



School of informatics

## WRITTEN EXAMINATION

Course: Data mining A1N

Examination

Course code: IT734A

Credits for written examination: 4.5

Date: 2026-01-07

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 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. What are the fundamental differences between descriptive and predictive data mining?
- b. Describe three major obstacles faced when working with real-world datasets in data mining projects.
- c. Explain how domain knowledge contributes to successful data mining outcomes.
- d. What role does data preprocessing play in the overall data mining workflow?
- e. Discuss potential biases that can emerge in data mining results and their societal implications.

### **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. Compare Random Forest and Decision Tree algorithms in terms of structure and performance characteristics.
- b. Explain the concept of support and confidence in association rule mining and their importance in evaluating rule quality.
- c. Describe the differences between partitional and hierarchical clustering approaches, providing examples of when each is most suitable.
- d. How do ensemble methods enhance the robustness of classification algorithms?
- e. What are the computational complexities associated with different clustering algorithms when dealing with large datasets?

### **Question 3 (10 Points)**

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

- a. Describe the mathematical foundation of Principal Component Analysis (PCA) and its role in dimensionality reduction.
- b. Explain the backpropagation algorithm in neural networks and how it enables learning from training data.
- c. How does the FP-Growth algorithm improve upon the Apriori algorithm for frequent pattern mining?

- d. Detail the working mechanism of Support Vector Machines (SVM) and explain the concept of kernel trick.
- e. Describe how DBSCAN clustering algorithm identifies clusters and handles noise points differently from K-means.

### **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. Analyzing customer purchase patterns to create targeted marketing segments without predefined categories.
  - b. Predicting the likelihood of equipment failure in manufacturing based on sensor readings and maintenance history.
  - c. Estimating the continuous value of carbon emissions from industrial processes using environmental factors.
  - d. Identifying different types of network intrusions from system log data with known attack categories.
  - e. Forecasting daily energy consumption for smart grid optimization based on historical usage and weather patterns.

### **Question 5 (10 Points)**

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

- a. What factors should guide the selection between supervised and unsupervised learning approaches for a given business problem?
- b. Explain the trade-offs between precision and recall in classification tasks and when each metric is more important.
- c. How do you validate the quality of clustering results when ground truth labels are not available?
- d. What are the common preprocessing steps in text mining?
- e. What is the difference between TF-IDF and modern transformer-based embeddings? Provide examples of when each approach is most effective.