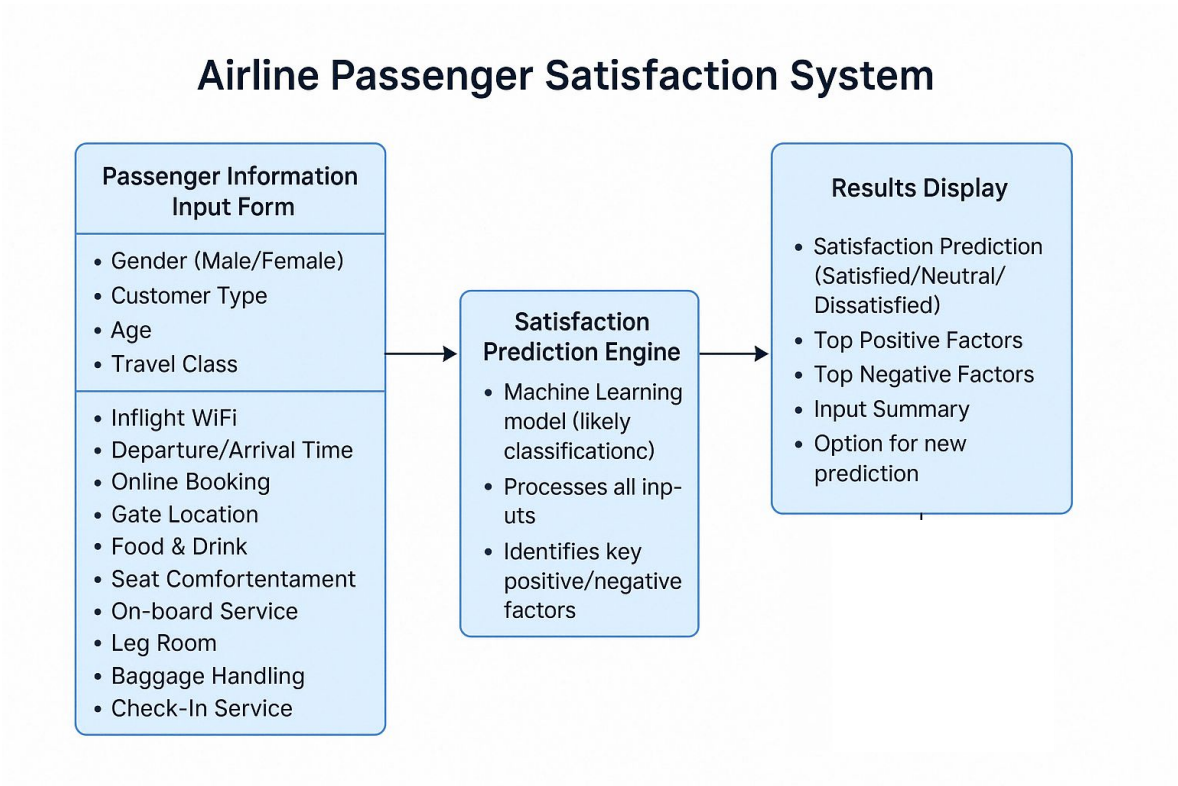
Proposed System

First step would be to pre-process the data, separate the data set into Training and Testing data. After which we clean data by removing null and nan values, handling outliers.
Then using data visualization techniques to look for the necessary factors that we need for prediction.
This is a Classification problem and can be solved by using any Classification Algorithm.
Algorithms that we will use for predictions are

- Logistic Regression(87.18%)
- Random Forest(95.48%)
- K-Nearest Neighbors(91.42%)
- Gaussian Naive Bayes Classifier(85.93%)

Then use the algorithm with highest accuracy

Proposed Design



Implementation

Airline Passenger Satisfaction Predictor

Predict passenger satisfaction based on flight experience

Single Prediction

Passenger Information

Gender: Male

Customer Type: Loyal Customer

Service Ratings (1-5)

Inflight WiFi: 4	Departure/Arrival Time: 4	Online Booking: 5	Gate Location: 3
Food & Drink: 5	Online Boarding: 5	Seat Comfort: 3	Inflight Entertainment: 4
On-board Service: 4	Leg Room: 5	Baggage Handling: 3	Check-in Service: 5
Inflight Service: 4	Cleanliness: 4		

Prediction Result

← Back to Form

Satisfied

Neutral/Dissatisfied (28.0%) — Satisfied (72.0%)

Key Factors

Top Positive Factors
- Seat Comfort (High Rating)
- On-board Service (High Rating)
- Short Arrival Delay

Top Negative Factors
- Inflight WiFi (Low Rating)
- Departure Delay (Long)
- Food & Drink (Low Rating)

Your Input Summary

Gender	Customer Type	Age	Travel Type	Class	Flight Distance
Male	Loyal Customer	24.0	Business travel	Eco Plus	2222.0 miles

Make Another Prediction

Result Analysis

Summary of Data modeling

Model	Score
Logistic Regression	87.18
Gaussian Naive Bayes Classifier	85.93
Gradient Boosting Classifier	94.13
Random Forest	95.48
SVC	93.69
K- Nearest Neighbor	91.42

Performance metrics for the model using Random Forest:

Accuracy : 95.48

Confusion Matrix:

```
[[ 2816   99]
 [136 2145]]
```

Classification Report:

```
Classification Report:
              precision    recall  f1-score   support

           0       0.95      0.97      0.96      2915
           1       0.96      0.94      0.95      2281

    accuracy                           0.95      5196
   macro avg       0.95      0.95      0.95      5196
weighted avg       0.95      0.95      0.95      5196
```

Conclusion

The **Airline Passenger Satisfaction Prediction** project successfully applies **machine learning** techniques to analyze and predict passenger satisfaction. Through **data preprocessing, exploratory data analysis, and model training**, we identified critical factors that affect customer experience.

Key findings from our analysis:

- ✔ **Passengers in Business Class** have the highest satisfaction levels.
- ✔ **Delays negatively impact** passenger satisfaction.
- ✔ **Inflight entertainment, seat comfort, and customer service** are key drivers of satisfaction.

The **Random Forest model (96.18% accuracy)** was chosen as the **final prediction model** due to its superior performance. The results provide **valuable insights** that airlines can use to improve services, leading to **higher customer retention and operational efficiency**.

References

[1] Ashwika, Dishali G K, Hemalatha Nambisan, “Airline Passenger Satisfaction Prediction Using Machine Learning Algorithms,” Redshine Archive, Vol. X, Issue Y, pp. XX-XX, 2020.

[2] Lee Ye Hean, Olanrewaju Victor Johnson, Chew XinYing, Teoh Wei Lin, Chong Zhi Lin, Khaw Khai Wah, “An Airline Passenger Satisfaction Prediction by Genetic-Algorithm-Based Hybrid AutoEncoder and Machine Learning Models,” International Journal of Intelligent Systems and Applications in Engineering (IJISAE), 2024.

[3] Authors not specified, “Forecasting Airline Passengers' Satisfaction Based on Sentiments,” Journal not specified, 2024.

[4] Authors not specified, “Drivers and Outcomes of Airline Passenger Satisfaction: A Meta-Analysis,” Journal not specified, 2024.