**Chengdu University of Technology Oxford Brookes College**

**Project Module (CHC 6096)**

**Weekly Report Sheet - 2023/2024 Academic Year**

|  |  |
| --- | --- |
| STUDENT NAME: | Wang Yutong (Ryan) |
| STUDENT NUMBER: | 202018010315 |
| SUPERVISOR NAME: | Joojo Walker |
| WEEK NUMBER | Week 7 |
| DATE: | 2023/11/17 |
| **Action plan for the current week:**   1. Download three datasets: Amazon Books, Good Reads, and Book-Crossing 2. Preprocessing of the Good Reads dataset. <https://github.com/FOMOKN/Project.git> 3. Study papers about RecVAE models. | |
| **Challenges and issues encountered in the week:**  Confused about feature engineering of datasets:   1. For columns with large outliers, such as ‘reviews’ and 'total ratings’, normalization was applied. However, such great values are extremely rare. Is it possible to remove them? Then use normalization 2. Checked that in the 'pages' column. Maximum value is 70,000 and Minimum value is 0. For books, they are unrealistic page numbers. Can they be deleted? 3. Need to find a suitable way to handle text capsules. For instance, a book may have multiple authors belonging to multiple categories. Temporary use of MultiLabelBinarizer. | |
| **Action plan for the next week:**   1. Complete preprocessing of all data sets 2. Training a model with a training set | |
| **Supervisor Feedback:** | |