

COSC2753 - Machine Learning

Assignment 2 – 2024A

Assessment Type	Group assignment. Submit online via Canvas → Assignments → Assignment 2. Marks awarded for meeting requirements as closely as possible. Clarifications/updates may be made via announcements/relevant discussion forums.
Due Date	Week 11, Sunday 19 May 2024, 23:59 pm Late submission: 20%/day, until 23 May 2024, 23:59 pm
Length	Report of no more than 05 pages plus up to 2 pages for appendices (excluding Cover Page, References List), written in font size 11, and source code + models
Group	Groups must be formed by no later than Week 7, and registered on Canvas. If there are issues with group work, you must inform the course coordinator by no later than 5.00pm Friday of Week 09, otherwise all group members will receive the same grade, with no exceptions.
Marks	40%

1 Overview

In this assignment you will design and create an end-to-end machine learning system for a real-world problem. This assignment is designed for you to apply and practice skills of critical analysis and evaluation to circumstances similar to those found in real-world problems. This is a group project (groups of 04 or 05). In this assignment you will:

- Analyze model requirements and constraints for the purpose of designing and implementing solutions to a learning challenge
- Selecting the appropriate ML techniques and applying them to solve a real-world ML problem.
- Analyzing the output of the above algorithm(s).
- Evaluate and compare approaches and algorithms on the basis of the nature of the problem/task being addressed
- Interpret abstract theoretical propositions, choose methodologies, justify conclusions and defend professional decisions to both IT and non-IT personnel via technical reports of a

professional standard and technical presentations

- Research how to extend the modelling techniques that are taught in class.
- Providing an ultimate judgment of the final trained model that you would use in a real-world setting.

To complete this assignment, you will require skills and knowledge of Machine Learning. However, you will be able to commence work on some sections. Thus, do the work you can initially, and continue to build in new features as you learn the relevant skills. *A machine learning model cannot be developed within a day or two. Therefore, start early.*

This assignment has three deliverables:

1. A report (of no more than 05 pages, plus up to 02 pages for appendices) critically analysing your approach and ultimate judgement.
2. Two (or three) final models for predicting, and the scripts to run them.
3. Your Python scripts or Jupyter notebooks and software used to build your learning system and produce the models and results with instructions on how to run them.

More detail is provided in Section.3, Assignment detail, below.

2 Learning Outcomes

This assessment relates to the following course learning outcomes (CLOs):

- **CLO 1:** Understand the fundamental concepts and algorithms of machine learning and applications.
- **CLO 2:** Understand a range of machine learning methods and the kinds of problem to which they are suited
- **CLO 3:** Set up a machine learning configuration, including processing data and performing feature engineering, for a range of applications.
- **CLO 4:** Apply machine learning software and tool-kits for diverse applications.

3 Task

Using machine learning in real-world settings involves a more than just running a data set through a particular algorithm. In this assignment, you will design, analyse and evaluate a complete machine learning system.

*The key aspect of this assignment is the **design, analysis, and evaluation** of your methodology, investigation, and results. This assignment focuses on both the accuracy of your model, and your understanding of your approach and model.*

For this assignment you have a **choice of your project**. You may **select the project listed in Section 4**, or you may **negotiate a project** with the **course coordinator** (providing that you are working in a group). Regardless of the problem you choose, you must conduct the following tasks:

1. You need to come up with an approach, where each element of the system is justified using data analysis, performance analysis and/or knowledge from relevant literature
2. Investigate various Machine Learning solutions to the problem
3. Make an ultimate judgement
4. Evaluate your ultimate judgement against independent testing data
- 5. Saving two/three models and scripts for running these two/three models.**
6. Produce a report of your design, investigation, evaluation and findings

3.1. Investigation

Your investigation will require you to design, use, analyse and evaluate an end-to-end machine learning system. You should consider a variety of techniques that have been discussed in class, and techniques you have researched. Your end-to-end system may consist of elements such as:

- A well justified evaluation framework
- Pre-processing the data set to make it suitable for providing to various machine learning algorithms
- Carefully selected and justified baseline model(s)
- Hyper-parameter setting and tuning to refine the model
- Analysing model and outputs and interpreting the trained models

*The details in this spec are the **minimum requirements**. A thorough investigation must consider more than the minimum to receive high grades.*

3.2. Ultimate Judgement

You must make an **ultimate judgement** of the “best” model that you would use and recommend for your particular project. It is up to you to determine the criteria by which you evaluate your model and determine what is means to be “the best model”.

For higher grades you must use techniques that go **beyond simple performance metric analysis** when making the **ultimate judgement**.

3.3. Independent Evaluation of your Ultimate Judgement

You may conduct an independent evaluation of your ultimate judgement. This can be conducted where possible by:

- Using data collected completely outside of the scope of your original training and evaluation
- Comparing your performance to other works in literature that use similar data

3.4. Approach, Critical Analysis and Report

You must compile a report analysing the approach you have taken in your investigation. Your report:

- **Must be no longer than 05 pages of text**
- **May contain an additional 02 pages for appendices**

- Use a single-column layout with no less than size 11pt font
- The appendices may only contain citations, figures, diagrams, or data tables that provide evidence to support the statements in your report, not the judgment.
- Include the name(s) and student id's of the student(s) who wrote the report.

Any over length content, or content outside of these requirements will not be marked. For example, if you report is too long, ONLY the first 3 pages of text will be read and marked.

In this report you should analyse elements such as:

- **Preprocessing data, including extra-collecting activities if necessary.** Note that if you do extra-collection or preprocessing step, make sure your new **dataset must be available (git, drive, ...) for evaluator to access.**
- Machine learning algorithms that you considered
- **Why** you selected these approaches?
- **Evaluations of the performance** of trained model(s)
- Your **ultimate judgement** with supporting analysis and evidence

This will allow us to understand your rationale. We encourage you to explore this problem and not just focus on maximising a single performance metric. By the end of your report, we should be convinced by your ultimate judgement and that you have considered all reasonable aspects in investigating your chosen problem.

The key aspect of this assignment isn't your code or model, but the thought process behind your work.

Remember that good analysis provides factual statements, evidence and justifications for conclusions that you draw. A statements such as:

"I did <xyz> because I felt that it was good"

is not analysis. This is an unjustified opinion. Instead, you should aim for statements such as:

"I did <xyz> because it is more efficient. It is more efficient because . . . "

4 Projects

4.1 Classify Images of Furniture

Assume you are a team of machine learning engineers working for an ecommerce furniture shop, where users can browse and navigate interior furniture items. You are required to build a Furniture Recommender that allows users who have recently moved to explore furniture on your ecommerce system at ease. Your systems should have a functionality to help users navigate to the category of the furniture item that users want to buy. In most of the current online shops, users should type the name of the items and browse from the list of the results. However, to enhance the quality of the searching results, our system provides an image based searching function, where the users can upload the images of the furniture item that they are looking for. The system will accomplish an image search and return the list of similar-styled furniture in favor from our dataset.

In the Furniture dataset, there are 06 categories: **beds - 6578 images; chairs - 22053 images; dressers - 7871 images; lamps - 32402 images; sofas - 4080 images; tables - 17100 images**, with total of **90084** images. For every category, **there are 17 interior styles**:

(a) Asian; (b) Beach; (c) Contemp; (d) Craftsman; (e) Eclectic; (f) Farmhouse; (g) Industrial
(h) Media; (i) Midcentury; (j) Modern; (k) Rustic; (l) Scandinavian; (m) Southwestern
(n) Traditional; (o) Transitional; (p) Tropical and (q) Victorian

You have three tasks in this project:

Task 1: *Classify images according to furniture category (beds; chairs; dressers; lamps; sofas; tables)*

Task 2: *Recommend 10 furniture items in our dataset **which is similar** to the input furniture item image from users. You are required to define **a metric of “similarity”** between two furniture items.*

Task 3: *(only for those aim HD) The extension of the model in Task 2, the recommended furniture items **must be in the same interior styles** with the style of the input images. In order to fulfill this task, you are required to build a model to recognize **the style of a furniture item**.*

Dataset: available on Canvas

Requirements

- You are required to do the pre-processing step on the Furniture dataset, including **extra-collection** if necessary.
- You must investigate at least one machine learning algorithms for each of the two tasks. That is, you must build at least **one model capable of predicting the type of furniture item images**, and at least **one model capable of showing 10 similar images**.
- **For every task, you must submit one different model.**
- You are not required to use separate type(s) of machine learning algorithms, however, a thorough investigation should consider different types of algorithms.
- You are required to fully train your own algorithms. You may not use pre-trained systems which are trained on other datasets (not given to you as part of this assignment).
- For higher grades (HD/DI) you must explore how the current status of the data will affect to the result of the models, how we can improve the models, and implement your suggestion to improve the models.
- Your final report must conduct an analysis and comparison between different model results, not only just one model.

Independent Evaluation

- **Your independent evaluation is** to research other works that have the same goals. Then you must compare and contrast your results to those other works.
- Using data collected **completely outside of the scope of your original training** and evaluation

4.2 Negotiated Project

To propose a project, please complete the project proposal template below and send an email

to the course coordinator (template is available on Canvas).

You may propose and negotiate a project and machine learning problem to investigate, with the course coordinator. This project must meet a number of constraints:

- The project must be of a suitable complexity and challenge that is similar to the suggested projects. As part of the negotiation, the scope and deliverables of the project will be set.
- If the project is using an existing data set, the problem should be phrased in a manner that can be solved by multiple machine learning methods, of which at least two methods will be investigated.
- If the project requires a data set to be generated, devised, or collected, this collection should require sufficient effort. In these cases, especially for reinforcement learning tasks, only one machine learning method may need to be investigated.
- The proposed project must be independent of previously or concurrently assessed work.

You may not conduct a project if you have already been assessed on the work, or are concurrently being assessed on the work.

In general, negotiations will take place via email, during tutorial/consultation hours, or by appointment.

All negotiated projects must be finalised by no later than 5pm Friday 03 May (Week 9). This is the absolute deadline.

If you wish to conduct a negotiated project, begin the negotiation process early. A negotiated project may be denied before the deadline if there is insufficient time for the negotiation process.

5 Submission

Only one member of your group should submit the necessary files. Please do not spread your submission across both members of your group. You will need to submit the following deliverables:

1. **The PDF/doc version of report.** Please name the report by following this convention:

COSC2753_A2_<Group number>_<studentID1_studentID2_...>.

2. Your **models & code** (Jupyter notebooks or Python scripts) used to perform your analysis. Should be a ZIP file containing all the support files. will be used for plagiarism checking
 - Putting two/three models and your code (including the scripts for running your models for predicting) in one folder.
 - Zip your folder before submitting it. Name the zip file by following the same convention above
 - Please make sure that the notebook must be readily opened and executed on any standard machine/setting. The best practice is to include a README file to instruct the user on how to set up the environment to run your file (e.g., put data files in which folder, install which libraries/packages etc.)

- Please note that your code will be checked for plagiarism by our specialised software and Turnitin too.
- Please submit your models and code in the separate Canvas page (link here)

3. Please name your report and source code zip file by following this convention:

COSC2753_A2_<studentID1_studentID2_...>

If your submission does not follow the name convention, the mark deduction will be applied.

The submission portal on canvas consists of ***two sub-pages***.

- First page for PDF/doc of your report submission – ***only PDF/doc file***
- The second page for models & code submission. Should be a ZIP file containing source code, models and all the support files. We strongly recommend you to attach a README file with instructions on how to run your application. Make sure that *your assignment can run only with the code included in your zip file!*

After the due date, you will have 5 days to submit your assignment as a late submission. Late submissions will incur a penalty of 20% per day. After these five days, Canvas will be closed and you will lose ALL the assignment marks.

Assessment declaration:

When you submit work electronically, you agree to the assessment declaration <https://www.rmit.edu.au/students/student-essentials/assessment-and-exams/assessment/assessment-declaration>

6 Teams

Group of 04 or 05 students

7 Academic integrity and plagiarism (standard warning)

Academic integrity is about honest presentation of your academic work. It means acknowledging the work of others while developing your own insights, knowledge and ideas. You should take extreme care that you have:

- Acknowledged words, data, diagrams, models, frameworks and/or ideas of others you have quoted (i.e. directly copied), summarized, paraphrased, discussed or mentioned in your assessment through the appropriate referencing methods
- Provided a reference list of the publication details so your reader can locate the source if necessary. This includes material taken from Internet sites. If you do not acknowledge the sources of your material, you may be accused of plagiarism because you have passed off the work and ideas of another person without appropriate referencing, as if they were your own.

RMIT University treats plagiarism as a very serious offence constituting misconduct. Plagiarism covers a variety of inappropriate behaviors, including:

- Failure to properly document a source
- Copyright material from the internet or databases
- Collusion between students

For further information on our policies and procedures, please refer to the following: <https://www.rmit.edu.au/students/student-essentials/rights-and-responsibilities/academic-integrity>.

8 Marking guidelines

A detailed rubric is attached on canvas.