1. **Name of Course:** Introduction to Machine Learning

2. **Instructor: Petra Kralj Novak**

3. **No. of Credits (US/ECTS): 4 ECTS**

4. **Academic term:** *Winter*

5. **Course level:** *BA / BSc*

6. **Relationship with other courses**: A prerequisite for this course are basic programming skills (e.g. Introduction to Programming in Python) and applied statistics knowledge (e.g. Fundamentals of Data Analysis).

7. **Course type:** *Mandatory*

8. **Name of the module in which the course is offered:** Data Science and Society

9. **Short description and the overall aim of the course:**
   This course aims to introduce students to basic principles of machine learning algorithms. In particular, we will look into basic principles of supervised and unsupervised machine learning through exemplars of algorithms of different families: heuristic search (decision trees), probabilistic classifiers (Naïve Bayes), lazy learning (KNN), methods for numeric prediction, association rule mining (Apriori), clustering, and fundamental principles of neural network architectures. Finally, we will look at mining unstructured data on a text example. Throughout the course, model evaluation and interpretation of results will be emphasized.

10. **The learning outcomes of the course:**
    After completing this course, the students will understand the basic theory underlying supervised and unsupervised machine learning. They will be able to formulate machine learning problems corresponding to different applications. They will understand a range of machine learning algorithms along with their strengths and weaknesses. They will be able to apply machine learning algorithms to solve problems of moderate complexity, evaluate their performance and properly interpret the results.

11. **Learning activities and teaching methods:** Lectures, hands-on exercises, project work.

12. **Assessment:**
    - Homework 20%
    - Project (presentation + report) 30%
    - Written exam 50%

13. **Course contents:**
    1. What is Machine Learning?
       - CRISP-DM
       - Basic classification example
       - Applying and evaluating decision trees
       - Tools: SciKit Learn and Orange

2. Classification with decision trees
    - Heuristic search, A*
    - Entropy and information gain
    - TDIDT, ID3
    - Confusion matrix, classification accuracy

3. Decision rules, classification evaluation
    - Decision tree pruning
    - Evaluation methods: cross validation, leave-one-out
    - The covering algorithm CN2

4. Probabilistic classification and classification evaluation
    - Naïve Bayes classifier
    - Cost-sensitive and unbalanced datasests
    - Precision, recall, F1-score, ROC, AUROC

5. Numeric prediction:
    - Linear regression,
    - Regression trees,
    - Model trees,
    - k-NN,
    - Numeric prediction evaluation: MAE, MSE, RMAE, RMSE, rRMSE, correlation coefficient

6. Clustering:
    - Hierarchical agglomerative clustering

7. Clustering 2:
    - K-means clustering
    - Evaluation and silhouette coefficient

8. Association rules, market basket analysis
    - Apriori algorithm
    - Support, confidence

9. Introduction to text mining

10. Introduction to neural networks
    - Perceptron, backpropagation idea
    - Use case: MNIST

11. Written exam

12. Seminar presentations

14. Further items:

The course will heavily rely on the book: Bramer, M., 2020. *Principles of Data Mining*. Springer, London.
Material available on Moodle.
Technical requirement: A computer with Python, ScikitLearn (https://scikit-learn.org/) and Orange (https://orangedatamining.com/)