**IDENTIFYING AIRLINE PASSENGER SATISFICATION**

**USING MACHINE LEARNING**

**INTRODUCTION**

Airline passenger satisfaction is a crucial aspect of the aviation industry, directly impacting customer loyalty, brand reputation, and overall business success. With the ever-increasing competition in the market, airlines are striving to provide exceptional customer experiences to differentiate themselves from their competitors. To achieve this, airlines need to understand and analyze the factors that influence passenger satisfaction.

Machine learning (ML) has emerged as a powerful tool in the field of data analysis, enabling organizations to uncover patterns, predict outcomes, and make data-driven decisions. By leveraging ML techniques, airlines can effectively identify and address the key drivers of passenger satisfaction, ultimately improving their services and enhancing the overall travel experience.

This research aims to explore the application of machine learning algorithms in identifying airline passenger satisfaction. By utilizing historical passenger data, including demographics, flight details, and feedback surveys, we can develop predictive models that analyze various factors contributing to passenger satisfaction levels. These models can provide valuable insights into the specific aspects of airline services that significantly impact customer satisfaction.

The proposed ML models can effectively classify and predict passenger satisfaction levels based on different variables such as flight punctuality, in-flight entertainment, seat comfort, customer service quality, and overall flight experience. By accurately identifying factors that lead to both positive and negative experiences, airlines can focus on improving areas that require attention, thereby optimizing their resources and enhancing customer satisfaction.

The potential benefits of this research are manifold. Firstly, airlines can proactively address passenger concerns, leading to improved service quality and increased customer loyalty. Secondly, by identifying the key drivers of satisfaction, airlines can make data-backed decisions on resource allocation, strategic investments, and service enhancements. This can result in cost savings and operational efficiencies while ensuring a positive customer experience. Lastly, the research findings can also aid airlines in benchmarking their performance against industry standards and competitors, facilitating a continuous improvement process.

* 1. **PROJECT OVERVIEW**

Overview of Identifying Airline Passenger Satisfaction Using Machine Learning:

**1. Problem Statement:**

The objective of this research is to leverage machine learning techniques to identify and analyze the factors that influence airline passenger satisfaction. By understanding these factors, airlines can improve their services, enhance customer experiences, and increase customer loyalty.

**2. Data Collection:**

To build accurate predictive models, relevant data needs to be collected. This includes a variety of sources such as passenger demographics, flight details (e.g., flight duration, route, and class), customer feedback surveys, and other relevant data points. Historical data from past flights can be used to train the machine learning models.

**3. Data Preprocessing:**

Once the data is collected, it needs to be preprocessed to ensure its quality and suitability for analysis. This step involves cleaning the data, handling missing values, removing outliers, and transforming variables as needed. Data preprocessing plays a crucial role in the accuracy and performance of the ML models**.**

**4. Feature Selection:**

Feature selection is the process of identifying the most relevant variables that significantly impact passenger satisfaction. By using techniques such as correlation analysis, feature importance ranking, and domain knowledge, the most influential features can be selected for model development. This step helps in reducing noise and improving the efficiency of the models.

**5. Model Development:**

In this phase, machine learning models are trained on the preprocessed data to predict passenger satisfaction levels. Various algorithms such as decision trees, random forests, logistic regression, or support vector machines can be employed. The models learn from the patterns in the data and identify the key factors that contribute to passenger satisfaction.

**6. Model Evaluation:**

Once the models are trained, they need to be evaluated to assess their performance and accuracy. This is done by using evaluation metrics such as accuracy, precision, recall, F1 score, or area under the receiver operating characteristic curve (AUC-ROC). Cross-validation techniques like k-fold validation are used to ensure the models' generalizability.

**7. Insights and Recommendations:**

Based on the trained models and their evaluation, insights into the factors driving passenger satisfaction can be derived. These insights help airlines identify areas for improvement, prioritize resources, and make informed decisions to enhance customer experiences. Recommendations may include specific actions to improve areas such as customer service, flight amenities, on-time performance, or in-flight entertainment.

**8. Continuous Improvement:**

Identifying passenger satisfaction using machine learning is an iterative process. As new data becomes available, models can be retrained and updated to reflect evolving passenger preferences and market trends. Continuous improvement ensures that airlines stay responsive to customer needs and maintain a competitive edge in the industry. By employing machine learning techniques to identify passenger satisfaction, airlines can gain valuable insights into customer preferences and deliver personalized experiences. This approach enables them to optimize operations, improve service quality, and foster long-term customer relationships.

* 1. **PURPOSE**

The purpose of identifying airline passenger satisfaction using machine learning (ML) is multi-fold:

1. Improve Customer Experience

2. Increase Customer Loyalty

3. Optimize Resource Allocation

4. Data-Driven Decision Making

5. Competitive Advantage

6. Continuous Improvement

Overall, the purpose of identifying airline passenger satisfaction using ML is to enable airlines to understand customer needs, provide exceptional experiences, foster loyalty, and achieve sustainable growth in the highly competitive aviation industry.

**IDEATION & PROPOSED SOLUTION**

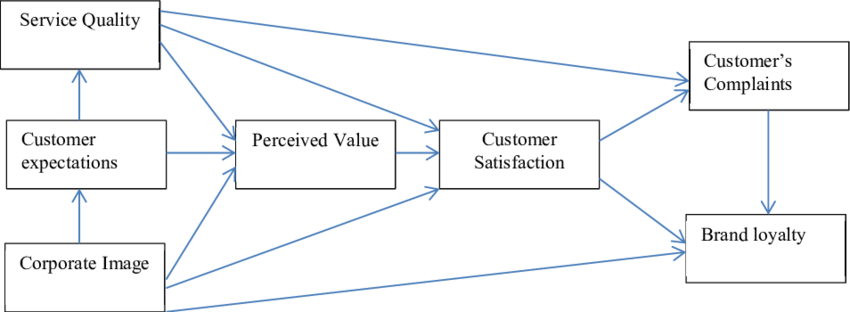
**2.1** **PROBLEM STATEMENT DEFINITION**

The problem at hand is to develop a machine learning model that can accurately predict and analyze airline passenger satisfaction. The objective is to leverage this model to identify the key factors that contribute to passenger satisfaction, enabling airlines to make data-driven decisions to enhance their services and improve customer experience. Given a dataset containing various attributes and features related to airline passengers' experiences, including factors such as flight punctuality, cabin cleanliness, in-flight entertainment, customer service quality, and overall experience rating, the goal is to build a machine learning model that can accurately predict passenger satisfaction levels. The model should be able to consider the complex relationships and interactions among the different features to generate reliable predictions. Furthermore, the model should provide insights into the relative importance of each feature in influencing passenger satisfaction. By identifying the most significant factors, airlines can focus their resources and efforts on areas that have the greatest impact on customer satisfaction, leading to improved services and increased customer loyalty. The solution to this problem will provide airlines with a powerful tool to analyze passenger satisfaction levels and gain a deeper understanding of the aspects that drive customer preferences. By leveraging machine learning techniques, airlines can proactively address potential issues, optimize their services, and ultimately enhance the overall travel experience for their passengers.

**2.2 EMPATHY MAP CANVAS**

An empathy map is a simple, easy-to-digest visual that captures knowledge about a user’s behaviours and attitudes.  It is a useful tool to helps teams better understand their users. Creating an effective solution requires understanding the true problem and the person who is experiencing it. The exercise of creating the map helps participants consider things from the user’s perspective along with his or her goals and challenges.

**AIRLINE CUSTOMER SATISFICATION MODEL:**

****

**2.3 Ideation & Brainstorming**

**IDEATION:**

**1. Sentiment Analysis:** Use natural language processing (NLP) techniques to analyze customer feedback, reviews, and social media posts related to airline experiences.

**2. Flight Experience Prediction:** Develop ML models that predict the likelihood of a positive or negative flight experience based on various factors such as flight duration, seat comfort, in-flight entertainment options, on-time performance, and customer service quality.

**3. Customer Segmentation**: Use ML clustering algorithms to segment airline passengers into different groups based on their preferences, demographics, and travel patterns.

**4. Feature Importance Ranking:** Employ ML techniques to determine the relative importance of different features in influencing passenger satisfaction.

**5. Anomaly Detection:** Utilize ML algorithms to detect anomalies in passenger feedback and behavior. This can help identify exceptional cases where passengers have significantly different experiences, allowing airlines to investigate and address any potential issues that may impact overall satisfaction.

**6. Recommender Systems**: Develop ML-based recommender systems that suggest personalized services and amenities to passengers based on their preferences and historical data.

**7. Predictive Maintenance:** ML models can analyze aircraft performance data and predict maintenance requirements.

**8. Dynamic Pricing:** Use ML algorithms to optimize pricing strategies based on passenger demand, flight occupancy, and market conditions. Pricing models can take into account factors that affect passenger satisfaction, such as fare affordability, value for money, and fare transparency**.**

**9. Image Recognition for Cabin Cleanliness:** Develop ML models that analyze images taken inside the aircraft cabins to assess cleanliness levels

**10. Real-time Customer Feedback Analysis**: Implement ML algorithms to analyze real-time customer feedback during flights, such as in-flight surveys or live chat interactions.

**BRAINSTORMING:**

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions.

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

**Graphical user interface, application

Description automatically generated**

**Step-2: Brainstorm, Idea Listing and Grouping**

**Graphical user interface, treemap chart

Description automatically generated**

**Step-3: Idea Prioritization**

**Diagram

Description automatically generated**

**2.4 Proposed Solution**

**1. Data Collection:** Gather relevant data from various sources, including passenger demographics, flight details (such as route, class, and duration), customer feedback surveys, social media sentiment, and other related data points.

**2. Data Preprocessing:** Clean and preprocess the collected data to ensure its quality and suitability for analysis.

**3. Feature Engineering:** Extract meaningful features from the collected data that are likely to impact passenger satisfaction. This can involve creating new variables or combining existing ones to capture important aspects of the travel experience, such as flight punctuality, seat comfort, in-flight amenities, and customer service quality.

**4. Model Selection:** Regression models, classification algorithms (such as logistic regression or decision trees), or ensemble methods (such as random forests or gradient boosting) can be considered depending on the nature of the problem and desired outputs.

**5. Model Training and Evaluation:** Split the preprocessed data into training and testing sets. Train the selected ML models on the training data and evaluate their performance using appropriate evaluation metrics such as accuracy, precision, recall, or F1 score.

**6. Feature Importance Analysis:** Techniques like feature importance ranking or permutation can be employed to identify the key factors that significantly influence passenger satisfaction.

**7. Prediction and Analysis:** Deploy the trained ML models to predict passenger satisfaction levels for new or unseen data.

**8. Insights and Recommendations:** These insights can be used to make data-driven recommendations for improving specific aspects of airline services, such as customer service training, on-time performance, in-flight amenities, or seating arrangements.

**9. Continuous Improvement:** Continuously update and refine the ML models as new data becomes available. Regularly analyze passenger feedback, monitor changing trends, and adapt the models to reflect evolving passenger preferences and market dynamics.

**3.REQUIREMENT ANALYSIS**

**3.1 Functional requirement**

**1. Data Integration:**

The system should be able to collect and integrate relevant data from various sources, including passenger demographics, flight details, customer feedback surveys, and social media sentiment.

**2. ML Model Development:**

The system should provide functionality to select and develop ML models appropriate for identifying passenger satisfaction, such as regression models or classification algorithms.

**3. Training and Evaluation:**

The system should enable the training of ML models using the collected and preprocessed data.

It should allow for the evaluation of model performance using appropriate evaluation metrics, such as accuracy, precision, recall, or F1 score.

**4. Real-time Prediction:**

The system should provide the capability to deploy trained ML models to predict passenger satisfaction levels for new or unseen data in real-time.

It should allow for efficient and accurate prediction processing to provide prompt feedback and insights.

**5. Feature Importance Analysis:**

The system should enable the analysis of feature importance to identify the key factors that significantly influence passenger satisfaction.

**6. Insights and Reporting:**

The system should provide functionality to derive actionable insights from the ML models and analysis results.

It should generate reports or visualizations that present the identified drivers of passenger satisfaction and provide recommendations for improvements.

**7. Model Maintenance and Update:**

The system should support continuous model maintenance and updates as new data becomes available.

**8. Integration with Feedback Channels:**

The system should integrate with feedback channels, such as customer surveys or social media monitoring tools, to collect real-time feedback and incorporate it into the analysis process.

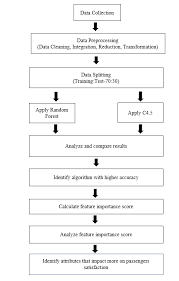
**9. Scalability and Performance:**

The system should be scalable to handle large volumes of data and perform computation effectively.

**10. Security and Privacy:**

The system should ensure the security and privacy of passenger data throughout the data integration, training, and analysis processes.

Compliance with relevant data protection regulations, such as GDPR, should be considered and implemented.



**3.2 Non-Functional requirements**

**1. Performance:**

The system should be able to process and analyze large volumes of data efficiently, ensuring fast response times for real-time predictions and analysis. ML model training and evaluation should be optimized for speed and scalability to handle increasing data sizes and complexity.

**2. Accuracy:**

The ML models used for identifying passenger satisfaction should exhibit high accuracy in predicting satisfaction levels based on the available data. The system should strive for minimal errors and maximize the precision of the predictions to provide reliable insights.

**3. Reliability:**

The system should be reliable and available for use at all times to ensure continuous analysis and prediction capabilities. It should be designed to handle potential failures or interruptions gracefully and recover quickly without losing important data or functionality.

**4. Scalability:**

The system should be designed to scale horizontally or vertically to accommodate increasing data volumes, user demands, and computational requirements. It should leverage distributed computing or cloud resources to support the growth of data and user load.

**5. Security:**

The system should ensure the security and confidentiality of passenger data throughout the entire process, including data integration, training, prediction, and analysis.

**6. Usability:**

The system should have a user-friendly interface that allows users, such as analysts or

stakeholders, to interact with the ML models and access relevant insights easily.

**7. Maintainability:**

The system should be designed and implemented in a modular and maintainable manner, allowing for easy updates, enhancements, and bug fixes. Proper documentation and code organization should be provided to facilitate system maintenance and future improvements.

**8. Ethical Considerations:**

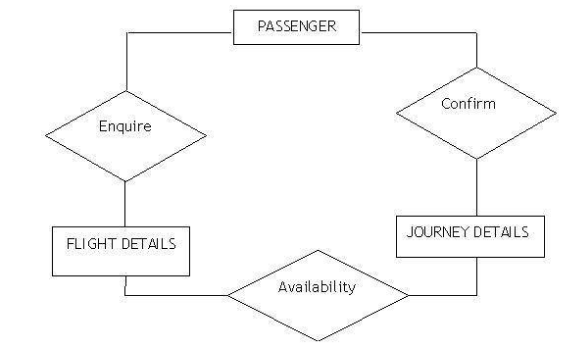
The system should adhere to ethical guidelines and principles when handling passenger data and making predictions. Bias detection and mitigation techniques should be employed to ensure fair and unbiased predictions, considering factors like age, gender, or race.

**9. Compatibility:**

The system should be compatible with different data formats, APIs, and technologies commonly used in the airline industry, ensuring seamless integration with existing systems and data sources.

**10. Compliance:**

The system should comply with relevant industry regulations and standards, such as data protection laws, privacy regulations, and aviation industry guidelines. It should adhere to best practices and ethical standards in handling passenger data and ensuring transparency in the analysis and prediction processes.



**4. PROJECT DESIGN**

Project Design for Identifying Airline Passenger Satisfaction Using ML:

1. Define the Problem

2. Data Collection

3. Data Preprocessing and Exploration

4. Model Selection

5. Model Development and Training

6. Feature Importance Analysis

7. Real-time Prediction and Analysis

8. Insights and Recommendations

9. Iterative Improvement

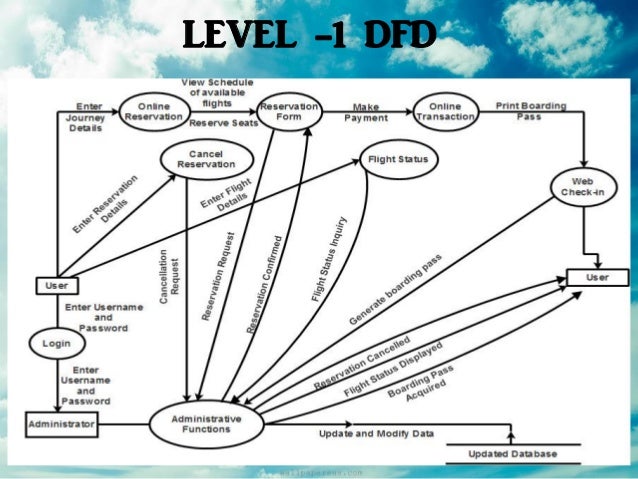
10. Deployment and Integration

11. Monitoring and Maintenance

12. Ethical Considerations

**4.1 Data Flow Diagrams**

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.



**4.2 Solution & Technical Architecture**

**Solution Overview:**

The solution for identifying airline passenger satisfaction using ML involves collecting and analyzing relevant data to predict and understand passenger satisfaction levels. It utilizes ML algorithms to train models, perform predictions, and generate insights for airlines to improve their services and enhance customer experiences.

**Technical Architecture:**

**1. Data Collection Layer:**

Integration with various data sources, such as customer feedback surveys, social media APIs, airline databases, and operational systems. Data collection mechanisms to gather passenger demographics, flight details, customer reviews, sentiment data, and other relevant information.

**2. Data Preprocessing and Feature Engineering:**

Data preprocessing modules to clean and transform the collected data, handling missing values, outliers, and data inconsistencies. Feature engineering techniques to extract meaningful features related to passenger satisfaction, such as flight duration, seat comfort, customer ratings, and service quality.

**3. ML Model Development and Training:**

Selection of ML algorithms suitable for the task, such as regression models, classification algorithms, or ensemble methods. Training of ML models using the preprocessed data, tuning hyperparameters, and utilizing cross-validation techniques for model evaluation.

**4. Real-time Prediction and Analysis:**

Deployment of trained ML models to predict passenger satisfaction levels for real-time or incoming data. Integration with prediction modules to process new data, generate predictions, and analyze passenger satisfaction scores.

**5. Insights Generation and Reporting:**

Derivation of actionable insights from the ML models and analysis results, identifying the key factors influencing passenger satisfaction. Generation of reports, visualizations, or dashboards to present the insights, trends, and recommendations to stakeholders and decision-makers.

**6. Integration and Scalability:**

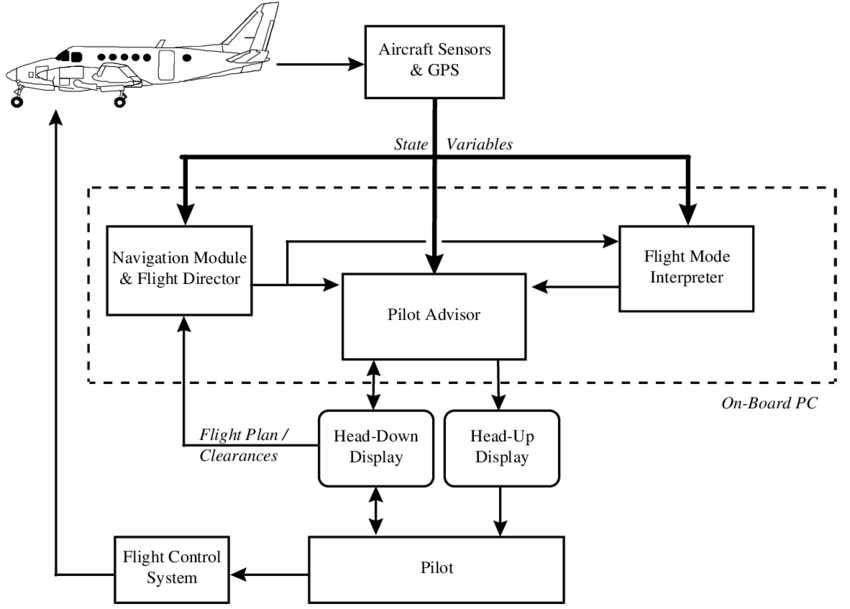
Integration with existing airline systems, databases, or APIs to leverage operational data and facilitate seamless data flow. Scalable infrastructure to handle large volumes of data and ensure optimal performance during training, prediction, and analysis.

**7. Security and Privacy:**

Implementation of robust security measures to protect passenger data and ensure compliance with data protection regulations. Encryption techniques, access controls, and secure communication protocols to safeguard sensitive information throughout the system.

**8. Monitoring and Maintenance:**

Monitoring mechanisms to track the performance, health, and accuracy of the deployed ML models in real-time. Regular model updates, bug fixes, and retraining cycles to maintain model performance and adapt to changing passenger preferences.

****

**4.3 User Stories**

1. As a data scientist, I want to preprocess and clean the collected data, handling missing values and outliers, to ensure accurate and reliable analysis.

2. As a data scientist, I want to develop and train ML models using the preprocessed data to predict passenger satisfaction levels based on various factors such as flight punctuality, seat comfort, and customer service quality.

3. As an airline manager, I want to receive real-time predictions of passenger satisfaction for new or upcoming flights, allowing me to address potential issues or concerns proactively.

4. As a data scientist, I want to generate visualizations and reports that present insights on passenger satisfaction trends, allowing stakeholders to understand the areas of improvement and make data-driven decisions.

5. As a customer service representative, I want access to a dashboard that shows real-time customer sentiment analysis from social media and customer feedback, enabling me to address specific concerns and provide personalized assistance.

6. As an airline executive, I want to receive regular reports and updates on passenger satisfaction scores, allowing me to evaluate the effectiveness of initiatives aimed at improving customer experiences.

7. As a marketing manager, I want to identify patterns and correlations between passenger satisfaction and customer loyalty or repeat bookings, enabling targeted marketing strategies to enhance customer retention.

8. As an IT administrator, I want to ensure the security and privacy of passenger data throughout the ML process, implementing measures such as data encryption, access controls, and compliance with data protection regulations.

9. As a data scientist, I want to continuously monitor the performance of the ML models and retrain them periodically to adapt to changing passenger preferences and evolving market dynamics.

10. As an airline operations manager, I want to integrate the ML solution with existing systems and databases to streamline data flow and ensure seamless integration into operational workflows.

**5. CODING & SOLUTIONING**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

from xgboost import XGBClassifier

from sklearn.metrics import accuracy\_score,mean\_squared\_error

from sklearn.ensemble import RandomForestClassifier

from sklearn.linear\_model import LogisticRegression

These import statements bring in the necessary libraries, classes, and functions required for data manipulation, visualization, preprocessing, and machine learning modeling in the subsequent code.

train\_data = pd.read\_csv("/content/drive/train 2.csv")

test\_data = pd.read\_csv("/content/drive/test 2.csv")

The code loads the data from the specified CSV files into pandas DataFrames, allowing for easy manipulation, analysis, and modeling of the data. The actual file paths may vary depending on the location and naming of the CSV files in your system.

train\_data.head()

By executing train\_data.head(), you will see the first five rows of the train\_data DataFrame, along with the column names and corresponding values. This is helpful in understanding the structure of the data, checking for missing values, and getting a general sense of the data's format

test\_data.head()

By executing test\_data.head(), you will see the first five rows of the test\_data DataFrame, including the column names and corresponding values. This allows you to quickly inspect the structure and content of the data.

train\_data.shape

When you access the shape attribute of a DataFrame in pandas, it returns a tuple representing the dimensions of the DataFrame. The tuple contains two values: the number of rows and the number of columns, respectively. By executing train\_data.shape, you will get an output that looks like (n\_rows, n\_columns), where n\_rows represents the number of rows in the DataFrame and n\_columns represents the number of columns.

test\_data.shape

When you access test\_data.shape, it returns a tuple containing two values: the number of rows and the number of columns in the DataFrame. The first element of the tuple represents the number of rows, and the second element represents the number of columns.

train\_data.isnull().sum()

test\_data.isnull().sum()

Calculate the sum of missing values in each column of the train\_data and test\_data DataFrames.

train\_data.dropna(inplace=True)

test\_data.dropna(inplace =True)

Remove any rows containing missing values from the train\_data and test\_data DataFrames using the dropna() function with inplace=True, which modifies the DataFrames in place.

train\_data.isnull().sum()

Checks for missing values again in the train\_data DataFrame to verify if all missing values have been dropped successfully.

def drops(data):

data.drop({"Unnamed: 0","id"},axis=1,inplace=True)

data["Gender"] = data["Gender"].map({"Male":1,"Female":0})

data["Customer Type"] = data["Customer Type"].map({"Loyal Customer":1,"disloyal Customer":0})

data["Type of Travel"] = data["Type of Travel"].map({"Business travel":1,"Personal Travel":0})

data = pd.get\_dummies(data,columns = ["Class"])

return data

This code defines a function named drops() that performs several data transformations and feature engineering steps on a given DataFrame data. It drops the columns "Unnamed: 0" and "id" using the drop() function, maps categorical variables to numerical values, and performs one-hot encoding for the "Class" column using pd.get\_dummies().

train\_data["Class"].value\_counts()

This line counts the number of occurrences of each unique value in the "Class" column of the train\_data DataFrame.

data = drops(train\_data)

data.head()

These lines call the drops() function on the train\_data DataFrame and assign the transformed DataFrame to the variable data. It then displays the first few rows of the data DataFrame.

data\_test = drops(test\_data)

data\_test.head()

These lines call the drops() function on the test\_data DataFrame and assign the transformed DataFrame to the variable data\_test.

ytrain\_target = data["satisfaction"]

xtrain\_feature = data.drop({"satisfaction"}, axis=1)

xtrain\_feature.head()

This code assigns the "satisfaction" column of the data DataFrame to the variable ytrain\_target. It drops the "satisfaction" column from the data DataFrame and assigns the resulting DataFrame to xtrain\_feature. The head() function is then used to display the first few rows of xtrain\_feature

ytest\_target = data\_test["satisfaction"]

xtest\_feature = data\_test.drop({"satisfaction"}, axis=1)

xtest\_feature.head()

The "satisfaction" column of the data\_test DataFrame to ytest\_target. It drops the "satisfaction" column from the data\_test DataFrame and assigns the resulting DataFrame to xtest\_feature. The head() function is then used to display the first few rows of xtest\_feature.

ytrain\_target.head()

ytrain\_target.value\_counts()

These lines display the first few rows of the ytrain\_target Series (containing the "satisfaction" values) and then perform a count of each unique value in the ytrain\_target Series.

y\_train = ytrain\_target.map({"neutral or dissatisfied":0, "satisfied":1})

y\_test = ytest\_target.map({"neutral or dissatisfied":0, "satisfied":1})

y\_train.head()

y\_test.head()

This code maps the categorical values in ytrain\_target and ytest\_target to numerical values using the map() function. It assigns 0 to "neutral or dissatisfied" and 1 to "satisfied". The transformed Series are assigned to y\_train and y\_test. The head() function is then used to display the first few rows of both Series.

sns.countplot(x="Online boarding", hue="satisfaction", data=train\_data)

sns.countplot(x="Customer Type", hue="satisfaction", data=train\_data)

This code maps the categorical values in ytrain\_target and ytest\_target to numerical values using the map() function. It assigns 0 to "neutral or dissatisfied" and 1 to "satisfied". The transformed Series are assigned to y\_train and y\_test. The head() function is then used to display the first few rows of both Series.

rfc = RandomForestClassifier()

rfc.fit(xtrain\_feature, y\_train)

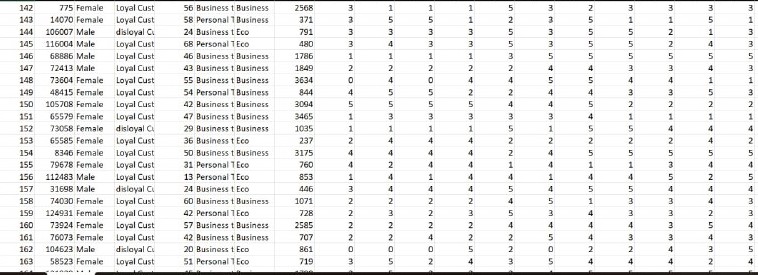
pred = rfc.predict(xtest\_feature)

accuracy\_score(y\_test, pred)

This code creates a RandomForestClassifier object named rfc and fits it to the training data. It then uses the trained model to make predictions on the test data and assigns the predictions to pred. Finally, it calculates the accuracy score by comparing the predicted values with the actual values (y\_test) using the accuracy\_score() function.

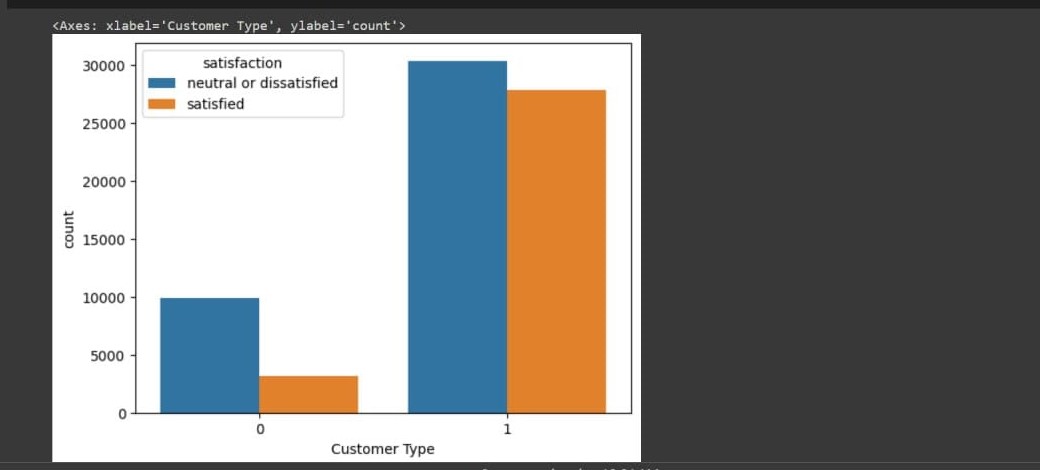
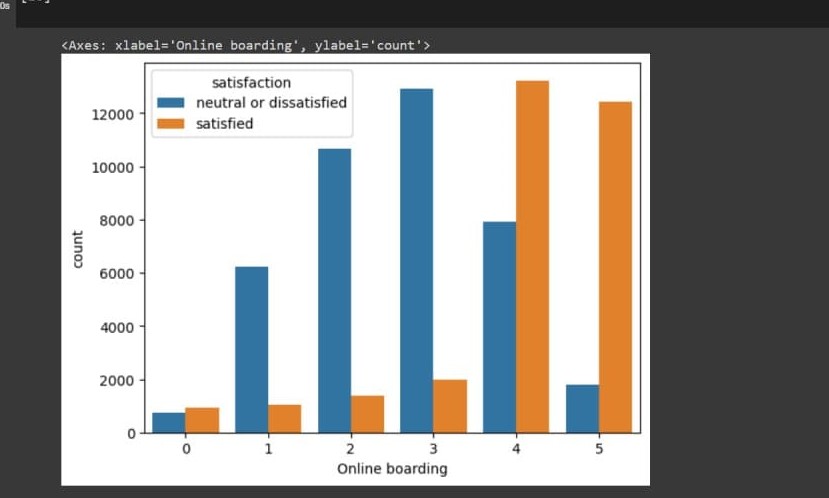
Overall, this code performs various data preprocessing steps, including handling missing values, feature engineering, and mapping categorical variables to numerical values. It also uses a Random Forest Classifier to train a model and make predictions on the test data, evaluating the accuracy of the model's predictions.

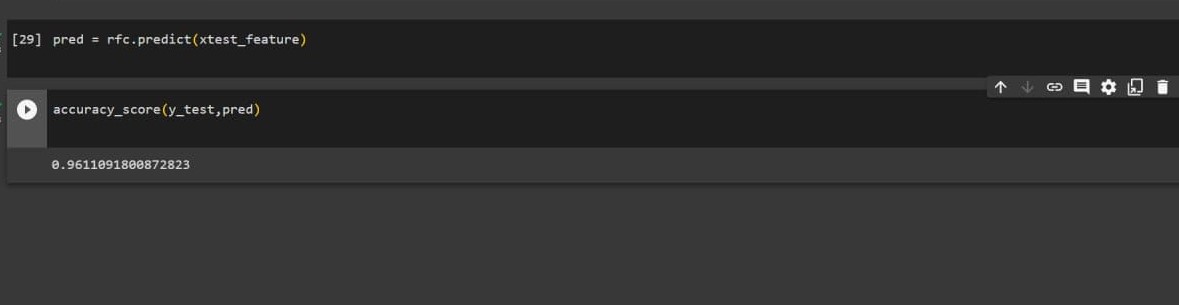
**DATASET SCHEMA:**

****

**6. RESULTS**

**6.1 Performance Metrics**

****

****

**7. ADVANTAGES & DISADVANTAGES**

**ADVANTAGES:**

**1. Enhanced Understanding:** ML algorithms provide a deeper understanding of passenger satisfaction by analyzing large volumes of data and identifying complex patterns and correlations that may not be easily discernible through traditional methods.

**2. Improved Accuracy:** ML models can make more accurate predictions of passenger satisfaction based on historical data, enabling airlines to proactively address potential issues and deliver personalized experiences.

**3. Real-time Insights:** ML-based systems can generate real-time insights on passenger satisfaction, allowing airlines to quickly identify and address emerging trends or issues, resulting in improved service quality.

**4. Data-driven Decision Making:** ML enables data-driven decision making by providing actionable insights and recommendations based on the analysis of passenger satisfaction factors, helping airlines prioritize improvements and allocate resources effectively.

**5. Customer-centric Approach:** ML-based systems help airlines gain a customer-centric perspective by considering multiple factors that contribute to passenger satisfaction, leading to tailored experiences and improved customer retention.

**DISADVANTAGES :**

**1. Data Limitations:** ML models heavily rely on the quality and availability of data. If the data used for training the models is incomplete, biased, or unrepresentative, it can lead to inaccurate predictions and biased insights.

**2. Interpretability Challenges:** ML models, particularly complex ones like deep learning models, can be challenging to interpret and explain. It may be difficult to understand the underlying factors that contribute to the model's predictions, reducing transparency and trust.

**3. Over-reliance on Historical Data:** ML models are trained on historical data, which may not fully capture future trends or changing passenger preferences. This can result in a lag in adapting to evolving customer expectations and behaviors.

**4. System Complexity and Cost:** Developing and implementing ML-based solutions require technical expertise, infrastructure, and computational resources, which can be costly and complex to set up and maintain.

**5. Ethical Considerations:** ML models may inadvertently introduce biases if not carefully designed and monitored. Biases based on factors such as age, gender, or race can negatively impact predictions and perpetuate discriminatory practices.

**8. CONCLUSION**

Identifying airline passenger satisfaction using ML brings numerous benefits and opportunities for airlines to enhance their services and improve customer experiences. ML algorithms enable the analysis of large volumes of data, providing deeper insights and predictive capabilities that can contribute to proactive decision making. By leveraging ML models, airlines can accurately predict passenger satisfaction levels, identify key factors influencing satisfaction, and prioritize improvements accordingly. Real-time monitoring and analysis allow for quick identification and resolution of issues, leading to enhanced service quality and personalized experiences. However, it is essential to address potential challenges such as data limitations, interpretability concerns, and ethical considerations. Ensuring data quality, transparency in model predictions, and ethical use of passenger data are crucial for building trust and avoiding biases. Overall, the use of ML in identifying airline passenger satisfaction offers a data-driven approach to understand customer preferences, optimize services, and drive customer loyalty. By continuously monitoring and adapting the ML models, airlines can stay responsive to changing passenger expectations and maintain high levels of satisfaction, ultimately benefiting both the airlines and their passengers.

**9. FUTURE SCOPE**

The future of identifying airline passenger satisfaction using ML holds great potential for innovation and customer-centric advancements. By leveraging advanced ML techniques, airlines can continue to enhance services, improve customer experiences, and maintain a competitive edge in the industry.

**10. APPENDIX**

**SOURCE CODE**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

from xgboost import XGBClassifier

from sklearn.metrics import accuracy\_score,mean\_squared\_error

from sklearn.ensemble import RandomForestClassifier

from sklearn.linear\_model import LogisticRegression

train\_data = pd.read\_csv("/content/drive/train 2.csv")

test\_data = pd.read\_csv("/content/drive /test 2.csv")

train\_data.head()

test\_data.head()

train\_data.shape

test\_data.shape

train\_data.isnull().sum()

test\_data.isnull().sum()

train\_data.dropna(inplace=True)

test\_data.dropna(inplace =True)

train\_data.isnull().sum()

def drops(data):

data.drop({"Unnamed: 0","id"},axis=1,inplace=True)

data["Gender"] = data["Gender"].map({"Male":1,"Female":0})

data["Customer Type"] = data["Customer Type"].map({"Loyal Customer":1,"disloyal Customer":0})

data["Type of Travel"] = data["Type of Travel"].map({"Business travel":1,"Personal Travel":0})

data = pd.get\_dummies(data,columns = ["Class"])

return data

train\_data["Class"].value\_counts()

data = drops(train\_data)

data.head()

data\_test = drops(test\_data)

data\_test.head()

ytrain\_target = data["satisfaction"]

xtrain\_feature = data.drop({"satisfaction"},axis=1)

xtrain\_feature.head()

ytest\_target = data\_test["satisfaction"]

xtest\_feature = data\_test.drop({"satisfaction"},axis=1)

xtest\_feature.head()

ytrain\_target.head()

ytrain\_target.value\_counts()

y\_train = ytrain\_target.map({"neutral or dissatisfied":0,"satisfied":1})

y\_test = ytest\_target.map({"neutral or dissatisfied":0,"satisfied":1})

y\_train.head()

y\_test.head()

sns.countplot(x="Online boarding",hue="satisfaction",data=train\_data)

sns.countplot(x="Customer Type",hue="satisfaction",data=train\_data)

rfc = RandomForestClassifier()

rfc.fit(xtrain\_feature,y\_train)

pred = rfc.predict(xtest\_feature)

accuracy\_score(y\_test,pred)

**PROJECT VIDEO DEMO LINK**

https://drive.google.com/file/d/1h6yOTgq1t7yviwMN2q9HotyWQVYbpDHO/view?usp=drivesdk