



TRANSFORMER-BASED SENTIMENT ANALYSIS CLASSIFICATION IN NATURAL LANGUAGE PROCESSING FOR BAHASA MELAYU

MOHD ASYRAF BIN ZULKALNAIN

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

MASTER OF SCIENCE IN ELECTRONIC ENGINEERING

2025



Faculty of Electronics and Computer Engineering

**TRANSFORMER-BASED SENTIMENT ANALYSIS
CLASSIFICATION IN NATURAL LANGUAGE PROCESSING FOR
BAHASA MELAYU**

Mohd Asyraf Bin Zulkalnain

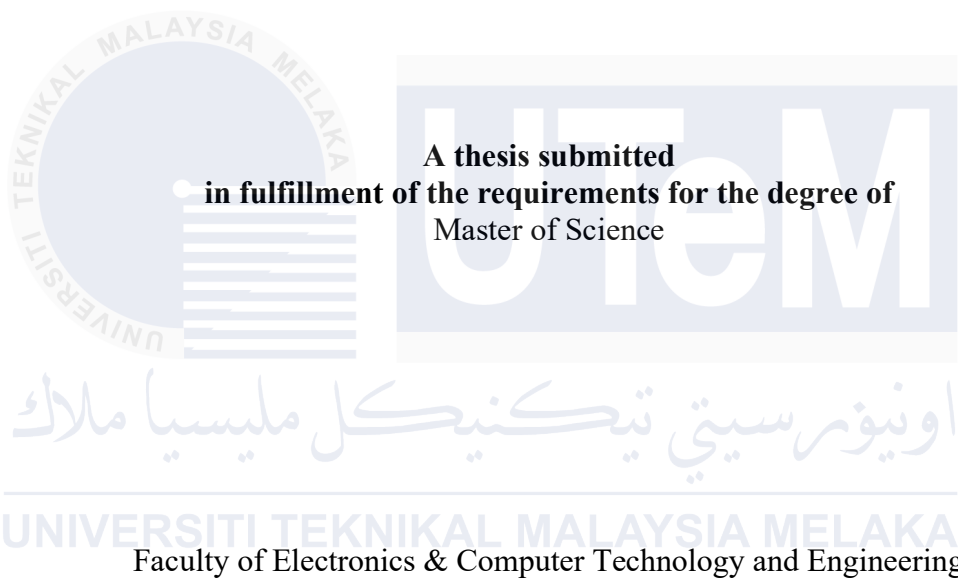


Master of Science in Electronic Engineering

2025

**TRANSFORMER-BASED SENTIMENT ANALYSIS CLASSIFICATION IN
NATURAL LANGUAGE PROCESSING FOR BAHASA MELAYU**

MOHD ASYRAF BIN ZULKALNAIN



UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2025

DECLARATION

I declare that this thesis entitled “Transformer-Based Sentiment Analysis Classification In Natural Language Processing For Bahasa Melayu” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.



Signature

Name



: Mohd Asyraf Bin Zulkalnain

اونيورسيٲي ٲيكنيكل ماليسيا ملاك
Date : 08.02.2025

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Master of Science in Electronic Engineering



Signature

.....

Supervisor Name

: Assoc. Prof. Dr Syafeeza binti Ahmad Radzi

Date

: 09.07.2024

اونيورسيٲى ٲيكنيكل ماليسيا ملاك
UNIVERSITI TEKNIKAL MALAYSIA MELAKA

DEDICATION

I dedicate this thesis to my beautiful parents, Zulkalnain bin Mohd Yussof and Norlela binti Mohd Nordin, and to my family, friends and colleagues. May Allah grant us good health and the strength in remembering Him.



ABSTRACT

Sentiment analysis in Bahasa Melayu leverages Natural Language Processing (NLP) to interpret opinions and emotional tone expressed in Malay texts. This research investigates the application of transformer-based deep learning models—Bidirectional Encoder Representations from Transformers (BERT), DistilBERT, BERT-multilingual, ALBERT, and BERT-CNN—for sentiment classification into positive, negative, and neutral categories. The study addresses challenges in Bahasa Melayu sentiment analysis, including limited annotated resources, linguistic nuances, and common mixed-language usage on platforms like social media. To train and evaluate the models, a large-scale Malay dataset (Malaya dataset) was used. Pretrained models from HuggingFace were fine-tuned using 10-fold cross-validation to improve generalization. Optimization methods such as data augmentation were also implemented. The evaluation considered not just accuracy but also precision, recall, F1 score, and computational efficiency. Among the models, BERT-CNN achieved the best performance, with 96.30% accuracy and consistently high scores across all sentiment classes. BERT also performed well, especially for neutral sentiment, reaching 89.5% accuracy but showed slightly lower recall in the positive class. DistilBERT offered competitive performance (88.96% accuracy) while being faster and more lightweight, making it suitable for deployment in resource-limited environments. BERT-multilingual showed balanced results with a peak accuracy of 89.84%, and ALBERT, despite having fewer parameters, reached 88.76% accuracy but underperformed in positive sentiment recall. The results demonstrate that transformer-based models outperform traditional machine learning and lexicon-based approaches, particularly in handling informal, mixed-language Malay text. The proposed models can support real-world applications such as analyzing consumer sentiment, public opinion, or social response to policies. This study contributes to advancing sentiment analysis for low-resource languages by offering comparative insights and effective model configurations, setting a solid foundation for further research and practical deployment.

**PENGELASAN ANALISIS SENTIMENT BERASASKAN TRANSFORMER DALAM
PEMROSESAN SEMULA JADI UNTUK BAHASA MELAYU**

ABSTRAK

Analisis sentimen dalam Bahasa Melayu memanfaatkan Pemprosesan Bahasa Asli (NLP) untuk mentafsir pendapat dan nada emosi yang dinyatakan dalam teks Melayu. Penyelidikan ini menyiasat aplikasi model pembelajaran mendalam berasaskan transformer—Perwakilan Pengekod Dwi Arah daripada Transformers (BERT), DistilBERT, BERT-berbilang bahasa, ALBERT dan BERT-CNN—untuk klasifikasi sentimen kepada kategori positif, negatif dan neutral. Kajian ini menangani cabaran dalam analisis sentimen Bahasa Melayu, termasuk sumber beranotasi terhad, nuansa linguistik dan penggunaan bahasa campuran biasa pada platform seperti media sosial. Untuk melatih dan menilai model, set data Melayu berskala besar (dataset Malaya) telah digunakan. Model pra-latihan daripada HuggingFace telah diperhalusi menggunakan pengesahan silang 10 kali ganda untuk meningkatkan generalisasi. Kaedah pengoptimuman seperti penambahan data juga dilaksanakan. Penilaian itu bukan sahaja mempertimbangkan ketepatan tetapi juga ketepatan, ingatan semula, skor F1, dan kecekapan pengiraan. Antara model, BERT-CNN mencapai prestasi terbaik, dengan ketepatan 96.30% dan skor tinggi secara konsisten merentas semua kelas sentimen. BERT juga menunjukkan prestasi yang baik, terutamanya untuk sentimen neutral, mencapai ketepatan 89.5% tetapi menunjukkan ingatan lebih rendah sedikit dalam kelas positif. DistilBERT menawarkan prestasi kompetitif (88.96% ketepatan) sambil lebih pantas dan lebih ringan, menjadikannya sesuai untuk digunakan dalam persekitaran terhad sumber. BERT-berbilang bahasa menunjukkan hasil yang seimbang dengan ketepatan puncak 89.84%, dan ALBERT, walaupun mempunyai parameter yang lebih sedikit, mencapai ketepatan 88.76% tetapi kurang berprestasi dalam ingatan sentimen positif. Keputusan menunjukkan bahawa model berasaskan pengubah mengatasi pembelajaran mesin tradisional dan pendekatan berasaskan leksikon, terutamanya dalam mengendalikan teks bahasa Melayu tidak formal dan bercampur. Model yang dicadangkan boleh menyokong aplikasi dunia nyata seperti menganalisis sentimen pengguna, pendapat umum atau tindak balas sosial terhadap dasar. Kajian ini menyumbang kepada memajukan analisis sentimen untuk bahasa sumber rendah dengan menawarkan cerapan perbandingan dan konfigurasi model yang berkesan, menetapkan asas yang kukuh untuk penyelidikan lanjut dan penggunaan praktikal.

ACKNOWLEDGEMENT

In the Name of Allah, the Most Gracious, the Most Merciful. First and foremost, I would like to take this opportunity to express my sincere acknowledgement to my beautiful parents who have raised me until I grow up to become a responsible person. My mother Norlela binti Mohd Nordin and my father Zulkalnain bin Mohd Yusoff. Also I would like to thank my family members, teachers, friends and colleagues for their teachings as they are my life's teachers. I would also like to express my sincere gratitude and acknowledgement to my supervisor, Dr. Syafeeza Ahmad Radzi for her guidance, support, knowledge and encouragement throughout the completion of this thesis.

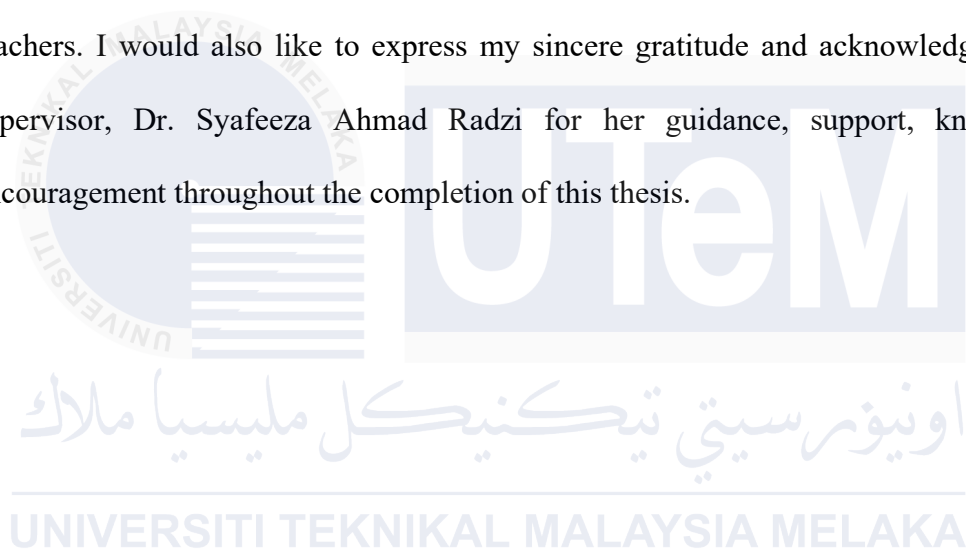


TABLE OF CONTENTS

	PAGES
DECLARATION	4
DEDICATION	6
ABSTRACT	7
ABSTRAK	8
ACKNOWLEDGEMENT	9
TABLE OF CONTENTS	10
LIST OF TABLES	12
LIST OF FIGURES	15
LIST OF ABBREVIATIONS	19
LIST OF SYMBOLS	21
LIST OF APPENDICES	22
LIST OF PUBLICATIONS	23
CHAPTER 1	24
INTRODUCTION	24
1.1 Background	24
1.2 Problem Statement	25
1.3 Research Question	27
1.4 Research Objective	28
1.5 Scope of Research	28
1.6 Thesis Outline	30
CHAPTER 2	31
LITERATURE REVIEW	31
2.1 Introduction	31
2.1.1 Natural Language Processing (NLP)	31
2.1.2 Sentiment Analysis (SA)	32
2.2 Lexicon-based and Machine Learning (ML) Methods in Sentiment Analysis	33
2.3 Types of Deep Learning Architecture	35
2.3.1 Deep Learning and Neural Networks in Sentiment Analysis	36
2.3.2 Previous works of Sentiment Analysis in Malay language	45
2.4 Transformer Architecture	54
2.4.1 Bidirectional Encoder Representations from Transformer (BERT)	55
2.5 Transfer learning	59
2.6 Common Tasks Used For BERT	60
2.7 Comparison between Lexicon-based, ML, CNN, Hybrid & BERT approaches from past researchers	64
2.8 Summary	68
CHAPTER 3	71
METHODOLOGY	71
3.1 Introduction	71
3.2 List of tools utilized in this study	74
3.2.1 Python	75
3.2.2 Jupyterlab	75

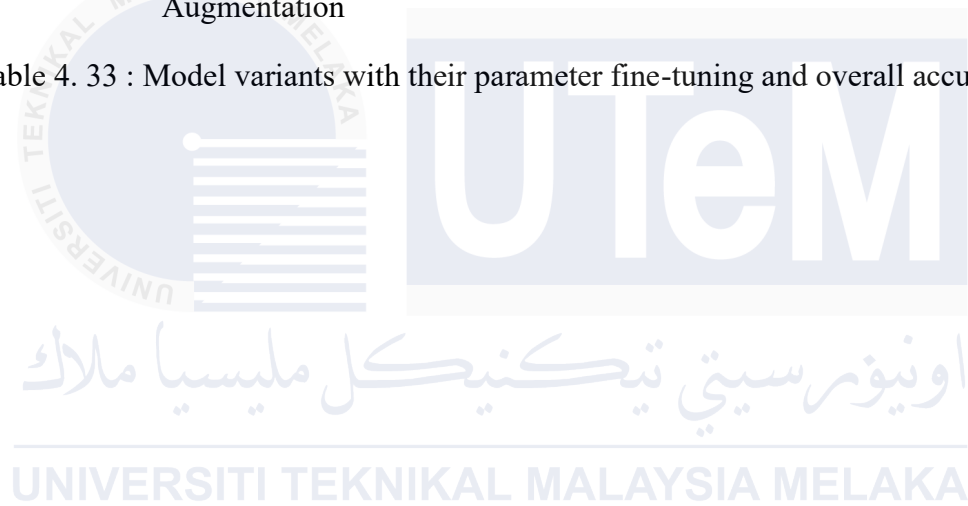
3.2.3	Pandas	75
3.2.4	NumPy	76
3.2.5	Matplotlib	76
3.2.6	Scikit-learn	77
3.2.7	Scipy	78
3.2.8	Torch	78
3.2.9	Transformers	79
3.3	The Proposed Methodology and Experimental setup	79
3.3.1	Data Preparation	81
3.3.2	Model Design	88
3.3.3	Performance Metrics and Data Analysis	93
3.4	Comparison of Model's Complexity	101
3.5	Summary	106
CHAPTER 4		108
RESULT AND DISCUSSION		108
4.1	Introduction	108
4.2	Main Architecture Results	108
4.2.1	BERT result	109
4.2.2	DistilBERT result	122
4.2.3	BERT-multilingual result	135
4.2.4	ALBERT result	149
4.2.5	BERT-CNN result	161
4.3	Comparisons: Main Architectures	177
4.3.1	Results of Model Training Before Augmentation	177
4.3.2	Results of Model Training After Augmentation	180
4.4	Improvements	184
4.5	Summary	187
CHAPTER 5		189
CONCLUSION AND RECOMMENDATIONS FOR FUTURE RESEARCH		189
5.1	Introduction	189
5.2	Summary of the Research Objectives	189
5.3	Research Contributions	190
5.4	Practical Implications and Beneficiaries	191
5.5	Limitations of the Present Study	192
5.6	Future Works	194
5.7	Summary	196
REFERENCES		197
APPENDICES		201
Appendix A: Malaya Dataset Sample (Training)		202
Appendix B: Malaya Dataset Sample (Testing)		203

LIST OF TABLES

TABLE	TITLE	PAGE
Table 1.1 :	Scope of Work	29
Table 2.1	Malay sentiment lexicon with its polarity	46
Table 2.2	Example of Malay stopwords and their translation in English	47
Table 2.3	Dataset used and accuracies obtained by BERT authors (Devlin et al., 2019)	64
Table 2.4	Comparison of Lexicon-based, ML, CNN, Hybrid & BERT approaches from past researchers that conducted Malay sentiment analysis	66
Table 3.1	List of utilized softwares and dependencies used and their version	74
Table 3.2	Summary of text cleaning operations that are performed in this research	82
Table 3.3	Example of Tokenization	85
Table 3.4	Data Augmentation strategies and their definitions	86
Table 3.5	Example of synonym dictionary in Malay	87
Table 3.6	Variations of BERT models used in this research	101
Table 3.7	Parameter settings of different experiments for BERT, DistilBERT, BERT-multilingual and ALBERT	103
Table 3.8	The parameters of CNN architecture used for Hybrid Malaya BERT-CNN for each experiment	105
Table 4. 1 :	BERT Experiment 1 Precision, Recall, F1-Score and Support	115
Table 4. 2 :	Sentence prediction performed from BERT Experiment 1	117
Table 4. 3 :	BERT Experiment 2 Precision, Recall, F1-Score and Support	118

Table 4. 4 : Sentence prediction performed from BERT Experiment 2	119
Table 4. 5 : BERT Experiment 3 Precision, Recall, F1-Score and Support	120
Table 4. 6 : Sentence prediction performed from BERT Experiment 3	121
Table 4. 7 : DistilBERT Experiment 1 Precision, Recall, F1-score and Support	128
Table 4. 8 : Sentence prediction performed from DistilBERT Experiment 1	130
Table 4. 9 : DistilBERT Experiment 2 Precision, Recall, F1-score and Support	131
Table 4. 10 : Sentence prediction performed from DistilBERT Experiment 2	132
Table 4. 11 : DistilBERT Experiment 3 Precision, Recall, F1-score and Support	133
Table 4. 12 : Sentence prediction performed from DistilBERT Experiment 3	134
Table 4. 13 : BERT-multilingual Experiment 1 Precision, Recall, F1-score and Support	142
Table 4. 14 : Sentence prediction performed from BERT-multilingual Experiment 1	144
Table 4. 15 : BERT-multilingual Experiment 2 Precision, Recall, F1-score and Support	145
Table 4. 16 : Sentence prediction performed from BERT-multilingual Experiment 2	146
Table 4. 17 : BERT-multilingual Experiment 3 Precision, Recall, F1-score and Support	147
Table 4. 18 : Sentence prediction performed from BERT-multilingual Experiment 3	148
Table 4. 19 : ALBERT Experiment 1 Precision, Recall, F1-score and Support	155
Table 4. 20 : Sentence prediction performed from ALBERT Experiment 1	156
Table 4. 21 : ALBERT Experiment 2 Precision, Recall, F1-score and Support	157
Table 4. 22 : Sentence prediction performed from ALBERT Experiment 2	158
Table 4. 23 : ALBERT Experiment 3 Precision, Recall, F1-score and Support	159
Table 4. 24 : Sentence prediction performed from ALBERT Experiment 3	160
Table 4. 25 : BERT-CNN Experiment 1 Precision, Recall, F1-score and Support	166

Table 4. 26 : BERT-CNN Experiment 2 Precision, Recall, F1-score and Support	168
Table 4. 27 : BERT-CNN Experiment 3 Precision, Recall, F1-score and Support	170
Table 4. 28 : BERT-CNN Experiment 4 Precision, Recall, F1-score and Support	172
Table 4. 29 : BERT-CNN Experiment 5 Precision, Recall, F1-score and Support	174
Table 4. 30 : Sentence prediction performed from BERT-CNN	176
Table 4. 31 : Performance comparison for all 5 model variants of BERT against two baseline models	177
Table 4. 32 : Performance comparison for all 5 model variants of BERT After Augmentation	180
Table 4. 33 : Model variants with their parameter fine-tuning and overall accuracy	185



LIST OF FIGURES

FIGURE	TITLE	PAGE
Figure 2.1 :	CNNs architecture (Yadav, 2020)	37
Figure 2.2 :	Example of pooling operation (Mahajan, P.,2020)	39
Figure 2.3 :	Recurrent neural network unrolled into a 4-layer neural network (Wang et al., 2016)	40
Figure 2.4 :	Chain structure of LSTMs (Hochreiter & Schmidhuber, 1997)	41
Figure 2.5 :	Internal structure of GRUs (Cho et al., 2014)	43
Figure 2.6 :	Transformer model architecture (Vaswani et al., 2017)	54
Figure 2.7 :	Left: BERT Pre-training architecture, Right: BERT Fine-tuning architecture (Devlin et al., 2019)	56
Figure 2.8 :	BERT base ($L=12$, $H=768$, $A=12$, Total Parameters=110M) and BERT large ($L=24$, $H=1024$, $A=16$, Total Parameters=340M)	59
Figure 2.9 :	BERT Mechanism (Devlin et al., 2019)	62
Figure 2.10 :	Next Sentence Prediction task in BERT training	63
Figure 3. 1:	Overall flowchart of the research	73
Figure 3. 2:	Training and Testing Data for 3 classes ‘Negative’, ‘Neutral’ and ‘Positive’	80
Figure 3. 3:	Class-wise data distribution for training data and testing data before and after augmentation	88
Figure 3. 4:	Hybrid Malaya BERT-CNN model	92
Figure 3. 5:	Example of confusion matrix	93
Figure 3. 6:	Example of Training and Validation loss vs Epoch Plot	95
Figure 3. 7:	Example of 10-fold cross validation (Dantas, J., 2020)	99
Figure 3. 8:	Training and validation losses for K -fold	100

Figure 4. 1 : Mean of 10-fold training and validation loss for BERT Experiment 1	1109
Figure 4. 2 : F1-Score (weighted) Across 10 Folds for BERT Experiment 1	110
Figure 4. 3 : Mean of 10-fold training and validation loss for BERT Experiment 2	111
Figure 4. 4 : F1-Score (weighted) Across 10 Folds for BERT Experiment 2	112
Figure 4. 5 : Mean of 10-fold training and validation losses for BERT Experiment 3	113
Figure 4. 6 : F1-Score (weighted) Across 10 Folds for BERT Experiment 3	114
Figure 4. 7 : Confusion Matrix for BERT Experiment 1 (Left: Before augmentation, Right: After augmentation)	115
Figure 4. 8 : Confusion Matrix for BERT Experiment 2 (Left: Before augmentation, Right: After augmentation)	117
Figure 4. 9 : Confusion Matrix for BERT Experiment 3 (Left: Before augmentation, Right: After augmentation)	119
Figure 4. 10 : Mean Training and Validation Loss for DistilBERT training for Experiment 1	122
Figure 4. 11 : DistilBERT Experiment 1 F1-Score (weighted) Across 10-Folds	123
Figure 4. 12 : Mean Training and Validation Loss Across 10-Folds for DistilBERT Experiment 2	124
Figure 4. 13 : DistilBERT F1-Score (weighted) Across 10 Folds for Experiment 2	125
Figure 4. 14 : Mean Training and Validation Loss Across 10-Folds for DistilBERT Experiment 3	126
Figure 4. 15 : DistilBERT F1-Score (weighted) Across 10 Folds for Experiment 3	127
Figure 4. 16 : Confusion Matrix for DistilBERT Experiment 1 (Left: Before augmentation, Right: After augmentation)	128
Figure 4. 17 : Confusion Matrix for DistilBERT Experiment 2 (Left: Before augmentation, Right: After augmentation)	131
Figure 4. 18 : Confusion Matrix for DistilBERT Experiment 3 (Left: Before augmentation, Right: After augmentation)	133
Figure 4. 19 : Mean Training and Validation Loss Across 10-folds for BERT-multilingual training for Experiment 1	136

Figure 4. 20 : BERT-multilingual F1-Score(weighted) Across 10 Folds for Experiment 1	137
Figure 4. 21 : Mean Training and Validation Loss Across 10-folds for BERT-multilingual training for Experiment 2	138
Figure 4. 22 : BERT-multilingual F1-Score(weighted) Across 10 Folds for Experiment 2	139
Figure 4. 23 : Mean Training and Validation Loss Across 10-folds for BERT-multilingual training for Experiment 3	140
Figure 4. 24 : BERT-multilingual F1-Score(weighted) Across 10 Folds for Experiment 3	141
Figure 4. 25 : Confusion Matrix for BERT-multilingual Experiment 1(Left: Before augmentation, Right: After augmentation)	142
Figure 4. 26 : Confusion Matrix for BERT-multilingual Experiment 2(Left: Before augmentation, Right: After augmentation)	144
Figure 4. 27 : Confusion Matrix for BERT-multilingual Experiment 3(Left: Before augmentation, Right: After augmentation)	147
Figure 4. 28 : Mean Training and Validation Loss Across 10-fold for ALBERT Experiment 1	149
Figure 4. 29 : ALBERT F1-Score(weighted) Across 10 Folds for Experiment 1	150
Figure 4. 30 : Mean Training and Validation Loss Across 10-fold for ALBERT training for Experiment 2	151
Figure 4. 31 : ALBERT F1-Score(weighted) Across 10 Folds for Experiment 2	152
Figure 4. 32 : Mean Training and Validation Loss Across 10-fold for ALBERT training for Experiment 3	153
Figure 4. 33 : ALBERT F1-Score(weighted) Across 10 Folds for Experiment 3	154
Figure 4. 34 : Confusion Matrix for ALBERT Experiment 1(Left: Before augmentation, Right: After augmentation)	155
Figure 4. 35 : Confusion Matrix for ALBERT Experiment 2(Left: Before augmentation, Right: After augmentation)	157
Figure 4. 36 : Confusion Matrix for ALBERT Experiment 3(Left: Before augmentation, Right: After augmentation)	159

Figure 4. 37 : BERT-CNN Training and Validation Loss for 5 Experiments	161
Figure 4. 38 : BERT-CNN Training and Validation Loss for 5 Experiments After Augmentation	163
Figure 4. 39 : BERT-CNN F1-Score (weighted)	164
Figure 4. 40 : BERT-CNN F1-Score (weighted) After Augmentation	165
Figure 4. 41 : Confusion matrix for BERT-CNN Experiment 1(Left: Before augmentation, Right: After augmentation)	166
Figure 4. 42 : Confusion matrix for BERT-CNN Experiment 2(Left: Before augmentation, Right: After augmentation)	168
Figure 4. 43 : Confusion matrix for BERT-CNN Experiment 3(Left: Before augmentation, Right: After augmentation)	170
Figure 4. 44 : Confusion matrix for BERT-CNN Experiment 4(Left: Before augmentation, Right: After augmentation)	172
Figure 4. 45 : Confusion matrix for BERT-CNN Experiment 5(Left: Before augmentation, Right: After augmentation)	174
Figure 4.46 : Model performance comparison Before and After Augmentation	182
Figure 4. 47 : Overall Accuracy Across Models with Different Experiments	187

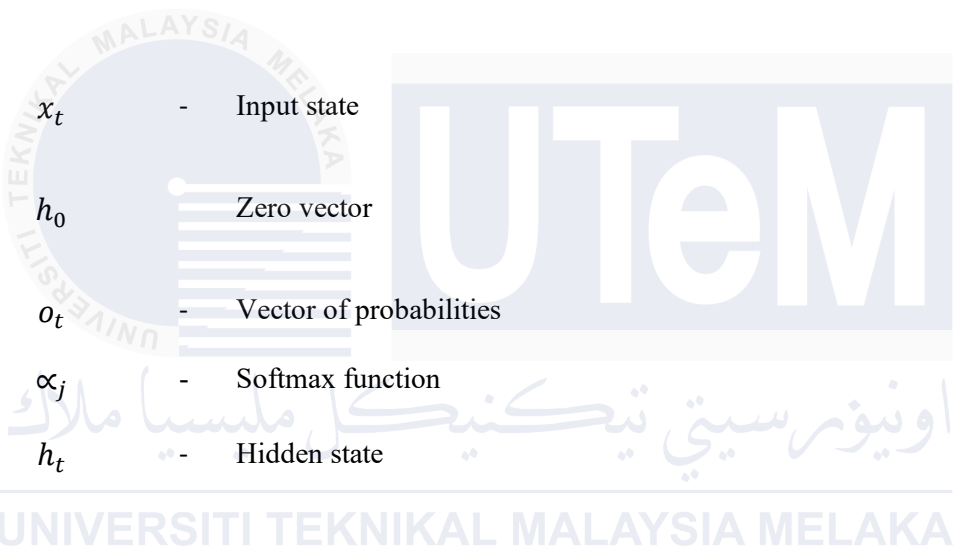
LIST OF ABBREVIATIONS

<i>AI</i>	-	Artificial Intelligence
<i>AP</i>	-	Average Precision
<i>API</i>	-	Application Programmable Interface
<i>ASCII</i>	-	American Standard Code for Information Interchange
<i>AUC-ROC</i>	-	Area Under the Receiver Operating Characteristic Curve
<i>BERT</i>	-	Bidirectional Encoder Representations from Transformers
<i>BPE</i>	-	Byte-Pair Encoding
<hr/>		
<i>CNN</i>	-	Convolutional Neural Network
<i>CPU</i>	-	Central Processing Unit
<i>CSV</i>	-	Comma-Separated Values
<i>EDA</i>	-	Easy Data Augmentation
<i>ELMo</i>	-	Embeddings from Language Models
<i>GLUE</i>	-	General Language Understanding Evaluation
<i>GPT</i>	-	Generative Pre-trained Transformer
<i>GPU</i>	-	Graphical Processing Unit
<i>GRU</i>	-	Gated Recurrent Unit
<i>GUI</i>	-	Graphical User Interface

<i>ICA</i>	-	Independent Component Analysis
<i>IDF</i>	-	Inverse Document Frequency
<i>KNN</i>	-	K-Nearest Neighbor
<i>LLMs</i>	-	Large Language Models
<i>LSTM</i>	-	Long Short-Term Memory
<i>M-BERT</i>	-	BERT-multilingual
<i>ML</i>	-	Machine Learning
<i>MLM</i>	-	Masked Language Model
<i>MRC</i>	-	Malay Reviews Corpus
<i>NB</i>	-	Naive Bayes
<i>NLP</i>	-	Natural Language Processing
<i>NLTK</i>	-	Natural Language Tool Kit
<i>NSP</i>	-	Next Sentence Prediction
<i>PCA</i>	-	Principal Component Analysis
<i>PRC</i>	-	Precision-Recall Curve
<i>RF</i>	-	Random Forest
<i>RNN</i>	-	Recurrent Neural Network
<i>SA</i>	-	Sentiment Analysis
<i>SQL</i>	-	Structured Query Language
<i>SQuAD</i>	-	Stanford Question and Answering Dataset
<i>SVM</i>	-	Support Vector Machines
<i>SWAG</i>	-	Situations with Adversarial Generation
<i>TF</i>	-	Term Frequency

$TF-IDF$	-	Term Frequency-Inverse Document Frequency
URL	-	Uniform Resource Locator

LIST OF SYMBOLS



x_t	-	Input state
h_0	-	Zero vector
o_t	-	Vector of probabilities
α_j	-	Softmax function
h_t	-	Hidden state

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

LIST OF APPENDICES

APPENDIX	TITLE	PAGE
Appendix A	Malaya Dataset Sample (Training)	202
Appendix B	Malaya Dataset Sample (Testing)	203



LIST OF PUBLICATIONS

The followings are the list of publications related to the work on this thesis:

Mohd Asyraf Zulkalnain, A.R., Syafeeza, W.H.M. and Rahaman, S., 2025. Evaluation of Transformer-Based Models for Sentiment Analysis in Bahasa Malaysia. *Journal of Telecommunication, Electronic and Computer Engineering*, Vol. X, No. X, pp. XX-XX. Available at: <https://doi.org/10.54554/jtec.202x.xx.xx.xxx>.



CHAPTER 1

INTRODUCTION

1.1 Background

Sentiment analysis is the process of determining the opinion or emotional tone of a given text and involves the use of Natural Language Processing (NLP). Natural language processing (NLP) focuses on techniques which enable computers to understand, interpret and generate human language. In addition to identifying particular feelings and viewpoints expressed in the text, this can also involve assessing the general sentiment of the text—whether it be positive, negative, or neutral. NLP also makes it possible to interpret words, such as sentiments.

Applications for sentiment analysis are numerous and include evaluating consumer reviews, identifying the tone of news articles, and reviewing the tone of social media posts. For instance, an organization may employ sentiment analysis to automatically categorize customer reviews as neutral, negative, or positive in order to instantly effectively determine the general perspective of its customers. Another example of sentiment analysis is detecting the sentiment of social media posts about a particular topic or brand. This can be useful for businesses, as it can help them understand how people are feeling about their products or services, and identify areas for improvement.