Classification
Project: Airline Passenger Referral Prediction

Project Description:

Business Context
In the dynamic and highly competitive airline industry, companies like IndiGo are continuously striving to enhance customer experience and build lasting loyalty. A critical component of this effort is understanding and predicting passenger referrals, which can significantly influence the airline's market position and reputation. The analysis of passenger reviews from 2006 to 2019 offers IndiGo a comprehensive view into the various facets of the travel experience as perceived by their customers. These reviews provide invaluable feedback on service quality, including comfort, service, and overall value, and play a key role in shaping potential customers' perceptions.
By leveraging this extensive dataset, IndiGo aims to develop a predictive model to identify which passengers are most likely to recommend the airline to others. This insight enables IndiGo to:
1. Enhance Customer Experience:
By pinpointing the factors that most influence positive referrals, IndiGo can focus on improving these aspects of their service, ensuring that they meet or exceed customer expectations.
2. Targeted Improvements:
The ability to predict referrals allows IndiGo to address specific areas needing improvement, whether it be in-flight comfort, customer service, or overall value for money.
3. Strategic Marketing:
Understanding referral patterns can help IndiGo tailor their marketing strategies to leverage positive word-of-mouth and foster a strong brand reputation.
4. Competitive Edge:
By continuously refining their services based on customer feedback, IndiGo can differentiate itself from competitors and strengthen its position in the market.

Main Libraries to be Used:

- Pandas for data manipulation, aggregation
- Matplotlib and Seaborn for visualisation and behaviour with respect to the target variable
- NumPy for computationally efficient operations
- Scikit Learn for model training, model optimization, and metrics calculation

Project should include:

1. **Problem Statement**
2. **Import libraries**
3. **Load dataset**
4. **Data cleaning**
   - Handle missing values

- ○ Convert data types
- ○ Remove duplicates
5. **Exploratory Data Analysis (EDA)**
   - ○ Visualize distributions
     - i. Univariate analysis
     - ii. Bivariate analysis
     - iii. Multivariate analysis
   - ○ Correlation analysis
   - ○ Feature-target relationships
6. **Outlier treatment**
   - ○ Boxplot
7. **Check distributions & apply transformations (if needed)**
   - ○ Skewness/Kurtosis
   - ○ Log Transformation, sqrt
8. **Feature engineering**
   - ○ Create new features
   - ○ One-hot encoding (for categorical)
9. **Split data into train/test sets**
10. **Train Logistic Regression model**
11. **Feature Scaling**
12. **Prediction using the algorithm**
13. **Evaluate model performance using Confusion Metrics**
    - ○ Accuracy
    - ○ Precision
    - ○ Recall
    - ○ F1 Score
14. **Perform the same steps for**
    - ○ Decision tree Classifier
    - ○ Random Forest Classifier
    - ○ Support Vector Machine Classifier
    - ○ K Nearest Neighbor Classifier
15. **Perform the Cross Validation using Cross_val_score for all the algorithms**
16. **Print the final Conclusion**

Link to Dataset: https://github.com/rahulinchal/SPPU

Note:
The data provided in the Github is the truncated one of the original data. Once the algorithm is done and the final observation is done from your end.
Try changing the data source to the original data and check for the accuracy.
Let me know if you have done that.

Data Description:

Fields	Description
Airline	Name of the airline
Overall	Overall points are given to the trip between 0 - 10
Author	Author of the trip
Reviewdate	Date of the Review Customer
Aircraft	Type of the aircraft
Traveller type	Type of the traveller (eg Business, Leisure)
Flight date	Cabin at the flight date flown
Cabin Service	Rated between 1-5
Foodbev	Rated between 1-5 entertainment: Rated between 1-5
Grounservice	Rated between 1-5
Value for money	Rated between 1-5
Recommended	Binary, Target value