

TABLE OF CONTENTS

TABLE OF CONTENTS.....	1
TABLE OF FIGURES.....	3
TABLE OF TABLES.....	5
ABSTRACT.....	6
CHAPTER I INTRODUCTION.....	7
1.1 Research Background.....	7
1.2 Research Problem.....	8
1.3 Research Scope.....	9
1.4 Research Objective.....	9
1.5 Research Advantage.....	10
CHAPTER II LITERATURE REVIEW.....	11
2.1 Recommendation System.....	11
2.2 Collaborative Filtering.....	12
2.3 Content-based Filtering.....	13
2.4 Hybrid Filtering.....	14
CHAPTER III THEORITICAL BASIS.....	18
3.1 Recommendation System.....	18
3.1.1 Collaborative Filtering (CF).....	19
3.1.2 Content-Based Filtering (CBF).....	24
3.1.3 Cascade Hybrid Filtering.....	37
3.2 Data Preprocessing.....	38
3.3 Evaluation.....	40
3.3.1 Mean Absolute Error (MAE).....	41
3.3.2 Root Mean Squared Error (RMSE).....	41
3.3.3 Topic Coherence.....	41
3.3.4 Confusion Matrix.....	42
3.3.5 User Study.....	43
3.3.6 Rest API Testing.....	44
3.3.7 Usability Testing.....	44
CHAPTER IV RESEARCH METHODOLOGY.....	45
4.1 Research Description.....	45
4.1.1 Literature Study.....	46
4.2 Tools and Material.....	46
4.2.1 Tools.....	47
4.2.2 Materials.....	48
4.3 Dataset and Database.....	48
4.3.1 Initial Data Acquisition and Pre-Processing.....	48
4.3.2 Data Pre-Processing for Neural Collaborative Filtering (NCF).....	49
4.3.3 Data Pre-processing in BERTopic.....	49
4.4 Database.....	50
CHAPTER V IMPLEMENTATION.....	53
5.1 Hybrid Recommendation Model.....	53
5.1.1 NCF Model.....	53



5.1.2	BERTopic Model.....	60
5.1.3	Cascade Hybrid Filtering Application.....	64
5.2	Software Design and Implementation.....	65
5.2.1	Requirements Specifications.....	66
5.2.2	Developing the Backend.....	69
5.2.3	Developing the Frontend.....	70
5.2.4	Initial Design Research.....	71
5.2.5	Mock-Up.....	73
5.2.6	Home Page.....	75
5.2.7	Search Page.....	77
5.2.8	Register.....	77
5.2.9	Log in page.....	78
5.2.10	Recommendation Page.....	78
5.2.11	Movie Information.....	79
5.2.12	Watchlist Page.....	81
5.2.13	Additional Features.....	82
CHAPTER VI RESULT AND DISCUSSION.....		84
6.1	NCF Result.....	84
6.2	BERTopic Result.....	84
6.3	Hybrid Recommendation Result.....	87
6.4	System Evaluation.....	89
6.4.1	NCF Model Evaluation.....	89
6.4.2	BERTopic Evaluation.....	90
6.4.3	Hybrid Model Evaluation.....	92
6.4.4	Software Evaluation.....	95
6.4.5	Rest API Testing.....	95
6.4.6	Usability Testing.....	95
CHAPTER VII CONCLUSION AND FUTURE WORKS.....		98
7.1	Future Works.....	98
REFERENCES.....		99
APPENDIXES.....		105

TABLE OF FIGURES

Figure 3.1 Concepts of user-based filtering and item-based filtering (Chen, Hui and Thaipsisutikul, 2020).....	19
Figure 3.2 Matrix Factorization (Google for Developers, 2022)	21
Figure 3.3 Neural Collaborative Filtering Framework (He et al., 2017)	22
Figure 3.4 Neural matrix factorization model (NeuMF), (He et al., 2017).....	23
Figure 3.5 The sequence of steps to create topic representations in BERTopic	25
Figure 3.6 BERT architecture.....	26
Figure 3.7 Illustration of MLM and NSP that is applied on sentences.	27
Figure 3.8 Pre-training and Fine-tuned BERT model.....	27
Figure 3.9 Simplexes (Readthedocs.io, 2018).....	29
Figure 3.10 HDBSCAN Visualization I.....	32
Figure 3.11 HDBSCAN Visualization II.....	32
Figure 3.12 HDBSCAN Visualization III	33
Figure 3.13 HDBSCAN Visualization I.....	34
Figure 3.14 CountVectorizer Visualisation.....	36
Figure 3.15 Cascade Hybrid Filtering (Chiang, 2021).....	38
Figure 3.16 Named Entity Recognition Visualization (Ekaterina Kochmar, 2020).....	39
Figure 3.17 Lemmatization Visualization (Turing.com, 2022)	40
Figure 4.1 Project Flow	46
Figure 4.2 ER Diagram.....	51
Figure 5.1 MF Model Architecture.....	54
Figure 5.2 Neural Network with MF model architecture	56
Figure 5.3 NCF Model Architecture	59
Figure 5.4 Cascade Hybrid Filtering Application	65
Figure 5.5 UK video streaming usage 2021 (Business of Apps, 2023)	71
Figure 5.6 Login Page Mock-Up.....	73
Figure 5.7 Register Page Mock-Up	73
Figure 5.8 Home Page Mock-up	74
Figure 5.9 Pop-up Mock-up.....	74
Figure 5.10 Watchlist Page	75



Figure 5.11 Recommendations Page.....	75
Figure 5.12 Carousel Home Page	76
Figure 5.13 Unauthenticated Users Home Page	76
Figure 5.14 Authenticated Users Home Page	76
Figure 5.15 Search Page	77
Figure 5.16 Register Page.....	78
Figure 5.17 Login Page	78
Figure 5.18 Recommendation Page.....	79
Figure 5.19 Poster Cards.....	80
Figure 5.20 In-depth Movie Information	81
Figure 5.21 Rating Visualisation	81
Figure 5.22 Watchlist Page	82
Figure 5.23 Genre-based Filtering.....	82
Figure 5.24 Load More Button.....	83
Figure 5.25 That's it Button.....	83
Figure 6.1 Highest Rated Movies of a User	84
Figure 6.2 10 Recommended Movies.....	84
Figure 6.3 The Topic Model Information.....	85
Figure 6.4 Document Probability Information	85
Figure 6.5 The final data frame result.....	86
Figure 6.6 Intertopic Distance Map.....	86
Figure 6.7 Topic Word Scores	87
Figure 6.8 Term Score Decline per Topic.....	87
Figure 6.9 Hybrid Result in Google Colab.....	88
Figure 6.10 Recommendation Page.....	88
Figure 6.11 Usability Testing Result.....	96

TABLE OF TABLES

Table 2.1 Previous Research to the Current Research.....	16
Table 3.1 Confusion Matrix	43
Table 4.1 List of Libraries.....	47
Table 5.1 Hyperparameter of MF	54
Table 5.2 Epoch result in MF model.....	55
Table 5.3 Hyperparameter of Neural Network with MF model	57
Table 5.4 Epoch result in the Neural Network with MF model.....	57
Table 5.5 Hyperparameter of the NCF model.....	59
Table 5.6 Epoch result in the NCF model	60
Table 5.7 BERTopic applied sub-models	61
Table 5.8 Sentence Transformer comparison (Sbert.net, 2022).....	62
Table 5.9 Parameter Settings and Reasons for UMAP	63
Table 5.10 Parameter Settings and Reasons for HDBSCAN	63
Table 5.11 Streaming Service Dark Mode Research	72
Table 6.1 MAE comparison for MF, Neural Network, and NCF	89
Table 6.2 Topic Coherence Evaluation with Different Fine-tuning Models.....	91
Table 6.3 Confusion Matrix Table Result.....	91
Table 6.4 NCF Recommendation System Evaluation	92
Table 6.5 BERTopic Recommendation System Evaluation	93
Table 6.6 Hybrid Recommendation System Evaluation	94
Table 6.7 Things to fix.....	97