T-809-DATA **DATAMINING AND MACHINE LEARNING** **8 ECTS**

Year of study:	First year MSc.
Semester:	Fall.
Level of course:	5. Second cycle, intermediate.
Type of course:	Core for all MSc programs in engineering.
Prerequisites:	No mandatory prerequisites. Basic knowledge of calculus, linear algebra, statistics and programming and statistics is assumed.
Schedule:	Runs for 12 weeks - 6 teaching hours a week.
Supervisor:	Jón Guðnason.
Lecturer:	Jón Guðnason and Eyjólfur Ingi Ásgeirsson.

Learning outcome:
Knowledge: After the course the students should be able to recall, describe and define, the following terms: Pattern recognition system, classifier design cycle and learning. Statistical pattern recognition, Bayesian decision theory, maximum likelihood and Bayesian parameter estimation. Linear models for classification. Principal component analysis. Multilayer neural networks. Nonparametric methods: k-nearest neighbours and Parzen kernels. Kernel methods and support vector machines. Unsupervised classification, K-means clustering, Gaussian mixture models and expectation maximization. Combination of classifiers, bagging and boosting.

Skills: After the course the students should be able to apply the data mining methods and implement the machine learning algorithms presented in the course using standard programming languages such as Python or Matlab and software packages such as scikit-learn and Weka.

Competence: After the course the students should be able to design a suitable machine learning algorithm for a real world problem, evaluate its performance, compare different designs and implementations and interpret the results. The students should also be able to present findings and new results in the subject.

Content: Pattern recognition system, classifier design cycle and learning. Statistical pattern recognition, Bayesian decision theory, maximum likelihood and Bayesian parameter estimation. Linear models for classification. Principal component analysis. Multilayer neural networks. Nonparametric methods: k-nearest neighbours and Parzen kernels. Kernel methods and support vector machines. Unsupervised classification, K-means clustering, Gaussian mixture models and expectation maximization. Combination of classifiers, bagging and boosting.

Reading material: Christopher Bishop. *Pattern Recognition and Machine Learning.* Springer Science and Business Media. 2006. Richard O. Duda, Peter E. Hart, David G. Stork. *Pattern Classification,* 2nd edition, John Wiley and Sons. 2001.
Teaching and learning activities: To be introduced at the beginning of the semester.
Assessment methods: To be introduced at the beginning of the semester.
Language of instruction: English.

All course descriptions may be subject to change. Revised information on the course schedule, reading material, teaching and learning activities, and assessment methods will be introduced in the learning management system Canvas at the beginning of the semester.