

# A Hybrid Multi-Modal Recommender System using Neural Collaborative Filtering and Content Based Filtering - Student Outline DMP

---

## 1. General guidelines

**PURPOSE OF THIS TEMPLATE** - The purpose of the Outline DMP is to indicate your initial plans for how your data will be collected, shared and stored, and to give you a chance to think about these data-focused aspects of the research process. As you begin doing your research, your data process may change, and it is perfectly acceptable to change your data management plan to accommodate the changes in your research process. Indicate below that you understand the purpose of completing this Outline DMP template.

- I understand the Outline DMP template is a projection of my anticipated data management planning requirements and should be updated as my project develops.

## 2. Authors and supervisors

**PROJECT NAME** - Replicate the title of your project, dissertation or thesis exactly as it appears in your proposal document.

A Hybrid Multi-Modal Recommender System using Neural Collaborative Filtering and Content Based Filtering

**PERSONAL DETAILS** - Indicate the name(s) and student number(s) of the student(s) who will be involved in this project, dissertation or thesis.

Pavan Singh (SNGPAV003)

**SUPERVISOR(S) DETAILS** - Indicate who will supervise this project, dissertation or thesis. If you do not yet have a supervisor, leave this section blank.

Ian Durbach (Primary Supervisor) Allan Clark (Co-Supervisor)

## 3. Data Collection/Generation

**COLLECTION OF ORIGINAL DATA** - Indicate whether or not you intend to gather/produce original data for your study, and provide a brief description of the kind of data you think you will collect. If you are unsure at this time, indicate what you think you are most likely to collect. If you are not intending to gather or collect your own data, declare that here.

- I do not intend to collect original data.

The dataset is a publicly available dataset.

**USE OF EXISTING DATA** - Indicate if you intend to re-use existing data, either from online searches or from datasets provided by your supervisor, lab, or funder. If you are not intending to re-use existing data, declare that here.

- I intend to reuse existing data in my study (described below).

The dataset is the Amazon Product Reviews Dataset. It is publicly available and widely renowned for providing a large amount of metadata and ratings for a variety of product categories. The data is held at: <https://s3.amazonaws.com/amazon-reviews-pds/readme.html>

**DATA SHARING** - Indicate whether or not you are intending to publish your research data. If you are, indicate where you are intending to publish your data and under what licensing conditions, such as Creative Commons. If you are not intending to

**publish your data, provide reasons and reference the appropriate ethical considerations, commercial applications/patenting ambition, or data re-use agreements that prevent you from publishing your data.**

- I intend to share my data (details below).

The model and all parameters will be made available to public. The data, models and scripts will be made available on GitHub.

## 4. Data Storage

**ANTICIPATED DATASET SIZE** - Indicate the estimated size of your completed dataset, and indicate whether or not you will need to access additional data storage facilities. If such storage is not provided by your unit or department, you may need to factor in the cost of purchasing additional storage space.

- 20GB or less

The implementation of the project is anticipated to be less than 20GB. No access to special storage facilities are required.

**DATA BACKUPS** - Indicate how you plan to ensure your data is secure and retrievable in case of errors or hardware failure. Describe what procedures you will put in place to back-up copies of your data and where they will be stored.

- I intend to backup my data using a commercial service provider.

All code produced in this project (the framework, model and decoder files) will be stored on Github following git best practices.

## 5. Data Centre(s)/Repositories

**DATA CENTRES/REPOSITORIES** - Once your project, dissertation or thesis is complete, it is advisable to curate and archive your completed dataset with an established data centre or repository. Note that you should archive your data even if you are not intending to publish it. Check with your supervisor or funder if you are required to deposit your data in a specific repository, or declare that you will deposit the data in ZivaHub (see the Guidance section).

- At the end of my study, I will deposit my data on a subject or disciplinary repository (details below).

All code produced in this project (the framework, model and data files) will be made available on Github.

**METADATA** - Metadata is descriptive information that others will need to make sense of your dataset. Metadata includes things like study descriptions or abstracts, study instruments (sample collection schedules, codebooks for variables, survey instruments, etc.), subject codes, and keywords. Indicate what metadata will accompany your curated dataset.

The entire project will be accompanied by a README markdown file that describes the entire project. This includes installation, execution, data-collection as well as data processing. A separate markdown file called DATA\_DESCRIPTION will describe each of the input and output variables of the project and will be linked directly from the README file.

## 6. Budget

**BUDGET** - Indicate any costs specifically relating to the management and curation of your data, such as purchasing additional storage space, digitisation of physical media, data storage or curation charges, and data audits. Most student research will be able to make use of free options provided by UCT and will not have to budget for data costs.

- I do not anticipate any data costs as my data is less than 10GB, and I will be using a storage system provided by UCT (UCT GoogleDrive, UCT OneDrive, Netstorage, ZivaHub, etc.) to curate my data.

Github does have any fees for a project of this size.

