Data Analytics 774 & 874
Pre-block Assignment Department of Industrial Engineering

Deadline: 3 October 2021, 23:59
Full Marks: [150]

Instructions

Follow the following instructions when completing this assignment:

- Answer each of the questions below in your own words.
- This is not group work, and questions are to be answered by students individually.
- Submit your answers as a pdf document. Please name this report `????????xxxPreblockAssignment.pdf`, where you replace the question marks with your student number or your initials and surname, and replace `xxx` with 774 (PGDip students) or 874 (M.Eng students). If you are registered as an industry course student, then leave the `xxx` as is. Please note that only pdf documents will be accepted, and that all assignments have to be submitted via Sunlearn for 774 and 874 students and Sunonline for industry course students.
- Please make sure that your initials, surname, student number and email address are given on the first page of your pdf document.

Section A: Expert Systems [82]

The main focus of this module is on approaches to gain an understanding of data and trends within datasets, obtained via application of exploratory data analytics and data mining tools. A study of data analytics is not complete without exploring expert systems, as the first rule-based artificial intelligence system. Though experts systems are one of the oldest artificial intelligence tools for decision support, expert systems remain very relevant. The contact sections during the block week will not explore expert systems. You will explore expert systems as part of your pre-block studies. To guide you through this exploration, answer each of the following questions. In order to answer these questions, you are expected to search for relevant material using Google.

1. Define the following terms:
   (a) Expert system (2)
   (b) Knowledge engineering (2)
   (c) C Language Integrated Production System (CLIPS) (2)
   (d) Inferencing (2)
   (e) Knowledge base (2)
2. An expert system consists of different elements that work in conjunction to perform the functions of the system. Draw a diagram that illustrates the structure of a rule-based expert system and explain the function of each of the elements. (12)
3. Explain the concepts of forward chaining and backwards chaining. (4)
4. Discuss five advantages of expert systems. (5)
5. Discuss three disadvantages of expert systems. (3)

6. Define each of the following methods of inference:
   (a) Deduction (2)
   (b) Induction (2)
   (c) Abduction (2)
   (d) Heuristics (2)
   (e) Generate and test (2)
   (f) Analogy (2)
7. Describe what is meant by each of the following knowledge classes:
   (a) Procedural (2)
   (b) Declarative (2)
   (c) Tacit (2)
8. Uncertainty arise sometimes due to interaction between different rules about the same information, which then are not always compatible. Name five reasons why this conflict can occur. (5)
9. Explain how certainty factors are utilized with an expert system. (3)
10. When was the first expert system developed, and who developed it? (2)
11. What was the first expert system called, and in which domain has it been applied? (2)
12. Do you think expert systems still have a place in industry? Justify your answer. (5)
13. The following questions are related to knowledge-based artificial neural networks:
    (a) What is a knowledge-based artificial neural network? (3)
    (b) Decribe the role that a knowledge-based artificial neural network can play in expert systems (4)
    (c) Which of an expert system's disadvantages are addressed by a knowledge-base artificial neural network's integration with an expert system?
    (d) (6)

Section B: Online Analytical Processing and Data Mining [68]

The focus of this question is to guide you towards an understanding of online analytical processing (OLAP) and data mining. To answer these questions, it is expected that you will google to find answers to the questions.

1. Provide a definition of OLAP. (3)
2. What is an OLAP cube? (3)
3. The OLAP cube enable four basic types of multidimensional data analysis, namely drill-down, roll up, slice and dice, and pivot. Describe each of these, and provide an example for each. (12)
4. What are the main differences between OLAP and online transaction processing (OLTP)? (4)
5. What is data mining? (2)
6. What are the differences between OLAP and data mining? Illustrate the differences with examples. (6)
7. What are the main steps within a data mining process? Describe each of these steps. (12)
8. Discuss (shortly) whether or not each of the following activities is a data mining task:
   (a) Dividing the customers of a company according to their profitability. (2)
   (b) Predicting the outcomes of tossing a (fair) pair of dice. (2)
   (c) Understanding the profile of customers who default their loan within the first year of the loan. (2)
9. Describe the following to terms to make the differences between these two terms clear: Exploratory data analysis and data mining (or knowledge discovery from data). (6)
10. Discuss an application where data mining has been used successfully in any industry. In your discussion, name the industry, discuss the tools used and the data, the objectives, and the main findings. (8)
11. Discuss the role that feature selection plays within knowledge extraction approaches (6)