

TABLE OF CONTENTS

TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
ABSTRACT.....	x
INTISARI.....	xi
CHAPTER I INTRODUCTION.....	1
1.1 The Background Problem	1
1.2 Problem Statement	4
1.3 The Limitations	4
1.4 Research Objective.....	4
1.5 The Benefits	5
CHAPTER II LITERATURE REVIEW.....	6
2.1 Related Works	6
2.2 Chosen Models.....	8
CHAPTER III THEORETICAL FOUNDATION.....	13
3.1 Aspect-Based Sentiment Analysis (ABSA).....	13
3.1.1 Topic Modelling.....	13
3.1.2 Sentiment Classification	14
3.2 Data Preprocessing.....	14
3.3 Bidirectional Encoder Representations from Transformers (BERT).....	15
3.3.1 BERTopic.....	16
3.3.1.1 Embedding with SentenceBERT (SBERT):	17
3.3.1.2 Dimensionality Reduction with UMAP:.....	18
3.3.1.3 Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN):.....	21
3.3.1.4 Class-based TF-IDF	22
3.4 Support Vector Machine (SVM).....	23
3.5 Logistic Regression (LR)	24
3.6 Evaluation Method	25
3.6.1 Topic Coherence	25
3.6.2 Performance Metrics for Classification	27
3.7 Cross-Validation	28
CHAPTER IV RESEARCH METHOD	30
4.1 Research Description	30
4.2 Dataset.....	32
4.3 Data Pre-processing	34
4.4 The Models	36
4.5 Performance Evaluation	38
CHAPTER V IMPLEMENTATION.....	39
5.1 Initial Setup & Dataset	39
5.2 Dataset Preprocessing & Analysis	40
5.3 Training the model	42
5.4 Aspect Classification.....	46
5.5 The Evaluation Method.....	47
CHAPTER VI RESULTS AND DISCUSSION.....	51

6.1	Multi-Aspect Dataset Case.....	51
6.1.1	Aspect Extraction and General Sentiment Classification Performance for Multi-Aspect.....	54
6.1.2	Sentiment Classification of Individual Topic Representation Sets for Multi-Aspect.....	56
6.1.3	Comparison Between the SVM and LR Models for Multi-Aspect.....	59
6.2	Single-Aspect Dataset Case	59
6.2.1	Aspect Extraction and General Sentiment Classification Performance for Single-Aspect	62
6.2.2	Sentiment Classification of Individual Topic Models for Single-Aspect	64
6.2.3	Comparison Between the SVM and LR Models for Single-Aspect	67
6.3	Comparison Between Cases	67
6.4	Comparison with Other Models.....	68
CHAPTER VII CONCLUSION AND SUGGESTIONS		69
7.1	Conclusion	69
7.2	Suggestions	70
REFERENCES.....		71

LIST OF FIGURES

Figure 1.1 Example Steam reviews distribution on Overwatch 2 (<i>Overwatch® 2 on Steam</i>).....	2
Figure 3.1 Preprocessing process for SVM Sentiment analysis (Satrya <i>et al.</i> , 2022)	15
Figure 3.2 BERT input embeddings by (Devlin <i>et al.</i> , 2019).....	16
Figure 3.3 Diagram of the base BERTopic process	16
Figure 3.4 A Structural model representation of SBERT (Reimers and Gurevych, 2019)	17
Figure 3.5 A visualisation of the effects of the parameters of min_dist and n set for the UMAP(McInnes <i>et al.</i> , 2020).....	21
Figure 3.6 Example of Topic Representation Words (Anoop and Asharaf, 2018).....	23
Figure 3.7 Illustration of SVM hyperplanes adapted from Zhang (2012)	23
Figure 3.8 Illustration of the logistic regression sigmoid function by (Jurafsky and Martin, 2025)	25
Figure 3.9 Topic coherence model overview by (Röder <i>et al.</i> , 2015)	25
Figure 3.10 Example of a C_V Coherence Score of BERTopic Model	27
Figure 3.11 Confusion Matrix.....	27
Figure 3.12 Confusion Matrix SVC from Putra and Dijaya (2023).....	28
Figure 3.13 10-fold Cross-validation example by Berrar (2019).....	29
Figure 4.1 General Overview Flow Diagram of the research	30
Figure 4.2 The Dataset overview	32
Figure 4.3 Preprocessing Process Diagram.....	35
Figure 4.4 The architecture of the model.....	37
Figure 5.1 Initial Setup collection.....	39
Figure 5.2 Case Folding and Data Cleaning	40
Figure 5.3 Tokenization Code.....	40
Figure 5.4 Stop word Removal	40
Figure 5.5 Lemmatisation Code.....	41
Figure 5.6 Removal of empty reviews and noise reviews	41
Figure 5.7 Analysis code to showcase the Distribution of Data within the Processed Dataset.....	42
Figure 5.8 Initial Setup of the Sentiment Function.....	43
Figure 5.9 Grid Search for the Best Performing SVM	44
Figure 5.10 Grid Search for the Best Performing Logistic Regression	45
Figure 5.11 Building a Supervised BERTopic.....	46
Figure 5.12 Saving the Models	46
Figure 5.13 Example of Merged Topic Models	47
Figure 5.14 Visualising the topic and target alignment	47
Figure 5.15 Classification report with switching	48
Figure 5.16 Classification report without switching	48
Figure 5.17 Coherence Score calculation.....	49
Figure 5.18 Collection of Execution and evaluation method of Aspect Classification from the merged topic representations.....	50
Figure 5.19 Evaluating the sentiment classification for the merged model.....	50

LIST OF TABLES

Table 2.1 Literature Review.....	10
Table 3.1 Table of Topic Coherence and Topic Diversity Score by (Grootendorst, 2022)	18
Table 4.1 Total reviews for each aspect part 1	34
Table 4.2 Total reviews for each aspect part 2.....	34
Table 4.3 Total sentiment label for each aspect part 1	34
Table 4.4 Total sentiment label for each aspect part 2	34
Table 4.5 Data Cleaning and Preprocessing Example Result	35
Table 6.1 Training multi-aspect Dataset Split.....	51
Table 6.2 Testing multi-aspect Dataset Split	52
Table 6.3 Topic representation of multi-aspect dataset	52
Table 6.4 BERTopic+SVM Merged model multi-aspect dataset performance	54
Table 6.5 BERTopic+LR Merged model Full dataset Performance.....	55
Table 6.6 Sentiment evaluation multi-aspect dataset of Individual BERTopic+SVM	56
Table 6.7 Sentiment evaluation multi-aspect dataset of Individual BERTopic+LR	58
Table 6.8 Training Single Aspect Dataset Split	60
Table 6.9 Testing Single Aspect Dataset Split.....	60
Table 6.10 Topic representation of single-aspect dataset	61
Table 6.11 BERTopic + SVM Merged model Single Aspect Dataset Performance	62
Table 6.12 BERTopic + LR merge model Single Aspect Dataset Performance	63
Table 6.13 BERTopic + SVM single aspect dataset sentiment performance	64
Table 6.14 BERTopic + LR single aspect dataset sentiment performance.....	65