



Middle East Technical University



Department of Computer Engineering

# CENG 414 Introduction to Data Mining

Spring 2023 - Assignment 2

Due: 7 May 2023 23:55

Submission: **via ODTUClass**

---

## 1 Overview

In this assignment, you are going to use Weka 3.8 <sup>1</sup> to do some tasks on a dataset. The aim is to make you familiar with certain machine learning algorithms and Weka. Weka is a tool that has collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset through Weka desktop application or they can be called from your own Java code.

## 2 Tasks

You will use the German Credit Risk dataset that is provided for you as an attachment in Odtuclass. Each entry in the dataset represents a person who takes a credit by a bank. Each person is classified as "good" or "bad" credit risks according to the set of attributes. Therefore, you are expected to classify instances according to their credit risks. You can find more information about the dataset in the following link. However, please download the attached dataset as it is slightly different from the one in the link <sup>2</sup>.

### 2.1 Preprocessing (20 Points)

In this section, you are expected to perform data transformations such as handling the missing values in the dataset. The mini-tasks are explained in the "ceng414\_hw1.ipynb" file. You must write your solutions in "ceng414\_hw1.ipynb". At the end of this phase, you must save the final dataset as "credit\_wo\_na.csv".

---

<sup>1</sup>[https://waikato.github.io/weka-wiki/downloading\\_weka/](https://waikato.github.io/weka-wiki/downloading_weka/)

<sup>2</sup>[https://archive.ics.uci.edu/ml/datasets/statlog+\(german+credit+data\)](https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data))

## 2.2 Multi-Layer Perceptron (35 Points)

For this task, you will use the "credit\_wo\_na.csv" dataset on Weka. Under the Classify tab in explorer window, choose MultilayerPerceptron classifier. Report your results under 5-fold cross validation. You will run the classifier with the default parameters and note them. Answer the following questions according to the run:

1. How many hidden layers and hidden nodes created?
2. Did Weka normalize the attributes? What is the effect of normalizing the attributes?
3. Which halting strategy did MLP use?

## 2.3 Decision Tree (25 Points)

Open the explorer in Weka GUI and open the "credit\_wo\_na.csv". Go to Classify tab and choose J48 classifier under trees. Report your results under 5-fold cross validation. Execute the classifier without changing default parameters. Besides, you should express the pruned tree, Summary and Detailed Accuracy By Class. In addition, put the visualization of the tree in your report.

## 2.4 Naive Bayes (20 Points)

Naive Bayes is a simple yet powerful machine learning algorithm used for classification tasks. It is based on Bayes' theorem and assumes that the features are conditionally independent given the class label. Open the explorer in Weka GUI and open the "credit\_wo\_na.csv". Go to Classify tab and choose NaiveBayes classifier from the list of available classifiers.

## 3 Submission

You are expected to submit a zip file which includes the following two documents:

- "ceng414.hw1.ipynb" file: You are expected to perform data transformations given in this file and submit your own implementation.
- Report: You are expected to assess the performance of these classifiers in your report including accuracy, precision, recall, and F1-measure. You can access these metrics from the "Result list" panel in the "Classify" tab. Besides, you need to answer the additional questions in Section 2. Your report must not exceed 3 pages. You must submit your report in pdf format.

## 4 Tutorials

- [Pandas](#)
- [Jupyter Notebook](#)
- [Weka](#)

## 5 Regulations

- Submission will be done via ODTUClass. You are expected to submit a zip file containing your code and report presenting the analysis result.
- Late submission is not allowed.
- We have zero tolerance policy for cheating. People involved in cheating will be punished according to the university regulations.