**CDSC 608 – Applied Machine Learning (7 ECTS)**

**Aims**

On completion of the course students will be expected to:

* Have a good understanding of the two numerical approaches to learning (optimization and integration) and how they relate to maximum likelihood and the Bayesian approach.
* Have an understanding of how to choose a probabilistic model to describe a particular type of data.
* Know how to evaluate a learned model in practice.
* Understand the role of machine learning in massive-scale automation.
* Have a good understanding of the problems that arise when dealing with very small and

**Course outline**

1. **Machine learning (ML) Basic Concepts**

* Supervised ML
  + Classification
  + Regression
* Unsupervised Learning
  + Clustering
  + Dimensionality Reduction

1. **Data Preparation and Feature Engineering**

* Data Acquisition
* Data Preprocessing
* Methods to impute missing values
* Outlier/Anomaly Detection
* Feature Engineering
* Feature Selection
* Overfitting/under-fitting
* bias/variance trade-off
* Learning Curve
* Mean Removal
* MinMax Scaling
* Binarization
* Label Encoder

1. **Model Selection and Evaluation**

* Performance Metrics
* Cross-validation
* Hyperparameter Tuning
* Model Selection

1. **Regression**

* Introduction
* Cost function and Gradient Descent
* Basic idea: Regression and its applications
* Linear regression
* Types of Errors and Better regression models
* Polynomial Linear Regression
* Regularization (Rigde and LASSO regression)

1. **Classification**

* Introduction
* Types and applications of classification algorithms
* Logistic regression
* SVM
* Random Forest
* k-Nearest Neighbors
* Naive Bayes
* Decision trees
  + Gini Index and Information Gain
* Ensembe techniques (XGboost, AdaBoost and RankBoost)

1. **Unsupervised Machine learning**

* Why unsupervised learning and its Applications
* Clustering algorithms
* K-means clustering
* Dimensionality Reduction
* Principal Component Analysis (PCA)

1. **Neural Networks**