Data Mining - EX1

Deadline: Wednesday, Mehr 18, 1402 - October 9, 2024

Exercise number 1

Scenario:

You are a data analyst working for a hospital. The hospital has accumulated a large dataset containing information on patients' medical histories, diagnoses, treatments, and outcomes. The goal is to analyze this data to improve patient care, optimize resource use, and enhance overall operational efficiency.

The dataset includes:

- **Patient Demographics**: Age, gender, address, occupation, etc.
- **Medical History**: Previous illnesses, surgeries, allergies, etc.
- **Diagnostic Results**: Blood tests, MRI, X-ray results, etc.
- **Treatments Administered**: Medications, surgeries, physiotherapy, etc.
- **Outcomes**: Recovery time, improvement in health, complications, etc.
- **Hospital Resource Usage**: Number of beds used, equipment utilization, nurse and doctor hours, etc.

Tasks:

For each of the data mining tasks listed below, determine whether the task is useful for the scenario described. Justify your decision briefly.

1. Description
2. Estimation
3. Prediction
4. Classification
5. Clustering
6. Association

Exercise number 2

For each of the following scenarios, identify which data mining task(s) (Description, Estimation, Prediction, Classification, Clustering, or Association) are most relevant. Provide a brief explanation of why you selected the task(s).

Question:

1. The New York Times wants to summarize the key trends in voter behavior during the last presidential election.
2. A tech company wants to predict the average monthly revenue for the next quarter based on historical sales data.
3. A bank wants to categorize customers into different credit risk groups (low, medium, high) based on their financial history.
4. A hospital wants to group patients with similar symptoms to identify patterns in diseases that haven't been diagnosed yet.
5. A retail store wants to find common combinations of products frequently bought together, such as "bread and butter."
6. A car manufacturer wants to forecast the likelihood of a specific car model being recalled based on its manufacturing characteristics.
