FIT5212 - Assignment 2

Marks	Worth 50 marks, and 25% of all marks for the unit
Due date	Due Week 12 – Lecture Date at 23:55 pm
Extension	An extension could be granted under some circumstances. A special consideration application form must be submitted. Please refer to the university webpage for special consideration.
Lateness	For all assessment items handed in after the official due date, and without an agreed extension, a 10% penalty applies to the student's mark for each day after the due date (including weekends) for up to 7 days. Assessment items handed in after 7 days without special consideration will not be considered.
Authorship	This is an individual assessment. All work must be your own. All submissions will be placed through Turnitin. This makes plagiarism remarkably easy to identify for us.
Submission	Submission is 5 files: ● one CSV file (for part 1), ● one CSV file (for part 2), ● one PDF discussion report (including discussion for Parts 1&2), ● one Jupyter notebook (for part 1), ● one PDF generated directly from the Jupyter notebook of Part 1 without cell output for the whole assessment. The files must be submitted via Moodle. All files will go through Turnitin for plagiarism detection.
Programming Language	Python in Jupyter Notebook

Task 1: Recommender System Challenge (70% Marks)

Description:

You are required to complete an **FIT5212-specific challenge in Kaggle**

https://www.kaggle.com/t/3b8c276ec6224453a9d7f32392b99913

This dataset is collected from an online e-commerce platform. People interact with the system by posting reviews and ratings for different items. You are required to predict item ratings for user/item pairs in the test set.

Data:

The user-item interaction data is the main data for this challenge. This data is further split into training and test sets.

- train.csv. The training dataset contains a set of user_item ratings between users and items. The users explicitly rated the items that they interacted with between 1 to 5.
- test.csv. Each user is provided with a list of 100 items in the test dataset, for each user, you will need to predict the ratings for all the items in their list.

Please train your recommender systems and generate the outputs for the test data.

Optional Data:

In addition to the interaction data (training data), you are also provided with the item's metadata. In this assessment, each item represents a book and some information about the book is provided in the books_metadata.csv file. The usage of this file is optional but if used correctly, it can improve the performance of your model. More information can be found here https://www.kaggle.com/t/3b8c276ec6224453a9d7f32392b99913

Requirements:

1. Participate in the challenge and make your submission. The maximum submission in Kaggle is 10 submissions per day.
2. This is an individual assignment. You have to finish it on your own.
3. In addition to the challenge, you have to finish a report on this challenge and submit it to Moodle.

Submission:

To Kaggle

- Kaggle submission, you need to submit your predictions on the test dataset on Kaggle.

To Moodle:

1. A csv file, "**studentID.csv**". Please replace studentID with your own student ID. The content should be the same as the file you have submitted to Kaggle. This file should be submitted in Moodle. We will double-check the files you have submitted to Kaggle

and Moodle. If the two files are not the same (i.e., the file submitted to Moodle cannot get the same score in Kaggle), your result is invalid, and you will fail the assignment.

2. A jupyter notebook, "**code_studentID.ipynb**". This notebook contains the code for Task 1 and Task 2. The notebook should be self-contained. If a third-party package is used, this package should be a well-known package and easy to install (e.g., install within a single command). This notebook should include both codes and outputs so that we can read and mark them.
3. A pdf file, "**code_stduentID.pdf**". This pdf is generated by cleaning all the output in the Jupyter Notebook and exporting it as a pdf file. This pdf will be passed in Turnitin for plagiarism check.
4. A pdf report, "**report_stduentID.pdf**". This pdf contains a more detailed analysis of the work. This file should show how you finished the task. Ideally, you should show what sort of algorithms you have considered, what kind of information you have used, and the reason for your choice of the corresponding algorithm to achieve the results you submitted to Kaggle. **Comparison for different algorithms should be included in this pdf report.** And detailed analysis of the results is encouraged. If you have used other algorithms/packages which are not covered in this unit, you should give a brief introduction to that algorithm/package. We expect the length of this task in the report to be between **8 to 10 pages** excluding references. This pdf will be passed in Turnitin for plagiarism check.

Marking:

- The Kaggle leaderboard only shows your scores on 50% of the test data. Your final score will be marked based on your CSV file submitted to Moodle for the whole test dataset.
- The methodology and report are set to 30% marks (out of 70%), and the prediction score accounts for 40% marks (out of 70%). So please prepare a good report and clearly describe your method to achieve the marks.

Task 2: ChatGPT (30% Marks)

Create an account to use ChatGPT's browser interface (https://chat.openai.com/chat). You are given two broad domains. In this part of the assignment you need to come up with 5 reasoning questions per each domain (could be: short answer/essay, reading comprehension, or multiple choice questions) for which ChatGPT **fails**. So you will have 10 questions in total (2 domains x 5 questions). Domains and examples:

- Domain 1: Business Sector (Choose any business sector you are familiar with. Examples: Legal, Health & Medicine, Finance & Insurance. If you are unsure whether something fits as a business sector, ask the CE.)
    - *Sebastian, a 17-year-old university student, purchased a pair of running shoes last year. He has since discovered that both shoes are faulty. The vendor told him that he cannot return the shoes. Sebastian wishes to have this dispute resolved and wants a final and binding decision that he is entitled to a refund. His friend suggests that he take the matter to Consumer Affairs Victoria (CAV). Advise him about whether CAV is the most appropriate body to help resolve this dispute.*
    - An infant develops jaundice 6 hours after birth. Which one of the following is the most likely diagnosis? A. Haemolytic disease of the newborn. B. Umbilical

sepsis. C. Physiological jaundice. D. Atresia of the bile ducts. E. Neonatal hepatitis.

- Domain 2: General (Commonsense, Logical, Temporal)
    - *There is an apple inside a blue box. There is also a red box inside the blue box. The red box has a lid. How can I get the apple?*
    - *Why did my grandmother not invite me to her birthday party when she was 10?*

The questions should NOT come from the internet or published articles, but if you were inspired by a resource cite (Use Chicago Citation Style: https://guides.lib.monash.edu/citing-referencing) it along each question in the report. Think outside the box, be creative, and do not just mimic what other people have done. For each question you will get 3% (1% for the question, 0.5% for the creativity in designing the question, and 1.5% for the analysis of patterns you observed, what strategies you employed, why do you think the model failed), adding to a total of 30% marks (out of 30%).

IMPORTANT NOTES:

For interacting with ChatGPT and reporting the question and answers in your report, you should be following this instructions:

- When interacting with ChatGPT: Each question should be typed into a New Chat (i.e., do NOT type more than 1 question into a single chat session). You can create a New Chat session by clicking on ``New Chat'' located in the top-left corner of the environment. If you do not know why this is necessary, ask the CE.
- In your report, each question and the given answer by ChatGPT should be provided as a screenshot from the ChatGPT environment (showing both the question and the answer). Otherwise you will not be getting any marks for the question.
- For **each single question**, you need to report the exact date and time (e.g., Question was asked on Saturday, 23 April 2023, 13:26) in your report. Otherwise you will not be getting any marks for the question. ChatGPT is constantly improving, so it is important to document your findings with an exact time stamp.

Recommended (Optional) Readings:

To better be prepared for the analysis part of the failed cases, you may want to read the following works and cite them or any other works that support your speculation for the ChatGPT's behaviour:

- Benchmarks for Automated Commonsense Reasoning: A survey (https://arxiv.org/pdf/2302.04752.pdf)
- LogiQA: A Challenge Dataset for Machine Reading Comprehension with Logical Reasoning (https://arxiv.org/abs/2007.08124)
- HellaSwag: Can a Machine Really Finish Your Sentence? (https://arxiv.org/pdf/1905.07830.pdf)
- PIQA: Reasoning about Physical Commonsense in Natural Language (https://arxiv.org/abs/1911.11641)
- TruthfulQA: Measuring How Models Mimic Human Falsehoods (https://arxiv.org/abs/2109.07958)
- :WINOGRANDE: An Adversarial Winograd Schema Challenge at Scale (https://arxiv.org/pdf/1907.10641.pdf)

- Online Resource: Gary Markus and Ernest Davis Experiments testing GPT-3's ability at commonsense reasoning: results. (https://cs.nyu.edu/~davise/papers/GPT3CompleteTests.html)
- Training Verifiers to Solve Math Word Problems (https://arxiv.org/pdf/2110.14168v2.pdf)
- Online Resource: Google Big Bench (https://github.com/google/BIG-bench/blob/main/bigbench/benchmark_tasks/README.md)
- Language Models are Few-Shot Learners, sec 3.5 (https://arxiv.org/pdf/2005.14165.pdf)
- PaLM: Scaling Language Modeling with Pathways, section 6 (https://arxiv.org/pdf/2204.02311.pdf)
- Sparks of Artificial General Intelligence: Early experiments with GPT-4 (https://arxiv.org/abs/2303.12712)
- Evaluating the Logical Reasoning Ability of ChatGPT and GPT-4 (https://arxiv.org/pdf/2304.03439.pdf)

IMPORTANT SUBMISSION NOTE FOR PART 2:

For part 2, in addition to the corresponding section in the report (in "**report_stduentID.pdf**"), you also need to submit a CSV file ("**studentID_chatgpt.csv**") containing 10 rows, with the following 5 column headers: Domain, Question, Correct Answer, ChatGPT Answer, Time-Date.