

UNIVERSITI MALAYA  
UNIVERSITY OF MALAYA

SESI AKADEMIK 2023/2024 : SEMESTER II  
ACADEMIC SESSION 2023/2024 : SEMESTER II

WQD7006 : *Machine Learning for Data Science*

Mar 2024 : Duration: 12th May 2024 (3.15 pm) to (4.30 pm)

---

**Mid Term Test: 15 %**

---

**INSTRUCTIONS TO CANDIDATES:**

Candidate is required to answer all questions.

**Case Study – Machine learning for airline customers' satisfaction**

"Airline success in the post-COVID-19 era will hinge on a combination of building consumer confidence and operational flexibility with changing schedules and routes. Airlines have a tremendous reputation for safety. That will be even more critical as passengers look to airlines for detailed and specific information about what's being done to keep them safe," says Michael Taylor, travel intelligence lead at J.D. Power.

In the aviation sector, investments and volume of businesses are increasing each day hence competition would increasingly deepen. Accordingly, airline companies must be financially strong to establish themselves and to survive in such a competitive world. At the same time, customers have been the most important stakeholder to the airline companies. With the intense competition in air transportation market, having satisfied customers is the key towards success.

Indeed, customers' satisfaction are products of multidimensional features such as inflight Wi-Fi service, departure and arrival delay in minutes, departure, and arrival time convenient, ease of online booking, gate location, food and drink, online boarding, seat comfort, baggage handling, cleanliness, check-in service and others. Hence, airline companies must critically examine those features to find the potential factor to produce quality airline services to the customers.

Meanwhile, machine learning (ML) is a trending tool to predict airline customers' satisfaction. Remarkably, this technique is pertinent as it is also apt to assess the accuracy of selected model and to identify various problems affecting the airline industry. These are the crucial assessment towards creating the above satisfactory at the same time the preferred airline in the world.

Use the given dataset to pre-process and apply machine learning (ML) models for predictions of airline customers' satisfaction. To perform the practical output, you may use Python or R programming.

---

### Dataset Details.

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 Wi-Fi service: Satisfaction level of the inflight Wi-Fi 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

#### Target

Satisfaction: Airline satisfaction level (Satisfaction, neutral or dissatisfaction)

---

Based on the given case, answer the following questions.

#### QUESTION 1 (3 Marks)

Discuss any **Three (3)** problems associated with the predictions of airline customers' satisfaction, given machine learning algorithms are not in use.

#### QUESTION 2 (12 Marks)

You have been tasked with predicting the level of airline customer satisfaction (Satisfaction, Neutral, or Dissatisfaction) using linear and multiple regression techniques.

a) Describe the exploratory data analysis you would perform to understand the data distribution and the relationship between the features and the target variable. Include at least two data visualizations to show the relationship between the features and the target variable and provide an explanation.

**(4 Marks)**

b) Apply any two different machine learning models and use up to three appropriate evaluation metrics to determine the best-performing model and explain why it outperformed the others.

**(8 Marks)**

**TAMAT  
END**