

Laboratory assignment

Component 1

Authors: Patricia Moga, Ciobanu Sergiu-Tudor

Group: 6

January 13, 2026

1 Problem definition

The main concept that our problem addresses is the level of satisfaction of airline passengers following a flight. After taking part in a questionnaire, passengers' responses provide the essential information for the airline to classify the level of satisfaction, which is categorized as "Satisfaction" and "Neutral or Dissatisfaction". To resolve the issue of determining this satisfaction level, a predictive system must be implemented. The system will assess the data, learn specific patterns based on previous surveys, and then use the knowledge to predict the satisfaction for new, unseen passengers based solely on their input features.

2 Problem specification

- *Input data:* 24 of which 22 relevant features from the Airline Passenger satisfaction Dataset
- *Description of the input data:* The input data offers a variety of information grouped in three main categories:
 - Passenger demographics: high-level data of the passenger and their travel reason, and the flight distance. This set comprises a wide range of values; meaning there are:
 - categorical nominal: Gender (Male, Female), Customer Type: Loyal/Disloyal customer, Type of Travel(Personal/Business Travel)
 - categorical ordinal: Class (Business, Eco, Eco Plus)
 - numerical: Age
 - Service ratings: passengers' subjective satisfaction with various services on a fixed interval (1 to 5, and 0 that means not rated). These are ordinal variables: Inflight wifi service, Ease of Online booking, Food and Drink, etc.
 - Flight Metrics: distance of the flight journey and the duration in minutes of delays at the time of departure and arrival, measured on a discrete or continuous scale output data
- *Preconditions:*
 - Dataset contains complete records with all 22 features for each passenger
 - Service ratings are within valid range (0-5), categorical variables match defined values, and numerical features are non-negative
- *Output Data:* The predicted binary class label for passenger satisfaction.

- *Postconditions:*
 - The binary satisfaction label (satisfied/dissatisfied) for each passenger is predicted based on their feature profile (For Supervised Clustering)
 - Each passenger is assigned to one specific cluster based on patterns in their features (For Unsupervised Clustering)

3 specification of the learning Tasks

3.1 Supervised Classification

- **Task** The task is to classify an unseen passenger profile into one of two predefined categories: "Satisfaction" or "Neutral or Dissatisfaction".
- **Performance** Performance is measured by the model's predictive quality on an unseen test dataset.
- **Experience** The model learns from the labeled dataset of passenger surveys, where each input record is paired with the correct satisfaction category.

3.2 Unsupervised Classification

- **Task** The algorithm groups passengers into distinct segments based on their profile statistics, travel information, and satisfaction ratings, and it would identify which aspects tend to be rated similarly together.
- **Performance** Performance is measured by how well-separated and compact the discovered passengers groups are, whether the algorithm can meaningfully reduce the number of service ratings into fewer key factors without losing essential information.
- **Experience** The algorithm learns from the complete passenger survey records, iteratively refining how it groups similar passengers and relates different features.

References