Knowledge Discovery in Databases mit Übung

Student Demo in Summer Semester 2024

Prof. Dr. Richard Lenz

Matriculation No.
(fill from left to right!):

□0	□0	□0	□0	□0	□0	□0	□0
□1	□1	□1	□1	□1	□1	□1	□1
□2	□2	□2	□2	□2	□2	□2	□2
□3	□3	□3	□3	□3	□3	□3	□3
□4	□4	□4	□4	□4	□4	□4	□4
□5	□5	□5	□5	□5	□5	□5	□5
□6	□6	□6	□6	□6	□6	□6	□6
□7	□7	□7	□7	□7	□7	□7	□7
□8	□8	□8	□8	□8	□8	□8	□8
□9	□9	□9	□9	□9	□9	□9	□9

Example:

Correct coding for fictious matriculation no.

00708152

■0	■0	□0	■0	□0	□0	□0	□0
□1	□1	□1	□1	□1	■1	□1	□1
□2	□2	□2	□2	□2	□2	□2	■2
□3	□3	□3	□3	□3	□3	□3	□3
□4	□4	□4	□4	□4	□4	□4	□4
□5	□5	□5	□5	□5	□5	■5	□5
□6	□6	□6	□6	□6	□6	□6	□6
□7	□7	■7	□7	□7	□7	□7	□7
□8	□8	□8	□8	■8	□8	□8	□8
□9	□9	□9	□9	□9	□9	□9	□9

Matriculation No. (again, for counter-check):

....

First name and Surname (in block letters):

....

Room:

....

Seat number (if available):

....

Signature:

By signing this, you certify that the information you have provided above is correct and that you have fully read, understood, and will comply with the processing notices on page 2.

....

Notes of the exam supervisors (do not fill in):

- □ Vorzeitige Abgabe (Uhrzeit):
- □ Ausgetreten (Uhrzeit):
- □ Teilnahme unter Vorbehalt (Unterschrift Kand.):
- □ Täuschungsversuch (Uhrzeit):
- □ Sonstiges:

Status (do not fill in):

- □ Demo

Processing notices concerning the whole exam

- **Exam details:**
  - The duration of the exam is 90 minutes.
  - Check your examination copy for completeness (in total 4 pages).

- **Utilities:**
  - You may bring any number of writing utensils (pens, rulers, erasers).
  - You should bring at least one correction roller.
  - You can prepare a single hand-written sheet (paper format A4, only front page), any printing is forbidden. The sheet's contents are not subject of any restrictions.
  - Any other documents or utilities are forbidden.

- **Workplace:**
  - Bags and wardrobe must be stored near the workplaces.
  - Electronic devices are to be switched off and stored in their pockets before the start of the exam.
  - Have your photo ID ready at your seat (identity check).

- **Questions:**
  - Use only document-proof pens in shades of dark blue or black for your answers. We reserve the right not to consider answers in pencil for correction.
  - Read the task carefully and pay attention to negations like **not** or whether you should identify **correct** or **wrong** answers.
  - **Multiple choice questions**:
    * Fill in the boxes completely.
    * Use only correction rollers for corrections. Make sure that you cover the box completely with correction tape. If you want the box to count again, paint over the correction tape at the position of the original box.
    * The use of correction fluid and ink eraser is not permitted for technical reasons.
  - **Free-response questions**:
    * Try to answer all questions unambiguously. Answers that are obviously incorrect or needless can cause a deduction of points.
    * The answers to the questions of this examination must be given in English. Answers in other languages will not be considered.
    * Do not answer outside of the template boxes! If you need more space, you can use the back pages of your exam copy. However, take care that your pen's colour does not shine through to the front. In the original template box refer to the back page where your solution can be found.
    * Clearly cross out any answers that are not to be evaluated.

- **Other notices:**
  - Do not remove the retaining clip of your examination copy and do not manipulate the codes at the upper margin, the grading fields and the calibration symbols in every corner. Each manipulation will be graded as an attempt to deceive and will lead to failing the exam.
  - We collect your examination copy and your prepared cheat sheet at the end of the exam. You are not allowed to take any sheet with you! We evaluate only answers on the original examination copy!
  - If you have to cancel the exam for health reasons, inform us immediately. You will then receive a form and must obtain a medical certificate from a medical officer appointed by the university.

The entire KDDmUe team wishes you lots of success!
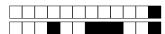
Demo Questions (Processing time approx. 10.5 min)

Question 1
Given is the outline of the **data mining pipeline** as seen by the machine learning and statistics community:

For each of the following steps, mark **which step** (1st-3th) of the data mining pipeline it is.
Mark one applicable answer per step (3 completely filled boxes in total).

☐ 1 ☐ 2 ☐ 3   **Step:** Data mining

☐ 1 ☐ 2 ☐ 3   **Step:** Post processing

☐ 1 ☐ 2 ☐ 3   **Step:** Data preprocessing

Question 2
For the following statements, mark whether they are **True** or **Wrong** with regard to unclean data.
Mark one applicable answer per statement (4 completely filled boxes in total).

☐ **True** ☐ **False**   An attribute named *Nme* is considered an incomplete attribute due to the missing *a*.

☐ **True** ☐ **False**   Naming all persons *Doe* is in most cases considered intentional.

☐ **True** ☐ **False**   Noise and outliers are in most cases exactly the same.

☐ **True** ☐ **False**   Using two grading schemes (e.g. A and 1) in one relation is an inconsistency.

Question 3
Which of the following methods are methods mining **closed itemsets** in specific?
Mark all applicable answers (1 - n completely filled boxes).

☐ FPGrowth   ☐ ECLAT   ☐ CLOSET   ☐ A Priori

Question 4
Given is a **mystery helper** function:

```
def mystery_helper(itemsets, minimal_support_count):
  itemset_list = ItemsetList()
  itemset_list.itemsets = [
    x for x in itemsets.itemsets if x.occurrence_count >= minimal_support_count
  ]
  return itemset_list
```

In which of the following **algorithms** is this function most likely to **be used**?
Mark one applicable answer (1 completely filled box).

☐ Z-Score normalization   ☐ RainForest
☐ A Priori   ☐ CHAMELEON

Question 5

Given is an initial vector:

$$X = (2, 2, 0, 2, 3, 5, 4, 4)$$

Use the discrete wavelet transform to compute the wavelet coefficents of this initial vector.

Only for grading (do not fill in!): □0 □1 □2 □3 □4 □5 □6 □7 □8