

Data Communication

Final project

Babol Noshirvani University of Technology (NIT)

Fall 2023

A dataset about flight satisfaction is provided to you (flight_satisfaction.csv).

1- Data preparation: Import the data (all data) then modify dataset

according to following process:

- Open the "project_student_data.ipynb" and replace STUID with your personal student ID then run the code. It will give you your exclusive features.
- Then you only need those selected features

2- Data cleaning: Handle missing or duplicate data, deal with outliers, and perform data type conversions.

3- Data Manipulation and transformation: Perform operations like selecting specific columns, filtering rows, adding or removing columns, and transforming data as needed. (consider the correlation between features as if it needed).

Beside those things you should consider normalization, scaling or etc.

4- Data Analysis:

- a) The purpose of this exercise is computing satisfaction of flight. in this part you will be asked to calculate this (you are free to choose neural network models or any other method).

For this case you should:

- Choose 2000 random rows and split it to test and train

- Define your model
- Train your model and fit it
- Do this steps with 5000 rows and compare it with first scenario

b) You should also order features from most effective to least one.

5- Data Visualization: Use libraries like Matplotlib, Seaborn, or Plotly to visualize the data.

6- Drawing conclusions and documentation: Based on the analysis and modeling results, draw conclusions and insights from the data.

Document your process, findings, and any important steps taken.

-Download files from links below:

- https://drive.google.com/file/d/1iXaZQpDuFsu73unlMXOoce1DbjXJDFcw/view?usp=drive_link
- https://drive.google.com/file/d/1z9hd9GUY6KsB2jcNmcRAbrEg3Jh7Ec1Z/view?usp=drive_link

-Using libraries like **Scikit-learn** is illegal.