

**Data.ml.320 Knowledge Mining and Big Data (5 cr)**

Exam 8.5.2024 / Ari Visa

Please, do not forget the Kaiku-feedback!

<p>1. Define terms: Data warehouse, Knowledge Mining, and Big Data</p> <p>An example of a definition: <b>Data:</b> Facts and things certainly known. Data are any facts, numbers, or text that can be processed by a computer.</p>	6p
<p>2. What is HADOOP? Are there alternatives to HADOOP? How do you use HADOOP when solving a clustering problem? What is the relation between cloud computing and HADOOP?</p>	6p
<p>3. What do you know about associative analysis?</p>	6p
<p>4. What is the difference between classification and prediction? You have 2 labeled samples of type <math>\mathbf{X} = (x_1, x_2, x_3)^T</math>, consisting of real numbers. You should make a predictive model. What kind of model do you use? Motivate your answer! What is the robustness of your solution?</p>	6p
<p>5. How do you define cluster analysis? How can you estimate the number of clusters? Motivate your answer. You have <math>1T (=10^{12})</math> samples of high dimensional data (dimension <math>&gt; 100</math>) available. What kind of clustering method do you use? Motivate your answer! How does the computer architecture influence your proposal?</p>	6p