

UCCD2063 Artificial Intelligence Techniques

Group Assignment

June 2023

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 (Friday) 5pm**. 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

In this assignment, you are required to implement a model that performs classification prediction for the given “*dataset.csv*” dataset. The objective is to generate machine learning models that can accurately predict whether a car driver will actually stop by a local store when he/she hears a promotion advertisement from the store on the car radio while driving, using the information provided in the dataset. The results can be helpful for determining the effectiveness of in-vehicle radio advertising.

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, and tuning of the learning algorithm in order to achieve better classification 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 21 features (driver’s demographic information, driving conditions, preferences, store promotion information etc.) and one prediction output. The data contain certain complexity with missing data. You should study the dataset carefully to explore the data in the dataset, and to 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 classification algorithm with default setting, but it should also include analysis of your learning algorithm by testing different kind of settings. For example, you can show how changing hyperparameters or training strategies affect your classification 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 as long as you understand how they work. Higher marks would be allocated for assignments that:

- Implement learning algorithms not taught in class AND you understand the algorithms that you are using.
- Implement two or more learning algorithms and compare their performances.
- 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). Must provide clear instruction on how to run your code and get the desired results as indicated in the report. Make sure your codes are error free so that the instructor is able to test your program and get the same results as 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 allocated 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:**
 - 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 training and validation: select suitable models for the data and perform training and validation, provide justification for your model selection and implementation. Give brief explanation of how the models work if necessary.
 - Model tuning and testing: fine tune the models and perform final testing.
3. Results and Discussion
 - Experimental results: summarize the training and testing results, perform in-depth analysis of the classification performance and errors of the tested models.
 - Discussion: compare the performance between different models and discuss about their strengths and weaknesses, investigate what features are important for the prediction, etc.
4. Conclusions
 - Summarize what have been done and findings.

4. Marking Criteria

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

Marking Criteria	
Criteria	Weight
Dataset and problem description	5%
Data exploration and data preparation	20%
Model training and validation	20%
Performance and error analysis	20%
Discussion & conclusion	10%
Report organization and results presentation	10%
Effort (e.g. In-depth analysis, self-learnt learning algorithm)	10%
Coding style (variable naming, comments, formatting, etc.)	5%
Total:	100%