

SGN-43006 Knowledge Mining and Big Data (5 cr)

Examination 2.11.2017 / Ari Visa

A: No Calculator

The questions are returned.

Please, do not forget the Kaiku-feedback!

1. Define terms: Data Mart, Data Mining, and Big Data	6p
2. What is HADOOP? What is the relation between HADOOP and MapReduce algorithms? How does MapReduce algorithm work? Are there any limitations when using HADOOP? Are there other techniques like HADOOP? If answer is yes then give an example.	6p
3. How to define association rule mining? List the types of associations! What is the difference between causality and correlation? If you have 100 Tbytes data available how do you start your association analysis?	6p
4. What is the difference between classification and prediction? You have 100 Gbytes process data available. Note, that you do not have the exactly right answers available. Then you have 1 Mbyte labeled data. You should make a process model. Is it possible? Motivate your answer!	6p
5. How do you define cluster analysis? What are the main players that define the quality of clustering. You have 1 Kbyte data that should be clustered. The dimensionality of the data is high. Which clustering approach do you choose? Motivate your answer!	6p