UNIVERSITY
OF SKÖVDE

School of informatics

WRITTEN EXAMINATION

Course: Data mining A1N

Examination

Course code: IT734A

Credits for written examination: 4.5

Date: 2025-03-13

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. State two key benefits of data mining and one major limitation.
b. Explain data mining's role in decision-making.
c. Identify a core advantage of data mining and assess data quality's impact.
d. How do scalability issues restrict data mining utility?
e. What computational challenges arise with large datasets?

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. Analyze the SVM algorithm's structure and its use in text categorization.
b. Evaluate the hierarchical clustering algorithm's structure and its scientific application.
c. State the purpose of logistic regression and its algorithmic structure.
d. Explain the utility of the Apriori algorithm in association analysis and its structural limitation.
e. Assess the k-means algorithm's structure and its use in market segmentation.

Question 3

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

a. Analyze the benefits and challenges of ensemble methods.
b. What is the process for implementing a recommender system?
c. What distinguishes hierarchical clustering from partition-based methods?
d. Describe the steps of a gradient descent algorithm and its role in optimization.
e. How do classification algorithms handle imbalanced datasets?

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.

  a. Determining public sentiment from social media posts about a new product launch.
  b. Forecasting quarterly sales revenue for a retail chain using past sales data and economic indicators.
  c. Grouping scientific research articles into groups based on similar topics.
  d. Estimating delivery times for online orders by analyzing distance, traffic conditions, and weather.
  e. Detecting anomalous network activity to identify potential cybersecurity threats.

Question 5

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

a. How do you compare the accuracy of classification models?
b. When would you favor decision trees over neural networks?
c. How does cross-validation support result reliability?
d. How do rare words and out-of-vocabulary terms affect text classification?
e. What ethical and privacy issues restrict the scope of text mining applications?