|  |  |
| --- | --- |
| **Module description/Learning goal** | DS II: Introduction to Machine Learning  This module introduces students to the fundamental concepts and techniques in Machine Learning. These concepts are of three kinds. First, it will introduce students to the different types of machine learning algorithms; supervised vs. unsupervised and how they can be applied. Second, the module introduces students to different types of output: regression versus classification. Lastly, the module will focus on applying and evaluating machine learning models using the Python programming language. |
| Code: | DS II |
| Title: | DS II: Introduction to Machine Learning |
| Percentage: | 26 |
| Learning Objectives: | 1) The student will gain an understanding of fundamental machine learning techniques for data analysis, specifically covering the following topics:   1. Linear regression 2. Logistic regression 3. Decision trees 4. K-means Clustering 5. K-Nearest-Neighbour 6. Neural Networks: Perceptron   2) The student learns to design and implement machine learning algorithms, specifically covering the following topics:   1. Thinking about machine learning problems: Why, what for and where to use machine learning? 2. Different types of output: Regression vs. classification 3. Different types of machine learning: Supervised vs. Unsupervised 4. Model evaluation: A) Train set vs. Validation set vs. Test set, B) Accuracy, Recall, Precision, C) Bias-variance trade-off |
| Intended Learning Outcome: | Design and Prototyping & Technical Knowledge and Analyses, DD 1,2, 3:  **The student understands different types of basic machine learning techniques for data analysis and can design and implement these models to solve a given use-case. Specifically, he/she can:**   * Form a clear data-driven research question which can be answered using a predictive or prescriptive analyses based on a machine learning technique; SDSDA04, SDSDA05 * Apply a variety of data analytics techniques (Supervised machine learning) for data analysis. SDSDA01 * Explore a solution to your research question by: fitting and evaluating a given model (e.g. train, test, iterate, validate); SDSDA01, SDSDA09 * Evaluate the model using a suitable performance metric given one of the covered quantitative techniques; SDSDA09 * Present a well-written/illustrated answer using appropriate data visualisations and tables for the given use-case and presenting format; SDSDA04, SDSDA05, SDSDA06   Note: The [Edison Data science framework](https://edison-project.eu/edison/edison-data-science-framework-edsf/) was used to create these learning outcomes. Only Data Science Analytics and Research Methods and Project Management competencies were included. Having an S at the start of a code means applying competencies is required here, a K at the start of a code means only a knowledge competency level is required compared to its counterpart code (e.g. ***S***DSDA04 includes knowledge level of ***K***DSDA04). |
| Pre-requirements: | DS I |
| Approach: | Students will be taught using the github platform where they can access self-study material and Datalab material produced or curated by the lecturers.  2 days a week of which:  Day 1) self-study day of 8 hour course day, structured as follow:   * Github instructions/Video-lecture for presenting and discussing the main concepts, using practical examples and online material: 30 – 1.5 hours * Interactive Workshop, at home or during Data Lab with lecturer support: 4 - 6 * Interactive Mock Assessment: create building-blocks, at home or in Data Lab: 1 – 3 Hours   2) Data lab day of 4 hours: Mini-project to apply material learned in self-study day to a use-case (Oosterhout Dataset). |
| Topics by week: | |  |  |  |  | | --- | --- | --- | --- | | **Week** | **Day** | **Lecturer** | **Module Topics** | | 2 | 4 | Bram | Introduction to Machine Learning:   * Predictive and Prescriptive analyses; an overview of methods and analytical techniques. * [Codecademy Machine Learning Course](https://www.codecademy.com/learn/machine-learning): Introduction to Machine Learning | | 3 | 4 | Bram | Regression:   * [Linear Regression](https://www.codecademy.com/learn/machine-learning) * [Multiple Regression](https://www.codecademy.com/learn/machine-learning) * Performance Metrics: B) Accuracy, Recall, Precision | | 3 | 5 | Bram | Create a regression-based predictive analysis – Half a day   * [Yelp Regression Project?](https://www.codecademy.com/learn/machine-learning) | | 4 | 4 | Nitin | Classification   * [Classification versus regression](https://www.codecademy.com/learn/machine-learning) * [Classification: K-Nearest Neighbors](https://www.codecademy.com/learn/machine-learning) | | 4 | 5 | Nitin | Create a classification based predictive-analysis – Half a day: [Breast Cancer Classifier](https://www.codecademy.com/learn/machine-learning) | | 5 | 4 | Bram | * [Logistic Regression](https://www.codecademy.com/learn/machine-learning) * [Decision Trees](https://www.codecademy.com/learn/machine-learning) | | 5 | 5 | Bram | Create a classification-based prescriptive analysis – Half a day: [Predicting Income with Random Forests](https://www.codecademy.com/learn/machine-learning) | | 6 | 4 | Nitin | Clustering:   * [K-Means](https://www.codecademy.com/learn/machine-learning) | | 6 | 5 | Nitin | Create a classification based-predictive analysis – Half a day: [Handwriting Recognition using K-means](https://www.codecademy.com/learn/machine-learning) | | 7 | 4 | Irene | Neural Networks:   * [Perceptron](https://www.codecademy.com/learn/machine-learning) | | 7 | 5 | Irene | Model Evaluation & Iteration:  A) Train set vs. Validation set vs. Test set, B) Bias-variance trade-off | | 8 | 3 | - | Business Case Presentation | | 8 | 4 | - | Business Case Presentation | | 8 | 5 | - | Business Case Presentation | |
| Assessment: | Design a predictive or prescriptive model for a given use-case using a machine learning algorithm. Apply, iterate upon and evaluate the model. Subsequently present your work by:   * 2-weekly evaluation - formative assessment: to evaluate students’ progress and models created in the datalab projects. * Final evaluation - summative assessment: Present your final model in an appropriate text, tabular and/or graphical visualisation by integrating this into the dashboard\* with the appropriate contextual information. Note that if dashboard integration fails you provide a ZIP file which contains a document which can refer to the files containing your model output on github instead.     The use of Python is encouraged, but students are free to use R to complete (parts of) the material for final assessment. |
| Literature/Resources: | * Python and data analytics libraries (pandas, numpy, mathplotlib, scipy, scikit-learn, etc.) * R and data analytics libraries (cran, ggplot2, dplyr, reshap2, etc.) * https://www.codecademy.com/courses/machine-learning * An introduction to Statistical Learning: <https://www.statlearning.com/> |
|  |  |