

University of Turku
Dept. of Information Technology

Data Mining
Examination Aug 29, 2011

1. Present a few real-life applications of data mining.
2. Explain the construction of a **naïve Bayesian classifier**, based on some training dataset. Discuss the handling of attributes which are (a) numeric, (b) nominal. Present the usage of the classifier to classify a new data instance.
3. Define the meaning of **association rules** in data mining. Define the concepts of **coverage** and **accuracy** of the derived rules. Mention some algorithmic approach to mine association rules efficiently from large datasets.
4. What does **validation** mean in the context of data mining? Explain the functioning of the following validation approaches:
 - (a) repeated holdout
 - (b) cross-validation
 - (c) leave-one-out validation