CI603 Data Mining

Tutorial 1

(solution)

1. Discuss whether or not each of the following activities is a data mining task.
   a. Monitoring the heart rate of a patient for abnormalities.

      Yes. We would build a model of the normal behaviour of the heart and raise an alarm when an unusual behaviour of the heart occurred.

   b. Dividing the customers of a company according to their gender.

      No. This is a simple database query.

   c. Predicting the outcomes of tossing a (fair) pair of dice.

      No. This is a probability calculation.

   d. Computing the total sales of a company.

      No. This is simple accounting

   e. Monitoring seismic waves for earthquake activities.

      Yes. In this case, we would build a model of different types of seismic waves behaviour associated with earthquake activities and raise an alarm when one of these different types of seismic activity was observed.

   f. Predicting the future stock price of a company using historical records.

      Yes. We would attempt to create a model that can predict the continuous value of the stock price.

   g. Extracting the frequencies of a sound wave.

      No. This is signal processing.

   h. Identifying the key attributes of clients attracted by a mortgage offer.

      Yes. In this case, we would identify the socio-demographic characteristics of these clients through exploratory data analysis to create profiles.

i. Increasing cross-selling to existing customers.

   Yes. In this case, we would create a recommendation system to offer additional products based on rules generated from an association analysis.

j. Sorting a student database based on student identification numbers.

   No. This is a simple database query.

k. Detecting fraudulent activity in telecommunication services.

   Yes. In this case, if we have predefined classes of fraud and historical data, we would use a logistic regression to distinguish between valid and not.

l. Classifying the current customers of a company according to their profitability.

   No. This is an accounting calculation, with application of a threshold.

2. Which statement is **NOT TRUE** regarding a data mining task?
   a. Clustering is a descriptive data mining task
   b. Classification is a predictive data mining task
   c. Regression is a descriptive data mining task
   d. Deviation detection is a predictive data mining task

3. Which data mining task can be used for predicting wind velocities as a function of temperature, humidity, air pressure, etc.?
   a. Cluster Analysis
   b. Regression
   c. Classification
   d. Sequential pattern discovery

4. Which of the following is not a data mining functionality?
   a. Characterisation and Discrimination
   b. Classification and regression
   c. Selection and interpretation
   d. Clustering and Analysis

5. Which data mining task is most suitable for a bank credit risk assessment for the granting of a mortgage?
    a. Neural network classification.
    b. Logistic regression.
    c. Clustering
    d. All of the above
    e. None of the above.

6. Which data mining task is most suitable for predicting the effect that current covid vaccines will have thirty years from now in the population under 12 who received the two doses in the UK.
    a. Logistic regression.
    b. Neural network classification.
    c. Clustering
    d. All of the above
    e. None of the above.

7. Which data mining task is most suitable for designing a supermarket layout to maximise sales?
    a. Classification.
    b. Clustering
    c. Affinity grouping.
    d. All of the above
    e. None of the above.