



UNIVERSITY
OF SKÖVDE

School of informatics

WRITTEN EXAMINATION

Course: Data mining A1N

Examination

Course code: IT734A

Credits for written examination: 4.5

Date: 2026-02-20

Examination time: 14:15 - 17:30

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 worth 2 points.
- 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 (10 Points)

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

- a. When does data mining provide real value, and when can it lead to poor decisions?
- b. How do data quality and data availability limit data mining projects?
- c. Evaluate the role of human judgment in validating and interpreting data mining results.
- d. Why does having more data not always improve data mining results?
- e. Why is continuous monitoring important for deployed data mining models?

Question 2 (10 Points)

[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 do constraints or thresholds affect the outcome of association analysis?
- b. Why do different clustering strategies produce different groupings on the same data?
- c. Why do some clustering algorithms struggle with large or high-dimensional datasets?
- d. How do ensemble methods improve prediction accuracy compared to single models?
- e. Describe k-means algorithm and MiniBatchKMeans.

Question 3 (10 Points)

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

- a. What is the main difference between classification and regression?
- b. How does a decision tree decide which attribute to split on?
- c. How do you choose the number of clusters in k-means?
- d. Why is data normalization important for distance-based algorithms?
- e. Detail the working mechanism of Convolutional Neural Network (CNN).

Question 4 (10 Points)

[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. Predict whether an email is spam or not spam
 - b. Estimate the price of a house based on its features (size, location, age)
 - c. Group customers based on their purchasing behavior
 - d. Predict whether a patient has diabetes based on medical tests
 - e. Forecast monthly sales for a store next year

Question 5 (10 Points)

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

- a. How can model evaluation metrics guide the choice between different classification algorithms?
- b. Describe methods to assess clustering quality when labels are unknown
- c. Compare word embeddings and traditional vectorization methods for text mining
- d. When would an ensemble method be preferred over a single model?
- e. What is the difference between TF-IDF and modern transformer-based embeddings?