Data Mining
HW 1

Omid Naeej Nejad
610301189

October 10, 2024

1 Exercise 1

1. Description

   The description involves summarizing and visualizing data to understand its main characteristics. e.g. plotting clustered bar chart to check interaction between side effects or effectiveness of a drug against patient's demographics.

2. Estimation

   Estimating the length of hospital stay based on patient demographics and medical history can help in resource planning and management.

3. Prediction

   Predicting patient outcomes, such as recovery times required after a specific surgery or improvement in health after a specific treatment administered, based on patient's demographics and medical history.

4. Classification

   Categorize patients into different risk groups (low, medium, high) based on patient's demographics, medical history and diagnostic results.

5. Clustering

Clustering patients with the most similarity in diagnostic results, patient demographics and medical history to identify people are at risk of diseases, optimizing administered treatments and hospital resources.

6. Association

   Association identifies relationships between variables. For example, finding associations between certain treatments and patient outcomes can help in understanding the effectiveness of different treatment protocols and optimizing them for better results.

2 Exercise 2

1. Description: This task involves summarizing and visualizing the data to identify key trends and patterns in voter behavior. Description helps in understanding the overall voting trends, demographic influences, and other significant factors.

2. Estimation & Prediction: Estimation is used to predict a continuous value, such as average monthly revenue. This task involves analyzing historical sales data to estimate future revenue.

3. If the data points are labeled, Classification, If not, Clustering.

   Classification assigns data into predefined categories. In this case, customers are categorized into different credit risk groups based on their financial history, which helps in risk management and decision-making.

   Clustering similar customers together without predefined categories. (Number of Clusters = 3)

4. Clustering: Clustering groups similar data points together without predefined categories. This task helps in identifying patterns and subgroups of patients with similar symptoms, which can be crucial for diagnosing new diseases and treatments.

5. Association: Association identifies relationships between variables. This task involves finding common combinations of products that are frequently bought together, which can help in optimizing product placement and marketing strategies.

6. Prediction: Prediction involves forecasting future outcomes based on historical data. This task helps in predicting the likelihood of a car model being recalled, which can be crucial for quality control and risk management.