CS-4038

Data Mining

Term Project

Group Members:

1) Ayesha Saqib 20F-0360
2) Mozeb Ahmed 20F-0161
3) Huma Mukhtar 20F-0139

Phase I Report:

1. Data Analysis:

The provided dataset contains information on students' assessment scores, including assignments, quizzes, mid-I scores, and final grades. The data is divided into seven sheets, each containing a different number of assignments and quizzes. For consistency, only the best 5 assignments and quizzes are included for each student before calculating their grades.

2. Preprocessing:

Handling Missing Values: Missing values in the dataset were filled with zeros to ensure data integrity.

Normalization: Assignment and quiz scores were normalized based on their total marks and corresponding weights to bring them to a uniform scale for analysis.

3. Results:

The analysis was performed using two classification algorithms:

  1) K Nearest Neighbour (KNN)
  2) Decision Tree.

K Nearest Neighbour (KNN):

- Accuracy: 92.9%

- Specificity: 100%
- Sensitivity: 83.3%

Decision Tree:

- Accuracy: 71.4%
- Specificity: 50%
- Sensitivity: 100%