

## Course Analysis

### ID2222 Data Mining (Datautvinning)

#### 2023/2024, period 2 (HT23)

#### Course information

<b>Course:</b>	ID2222 Data Mining (Datautvinning) <a href="https://www.kth.se/student/kurser/kurs/ID2222?l=en">https://www.kth.se/student/kurser/kurs/ID2222?l=en</a>
<b>Credits:</b>	7,5 hp
<b>Level:</b>	Second cycle
<b>Requirements:</b>	LAB1 - Programming Assignments, 3.0 credits, grading scale: P, F TEN1 - Examination, 4.5 credits, grading scale: A, B, C, D, E, FX, F
<b>Academic Year:</b>	2023/2024
<b>Period:</b>	period 2 (HT23)
<b>Course coordinator:</b>	Vladimir Vlassov, Professor, <a href="mailto:vladv@kth.se">vladv@kth.se</a> , tel. 08-790 4115
<b>Examiners:</b>	Sarunas Girdzijauskas, Professor, <a href="mailto:sarunasg@kth.se">sarunasg@kth.se</a> , tel. 08-790 4175 Vladimir Vlassov, Professor, <a href="mailto:vladv@kth.se">vladv@kth.se</a> , tel. 08-790 4115
<b>Teaching Assistants</b>	Ahmad Al-Shishtawy, Lecturer, <a href="mailto:ahmadas@kth.se">ahmadas@kth.se</a> , 08-790 4242 Sameen Mansha, PhD student, <a href="mailto:sameen@kth.se">sameen@kth.se</a> Fabian Schmidt, PhD student, <a href="mailto:fschm@kth.se">fschm@kth.se</a> Sina Sheikholeslami, PhD student, <a href="mailto:sinash@kth.se">sinash@kth.se</a>
<b>Language:</b>	English
<b>Canvas (VT24):</b>	<a href="https://canvas.kth.se/courses/42990">https://canvas.kth.se/courses/42990</a>

#### Quantitative Data

<b>Total number of registered students:</b>	<b>124 students</b>
<b>The number of students completed the course</b>	<b>50 students (74 % of 124 st)</b> Grade Statistics (of 50 st): A: 12 (24%); B: 10 (20%); C: 11 (22%); D: 11 (22%); E: 11 (12%)
<b>Exam statistics</b>	<b>103 students (83 % of 124 st) attended the exams</b> <b>98 students (95 % of 103 st) passed the exam</b> Grade Statistics (of 103 st): A: 39 (38%); B: 17 (17%); C: 12 (12%); D: 16 (16%); E: 14 (14%); F: 5 (5%)

#### Course activities

- 18 lectures; 5 weekly homework programming assignments
- All course activities, namely lectures, reporting and help sessions, and the written exam, were conducted on the KTH Kista campus (Electrum, Kistagången 16, Kista);

#### Examination:

The examination consists of a written examination (TEN1) and programming assignments (LAB1). The final grade for the course is based on the performance of the written exam and the programming assignments (bonus points). Assignments graded P/F are to be done in groups of 2 students and presented in person at specially appointed reporting sessions in Electrum, Kista. The ID2222 exam is a proctored on-campus computer-based closed-book exam in Canvas. The exam consists of questions of different types, e.g., Multiple Choice, Multiple Answer, True/False, and Numeric.

### Bonus Policy

- Each homework assignment is awarded three bonus points (+1 point for an optional task) if submitted on time (i.e., before or on the due date), presented within a week after the deadline, and accepted (passed). The number of bonus points can be reduced for errors or inefficient solutions.
- Assignments (homework and the project) turned in after the corresponding deadlines will not be granted bonus points.

### ***Changes since the previous occasion of the course in the 2022/2023 academic year, HT22***

- Revised some lectures (e.g., Introduction, Frequent Itemsets, Link analysis), mostly to include up-to-date material, fix a few typos, and improve presentation.
- The new lecture, “Data Stream Processing and Analytics: Introduction to Apache Spark Streaming,” has been developed by Ahmad Al-Shishtawy to replace the guest lecture on Apache Flink.
- The lecture on PageRank was redeveloped and expanded.

### ***Proposals of Concrete Improvements for the Next Course Occasion***

- **Homework programming assignments:** Consider revising the third homework on Mining data streams, namely updating the relatively old papers for the homework task.
- **Videos:** If time allows, pre-record new versions of the lecture videos or record the lectures in the process of the course.

### ***Comments by Students:***

- It is difficult to start with the assignments because of a lack of experience in Apache Spark. It would be good to allow the implementation of programming assignments in a language and a programming environment of our choice.