

UCCD2063 Artificial Intelligence Techniques

Group Assignment

June 2024

1. General Instructions

The general guideline for this assignment is as follows:

1. The total mark for the assignment is 100 and contributes **20%** to the total grade.
2. This is a **group assignment**. Each group should have **3 students** (no need to be in the same practical group).
3. Each group has to submit ONE **Zip file** containing the **report (.pdf)** and **code (.ipynb)** to WBLE before the deadline. Name your zip file using your group number (e.g. G01.zip)
4. The deadline for the submission of reports is on **September 8 (Sunday) 10pm**. Late submissions may be subjected to a penalty of up to 50% of the total assignment mark.
5. Evidence of plagiarism will be taken seriously and University regulations will be applied fully to such cases, in addition to ZERO marks being awarded to all parties involved.

2. Introduction

The objective of this assignment is to train machine learning models that can accurately predict the cardiovascular risk (low, medium, high) of an individual, using the information provided in the “*dataset.csv*” dataset. The results can be helpful for cardiovascular disease diagnosis and prevention.

The purpose of the assignment is to let students have hands-on experience solving real-world problems using machine learning techniques. Students will learn to perform data exploration, pre-processing, model selection, model training, and tuning of the learning algorithm in order to achieve better prediction accuracy.

3. Requirements

The following are requirements as well as suggestions for things to consider in your assignment.

3.1 Dataset

The “*dataset.csv*” dataset consists of 17 features (an individual’s demographic information, eating habits, physical activities etc.) and one prediction output for three levels of cardiovascular risk. The data contain certain complexity with missing data and noise. You should study the dataset carefully to explore the data in the dataset, study the features and propose what data preprocessing steps are needed to process the dataset so that it is suitable for the prediction task.

3.2 Experiment

Since the objective of this assignment is for you to learn, your experiments should not only run on one prediction algorithm with default setting. Instead, you should explore and find the best model for the task, and should also include analysis of the models by testing different kind of settings. For example, you can show how changing hyperparameters or training strategies affect your prediction results, or show whether a model underfits/overfits the data and the effect of regularization. You can also show the effect of feature selection on your prediction accuracy, and apply appropriate validation and testing steps to verify your results.

3.3 Coding

You must use **Python 3.x** and related packages in your implementation. You may implement any learning algorithms of your choice, not limited to those you have learned in class and make sure you understand how they work. You should:

- a. Implement three or more learning algorithms and compare their performances.
- b. Analyse the results and optimize the performances of your models for the dataset.

Your program must be structured into several parts (e.g. preprocessing, visualizing, training and validation, tuning and testing) and well commented.

Python code must be submitted in the form of Notebook file (.ipynb). Your codes must be error free so that the instructor is able to test the program and get the same results as indicated in your report.

3.4 Report

Your report should include the followings:

- A **title page** stating the title of the assignment, student names and student IDs, and a list of the effort and contribution of each team member in terms of percentage as follow.
Marks will be adjusted based on a team member's contribution.

For example:

Student Name:	Student 1	Student 2	Student 3
Student ID:			
Contribution:	30%	30%	40%
Signature:			

- **Suggested chapter structure:**
 1. Introduction
 - Background: describe about the problem you want to solve, the approaches for such a problem, etc.
 - Objectives: the purposes of your project.

2. Methods
 - Dataset description: describe the dataset.
 - Data exploration and visualization: explore the content of the dataset to better understand the data.
 - Data pre-processing: perform pre-processing and cleaning operations on the dataset, provide justification for your implementation.
 - Model selection: select suitable models for the dataset and provide justification for your selection. Give brief explanation of how the models work.
 - Model training and validation: perform model training and validate the performance of the models.
 - Model tuning and testing: fine tune the models and perform final testing.
3. Results and Discussion
 - Summarize the training and testing results.
 - Perform in-depth analysis of the prediction performance and errors of the models.
 - Compare the performance between different models using tables and figures.
 - Discuss their strengths and weaknesses.
 - Investigate what features are important for the prediction, etc.
4. Conclusions
 - Summarize what have been done and the findings.

Note: do not simply put screenshots of your notebook file outputs and codes in the report.

4. Marking Criteria

The grading of your assignment will be based on the following criteria:

Marking Criteria	
Criteria	Weight
Introduction and problem description	5%
Data exploration and visualization	10%
Data pre-processing	10%
Model selection and justification	5%
Model training and validation	20%
Performance analysis and fine-tuning	20%
Result discussion and findings	20%
Report format and writing	5%
Coding (comments, program organization, etc.)	5%
Total:	100%