UNIVERSITY
OF SKÖVDE

School of informatics

WRITTEN EXAMINATION

Course: Data mining A1N

Examination

Course code: IT734A

Credits for written examination: 4.5

Date: 2024-03-01

Examination time: 08:15 - 11:30

Examination responsible: Addi Ait-Mlouk

Teachers concerned

Aid at the exam/appendices

Other

Instructions

- ☐ Take a new sheet of paper for each teacher.
- ☐ Take a new sheet of paper when starting a new question.
- ☐ Write only on one side of the paper.
- ☒ Write your name and personal ID No. on all pages you hand in.
- ☒ Use page numbering.
- ☒ Don´t use a red pen.
- ☒ Mark answered questions with a cross on the cover sheet.

Grade points: Each question is graded 0-10 points. To pass the exam, you need a minimum of 5 points on each question (more details on the next page).

Examination results should be made public within 18 working days

Good luck!

Total number of pages

Questions

- The exam has five questions, one for each course objective.
- Each question has sub-questions (a, b, c, ...)
- Each question is graded with up to 10 points.
- To pass a question, you need to have at least 5 points on the question.
- To pass the exam, you need to have passed all questions.
- The maximum number of points on the exam is 50.

Grading

If your score on any question is below 5 points, your grade will be U (Fail). If you have at least 5 points on each question, your grade is determined using the sum of points as follows:

Points	Grade	Percentage
45-50	A	90-100
40-44	B	80-89
35-39	C	70-79
30-34	D	60-69
25-29	E	50-59
0-24	F	0-49

A (Excellent), B (Very good), C (Good), D (Satisfactory), E (Sufficient) or F (Fail)

Don't forget to motivate all your answers!

Good luck!

Question 1

[Course objective: critically reflect and describe utility, problems and limitations of data mining]

a. What are the main challenges faced when dealing with unstructured data in data mining projects?
b. Describe the difference between supervised and unsupervised learning algorithms?

c. How does data preprocessing contribute to the success of data mining tasks, and what techniques are commonly employed in this stage?
d. Discuss the ethical considerations that arise in the context of data mining.
e. In what ways can data visualization techniques enhance the interpretation and communication of insights derived from data mining results?

Question 2

[Course objective: critically reflect and describe data mining algorithms within the classification, association analysis and cluster analysis, with respect to application and structure]

a. How does the Apriori algorithm work in association analysis, and what are its key components?
b. Explain the concept of overfitting in the context of classification algorithms, and discuss strategies to mitigate it.
c. Describe the K-nearest neighbors (KNN) algorithm for classification, and discuss its advantages and limitations.
d. What are the main steps involved in the Naive Bayes classification algorithm?
e. Discuss the concept of feature selection and its importance in optimizing the performance of classification models.

Question 3

[Course objective: implement and explain basic data mining algorithms]

Choose one classification algorithm presented in the course (e.g., Random Forest, Decision Tree, KNN, etc.) and answer the following questions:

a. How does the chosen algorithm make predictions for new data points?
b. Explain the role of individual features and decision rules in making these predictions.
c. Explain how you would evaluate the effectiveness of the chosen algorithm
d. How do different parameter settings influence the performance and characteristics of the algorithm?

e. Discuss the strengths and limitations of the algorithm's structure in terms of bias, interpretability, and handling complex data relationships

Question 4

[Course objective: identify and describe problems where data mining is relevant]

- Given the five following data mining problems, classify them as classification, regression or clustering problems. Motivate your answer.

a. Identifying fraudulent transactions in financial data using transaction history and behavioral patterns.
b. Analyzing customer feedback and sentiment to predict customer churn or loyalty in a subscription-based service.
c. Grouping news articles without predefined categories
d. Predicting student performance based on academic history and demographics
e. Classifying customer support tickets into different categories based on the reported issue and customer interaction history.

Question 5

[Course objective: select suitable data mining algorithms for solving such problems and analyze, compare and evaluate results]

a. What is text mining, and how is it used to extract insights from unstructured text data?
b. What are the key preprocessing steps in text mining, and how do they enhance analysis results?
c. Compare supervised and unsupervised learning in text mining, and give examples of their applications.
d. Explain the concept of a neural network and how it mimics the human brain in processing information.
e. What distinguishes convolutional neural networks (CNNs) from traditional neural networks, and what advantages do CNNs offer in image recognition tasks?