Python ml classification project

Airline customer satisfaction level

Decision Tree

Project Description:

In this particular project, we are using a dataset that contains information like, Gender, Customer_Type, Age, Type_of_Travel, Class etc and using that to predict the customer satisfaction level.
However, before you go ahead and make a prediction, it is advised that you first pre-process the data, since it may contain some irregularities and noise.
In addition, try various tricks and techniques in order to gain the best accuracy in your predictions.

Data details:

- **id** : Unique id number to each passenger.
- **Gender**: Gender of the passengers (Female, Male)
- **Customer Type**: The customer type (Loyal customer, disloyal customer)
- **Age**: The actual age of the passengers
- **Type of Travel**: Purpose of the flight of the passengers (Personal Travel, Business Travel)
- **Class**: Travel class in the plane of the passengers (Business, Eco, Eco Plus)
- **Flight distance**: The flight distance of this journey
- **Inflight wifi service**: Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)
- **Departure/Arrival time convenient**: Satisfaction level of Departure/Arrival time convenient
- **Ease of Online booking**: Satisfaction level of online booking
- **Gate location**: Satisfaction level of Gate location
- **Food and drink**: Satisfaction level of Food and drink
- **Online boarding**: Satisfaction level of online boarding
- **Seat comfort**: Satisfaction level of Seat comfort
- **Inflight entertainment**: Satisfaction level of inflight entertainment
- **On-board service**: Satisfaction level of On-board service
- **Leg room service**: Satisfaction level of Leg room service
- **Baggage handling**: Satisfaction level of baggage handling
- **Check-in service**: Satisfaction level of Check-in service
- **Inflight service**: Satisfaction level of inflight service
- **Cleanliness**: Satisfaction level of Cleanliness

- **Departure Delay in Minutes**: Minutes delayed when departure
- **Arrival Delay in Minutes**: Minutes delayed when Arrival
- **Satisfaction**: Airline satisfaction level(Satisfaction, neutral or dissatisfaction)

Part-1: data Exploration and Pre-processing

1) load the given dataset
2) print all the column names
3) describe the data
4) Drop the column 'Unnamed'
5) Replace all the " " in column with "_"
6) Give label to a satisfaction column value without using any encoding method
7) Plot the number of satisfied customers and the number of unsatisfied customers
8) find the mean value of satisfaction of male and female customers
9) find the mean value of satisfaction of customers with respect to Age.
10) find the mean value of satisfaction of customers with respect to Food_and_drink.
11) Display a boxplot for Flight_Distance
12) Display a boxplot for Checkin_service
13) Find all the Null values
14) Drop all the na values
15) Find the unique values in Flight_Distance

Part-2: Working with models

1) Perform encoding in columns Gender, Customer_Type, Type_of_Travel, and Class.
2) Drop the column id and unnamed:_0.1
3) Create the features and target Data
4) Perform scaling on features data
5) Split the data in training and testing sets
6) Fit the decision tree model with various parameters
7) Create a function to display precision score, recall score, accuracy, classification report, confusion matrix, F1 Score.