



UNDERGRADUATE THESIS

IMPLEMENTATION OF A BROWSER EXTENSION FOR BILINGUAL CYBERBULLYING DETECTION ON SOCIAL MEDIA USING BILSTM AND LIME EXPLANATION

MOCHAMMAD DAFFA FAIQ HUSIN SYAHPUTRA
NPM 22081010161

THESIS ADVISORS

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
Yisti Vita Via, S.ST., M.Kom.

**MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FACULTY OF COMPUTER SCIENCE
INFORMATICS STUDY PROGRAM
SURABAYA
2026**

APPROVAL SHEET

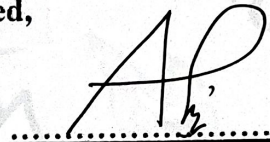
IMPLEMENTATION OF A BROWSER EXTENSION FOR BILINGUAL CYBERBULLYING DETECTION ON SOCIAL MEDIA USING BILSTM AND LIME EXPLANATION

By:
MOCHAMMAD DAFFA FAIQ HUSIN SYAHPUTRA
NPM. 22081010161

Has been defended before, and accepted by, the Board of Assessors of the Thesis Examination of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, on **July 10, 2026**:

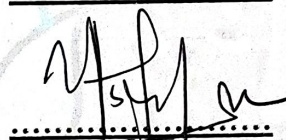
Approved,

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
NPT. 222198 60 816400



(Advisor I)

Yisti Vita Via, S.ST., M.Kom.
NIP. 19860425 202121 2 001



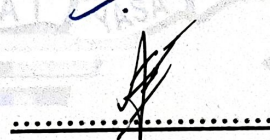
(Advisor II)

Dr. Intan Yuniar Purbasari, S.Kom., MSc.
NIP. 19800602 202521 2 029



(Head Assessor)

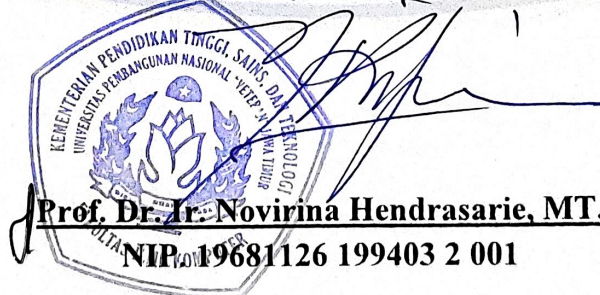
Budi Mukhamad Mulyo, S.Kom., M.T.
NIP. 19891118 202406 1 003



(Assessor I)

Acknowledge by,

Dean of the Faculty of Computer Science



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

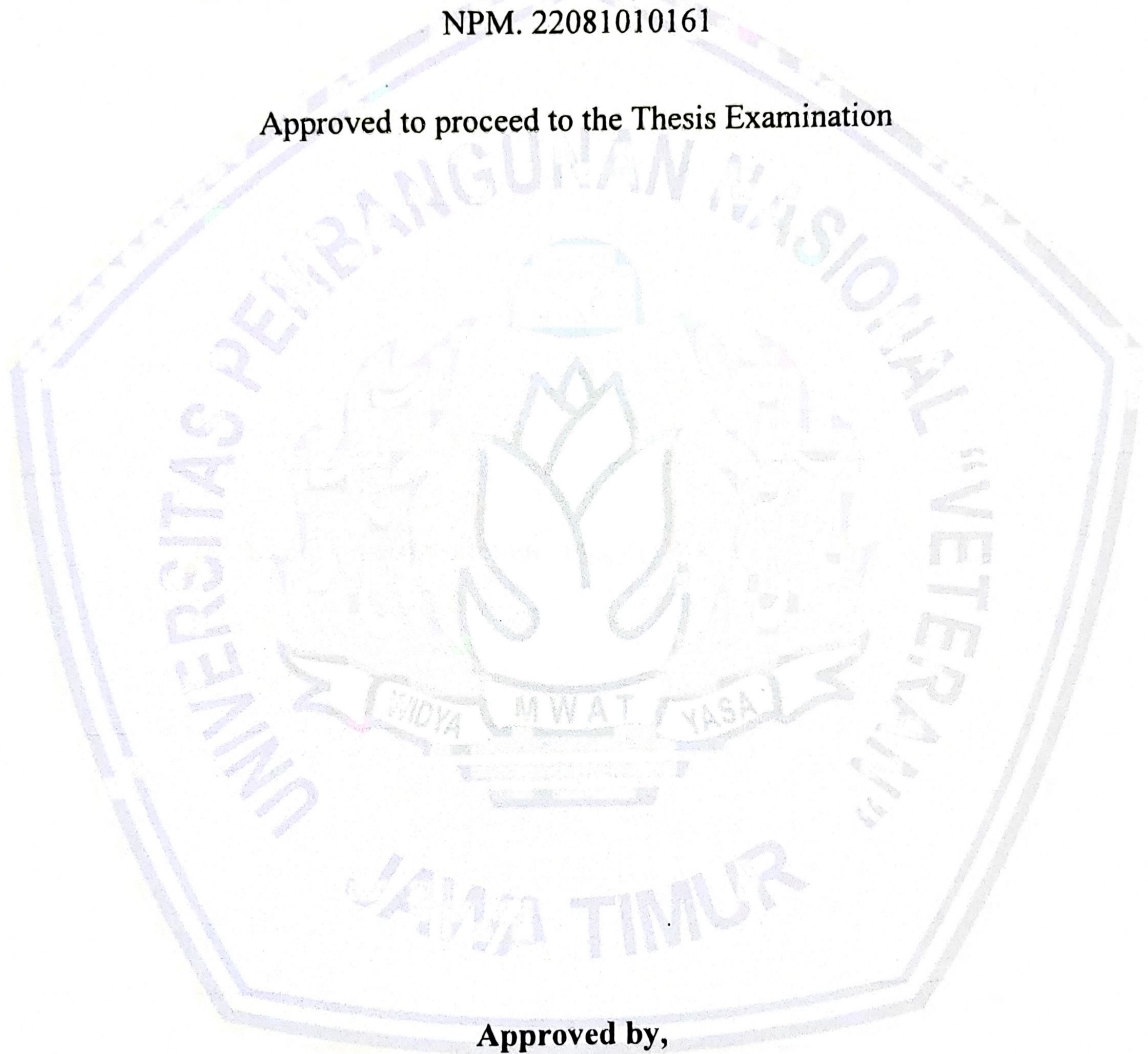
APPROVAL SHEET

IMPLEMENTATION OF A BROWSER EXTENSION FOR BILINGUAL CYBERBULLYING DETECTION ON SOCIAL MEDIA USING BILSTM AND LIME EXPLANATION

By:

MOCHAMMAD DAFFA FAIQ HUSIN SYAHPUTRA
NPM. 22081010161

Approved to proceed to the Thesis Examination



Approved by,

**Coordinator of Informatics Study Program
Faculty of Computer Science**

A handwritten signature in blue ink, appearing to read 'Intan', is placed above the name of the coordinator.

Dr. Intan Yuniar Purbasari, S.Kom., MSc.
NIP. 19800602 202521 2 029

STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name : Mochammad Daffa Faiq Husin Syahputra
NPM : 22081010161
Degree Program : Bachelor (S1)
Study Program : Informatics
Faculty : Faculty of Computer Science

Hereby declares that this undergraduate thesis contains no part of any other scientific work that has been submitted to obtain an academic degree at any higher education institution. Furthermore, it does not contain any work or opinions previously written or published by others, except for those which are explicitly cited in this thesis and listed completely in references.

And I declare that this scientific document is free from elements of plagiarism. If in the future indications of plagiarism are found in this Thesis, I am willing to accept sanctions in accordance with the applicable laws and regulations.

Thus, I made this statement without any coercion from anyone and to be used as it should.



Surabaya, July 15, 2026
Declarant,



MOCHAMMAD DAFFA FAIQ H. S.
NPM. 22081010161

ABSTRACT

Student Name / NPM : Mochammad Daffa Faiq H. S. / 22081010161
Thesis Title : Implementation Of A Browser Extension For
Bilingual Cyberbullying Detection On Social
Media Using BiLSTM And LIME Explanation
Advisor : 1. Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
2. Yisti Vita Via, S.ST., M.Kom.

Cyberbullying has become an increasingly prevalent issue due to the rapid growth of social media use. Although Instagram has implemented various protective features, the available mechanisms are not yet fully capable of protecting users from the risks of cyberbullying, particularly within comment sections that enable direct interaction. At the same time, the increasing prevalence of mixed Indonesian and English language usage on social media adds complexity to the cyberbullying detection process. In response to these challenges, this study proposes the development of a browser extension capable of automatically detecting and filtering cyberbullying-related comments in Instagram using the Bidirectional Long Short-Term Memory (BiLSTM) method integrated with Local Interpretable Model-Agnostic Explanations (LIME). This study utilizes a dataset gathered from multiple sources, comprising 25,276 Indonesian and English text samples classified into two classes, normal and bullying. The experimental results show that the best performing model attained an accuracy of 95.97%, a precision of 95.98%, a recall of 95.97%, and an F1-score of 95.96%. These findings demonstrate that the BiLSTM model can effectively distinguish between normal and bullying comments while maintaining high classification accuracy and strong generalization across bilingual datasets. Furthermore, the integration of LIME enables the system to provide explanations for its predictions by highlighting the words or phrases that contribute most significantly to the prediction process. Therefore, the proposed system not only performs automates cyberbullying detection but also improves the transparency and interpretability of its predictions for users.

Keywords: Bidirectional Long Short-Term Memory, Cyberbullying, FastText, Local Interpretable Model-Agnostic Explanations, Natural Language Processing

ACKNOWLEDGEMENTS

The author would like to express sincere gratitude to Almighty God for His abundant blessings, grace, and guidance, which have enabled the successful completion of this undergraduate thesis entitled, **“Implementation Of A Browser Extension For Bilingual Cyberbullying Detection On Social Media Using BiLSTM And LIME Explanation”**. This thesis is submitted in partial fulfillment of the requirements for the Bachelor’s degree in Informatics and serves as a contribution to research in the field of Natural Language Processing (NLP), particularly in bilingual cyberbullying detection and explainable artificial intelligence.

The author acknowledges that this thesis could not have been completed without the invaluable guidance, support, encouragement, and assistance provided by numerous individuals and institutions. Accordingly, the author wishes to extend sincere appreciation and heartfelt gratitude to:

1. Mrs. Prof. Dr. Ir. Novirina Hendrasarie, MT., as the Dean of the Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Mrs. Dr. Intan Yuniar Purbasari, S.Kom. MSc., as the Coordinator of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Mrs. Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT., as the First Thesis Advisor, for the invaluable guidance, time, knowledge, and continuous support provided throughout the preparation and completion of this thesis.
4. Mrs. Yisti Vita Via, S.ST., M.Kom., as the Second Thesis Advisor, for generously sharing time, insights, constructive feedback, and valuable suggestions that greatly contributed to the improvement of this thesis.
5. Mrs. Dr. Intan Yuniar Purbasari, S.Kom., MSc., as the Head Assessor, for the valuable comments and recommendations provided during the thesis examination, which significantly enhanced the quality of this work.
6. Mr. Budi Mukhamad Mulyo, S.Kom., M.T., as the Thesis Assessor, for the constructive suggestions, particularly regarding the technical aspects of this research, which greatly contributed to the refinement and improvement of this thesis.
7. All lecturers and administrative staff of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur, for

their knowledge, guidance, and assistance throughout the author's academic journey.

8. The author's beloved parents, for their unconditional love, unwavering encouragement, moral support, financial assistance, and countless sacrifices. This achievement would not have been possible without their continuous prayers and steadfast belief in the author.
9. Azzazahra Marega Dwita Putri, for her unwavering encouragement, motivation, and support as a close friend and companion throughout the completion of this thesis.
10. Everyone who has contributed, directly or indirectly, but whose names cannot all be mentioned individually, for their support, assistance, and companionship throughout the author's academic journey.

The author recognizes that this thesis is not without limitations and welcomes constructive comments and suggestions for its further improvement. It is hoped that that this study will make a meaningful contribution to the development of NLP, especially in bilingual cyberbullying detection and explainable artificial intelligence, while also serving as a valuable reference for future studies in related fields.

Surabaya, July 15th 2026

Author

TABLE OF CONTENTS

APPROVAL SHEET	i
APPROVAL SHEET	iii
STATEMENT OF ORIGINALITY	v
ABSTRACT	vii
ACKNOWLEDGEMENTS.....	ix
TABLE OF CONTENTS.....	xi
TABLE OF FIGURES.....	xv
LIST OF TABLES	xix
CHAPTER I INTRODUCTION.....	1
1.1. Background	1
1.2. Problem Statements.....	4
1.3. Research Objectives	4
1.4. Research Contributions	4
1.5. Research Limitations.....	5
CHAPTER II LITERATURE REVIEW	7
2.1. Previous Studies	7
2.2. Natural Language Processing (NLP)	11
2.3. Word Embedding	12
2.4. Long Short-Term Memory (LSTM)	17
2.4.1. Forget Gate.....	18
2.4.2. Input Gate.....	19
2.4.3. Output Gate	21
2.5. Bidirectional Long Short-Term Memory (BiLSTM).....	22
2.6. Fully Connected Layer.....	24
2.7. Hyperparameter.....	25
2.7.1. Learning Rate	25
2.7.2. Epoch	26
2.7.3. Batch Size	26
2.7.4. Hidden Unit.....	27
2.7.5. Dropout Rate	27
2.7.6. Optimizer	28
2.7.7. Sequence Length	28

2.7.8. Threshold	30
2.8. Evaluation Metrics	31
2.9. Parameter Tuning Methods	32
2.10. Early Stopping.....	33
2.11. Local Interpretable Model-agnostic Explanations (LIME).....	34
2.12. Web Browser.....	37
2.13. Browser Extension	38
2.14. HTML	39
2.15. CSS.....	40
2.16. JavaScript	40
2.17. Document Object Model (DOM)	41
2.18. REST API	41
CHAPTER III RESEARCH METHODOLOGY	43
3.1. Research Methods	43
3.2. Literature Study.....	43
3.3. System Design Analysis.....	44
3.4. Data Collection.....	45
3.5. Data Preprocessing.....	52
3.5.1. Case Folding	53
3.5.2. Text Cleaning	54
3.5.3. Tokenizing	56
3.5.4. Normalization.....	56
3.5.5. Stopword Removal.....	58
3.5.6. Lemmatization	59
3.6. Feature Extraction	59
3.7. Model Design.....	63
3.7.1. Spatial Dropout 1D Layer	64
3.7.2. Bidirectional LSTM Layer.....	65
3.7.3. Fully Connected Layer.....	67
3.8. Testing and Evaluation.....	69
3.9. Explainable AI	70
3.10. Backend Design	75
3.11. Browser Extension Design.....	76
CHAPTER IV RESULTS AND DISCUSSION	79

4.1. Data Collection.....	79
4.2. Data Preprocessing.....	90
4.3. Feature Extraction	100
4.4. Model Implementation	102
4.5. Sequence Length (Percentile) Evaluation	104
4.5.1. 20th Percentile.....	105
4.5.2. 50th Percentile.....	106
4.5.3. 90th Percentile.....	108
4.6. Data Split Evaluation	110
4.6.1. 60:20:20 Split Ratio	110
4.6.2. 70:20:10 Split Ratio	112
4.6.3. 80:10:10 Split Ratio	114
4.7. Batch Size Evaluation	116
4.7.1. Batch Size 16	116
4.7.2. Batch Size 32	118
4.7.3. Batch Size 64	120
4.8. Learning Rate Evaluation.....	122
4.8.1. Learning Rate 0.01	122
4.8.2. Learning Rate 0.001	124
4.8.3. Learning Rate 0.0001	126
4.9. LSTM Hidden Unit Evaluation.....	128
4.9.1. LSTM Hidden Unit 32	129
4.9.2. LSTM Hidden Unit 64	130
4.9.3. LSTM Hidden Unit 128	132
4.10. Dropout Rate Evaluation.....	134
4.10.1. Dropout Rate 0.2	135
4.10.2. Dropout Rate 0.3	136
4.10.3. Dropout Rate 0.4	138
4.11. Threshold Evaluation	140
4.11.1. Threshold 0.3	141
4.11.2. Threshold 0.5	142
4.11.3. Threshold 0.7	144
4.12. Analysis of Model Evaluation Results.....	146
4.13. Explainable AI Implementation	147

4.14. Backend Implementation	153
4.15. Browser Extension Implementation.....	157
CHAPTER V CONCLUSIONS AND SUGGESTIONS.....	165
5.1. Conclusions	165
5.2. Suggestions	166
BIBLIOGRAPHY	167
ATTACHMENTS	177

TABLE OF FIGURES

Figure 2.1 Relationship of NLP to Computer Science and Linguistics [28]	12
Figure 2.2 Word Embedding for Word Representation	13
Figure 2.3 Illustration of Character n-grams [37]	14
Figure 2.4 Illustration of CBOW [37]	15
Figure 2.5 Representation of the Target Word [37]	15
Figure 2.6 Illustration of Negative Sampling [37]	16
Figure 2.7 LSTM Architecture [42]	17
Figure 2.8 Structure of the Forget Gate [44]	18
Figure 2.9 Structure of the Input Gate [44]	19
Figure 2.10 Structure of the Output Gate [44]	21
Figure 2.11 BiLSTM Architecture [46]	23
Figure 2.12 Fully Connected Layer [47]	24
Figure 2.13 Structure of the Confusion Matrix [64]	31
Figure 2.14 Local Explanation Visualization [71]	34
Figure 2.15 Binary Bag of Words [72]	35
Figure 2.16 Perturbation Illustration [71]	36
Figure 2.17 Browser Market Share in Indonesia [77]	38
Figure 2.18 Browser Extension Architecture [79]	39
Figure 2.19 HTML Representation in the DOM [83]	41
Figure 3.1 Research Stages	43
Figure 3.2 System Design	44
Figure 3.3 Dataset Standardization and Integration Workflow	50
Figure 3.4 Data Preprocessing Pipeline	53
Figure 3.5 Proposed Model Architecture	63
Figure 3.6 Browser Extension Interface Wireframe (a) Control Interface, (b) Analysis Interface	76
Figure 3.7 Comment Analysis Option Interface	76
Figure 3.8 Illustration of Final Processing Results	77
Figure 4.1 Dataset A Structure Standardization Result	80
Figure 4.2 Dataset B Structure Standardization Result	81
Figure 4.3 Dataset C Structure Standardization Result	82
Figure 4.4 Dataset D Structure Standardization Result	83

Figure 4.5 Dataset E Structure Standardization Result.....	84
Figure 4.6 Dataset F Structure Standardization Result	85
Figure 4.7 Dataset G Structure Standardization Result	86
Figure 4.8 Results of the Indonesian Dataset Balancing Process	89
Figure 4.9 Final Dataset Integration Result	90
Figure 4.10 Case Folding Result.....	91
Figure 4.11 Text Cleaning Result	93
Figure 4.12 Tokenizing Result.....	93
Figure 4.13 Result of the Normalization Dictionary Integration	95
Figure 4.14 Normalization Result.....	96
Figure 4.15 Stopword Removal Result.....	97
Figure 4.16 Lemmatization Result.....	98
Figure 4.17 Dataset Composition After Balancing	99
Figure 4.18 Feature Extraction Output.....	102
Figure 4.19 Model Training Example.....	104
Figure 4.20 Accuracy and Loss Curves for the 20th Percentile.....	105
Figure 4.21 Confusion Matrix for the 20th Percentile	106
Figure 4.22 Classification Report for the 20th Percentile.....	106
Figure 4.23 Accuracy and Loss Curves for the 50th Percentile.....	107
Figure 4.24 Confusion Matrix for the 50th Percentile	107
Figure 4.25 Classification Report for the 50th Percentile.....	108
Figure 4.26 Accuracy and Loss Curves for the 90th Percentile.....	108
Figure 4.27 Confusion Matrix for the 90th Percentile	109
Figure 4.28 Classification Report for the 90th Percentile.....	109
Figure 4.29 Accuracy and Loss Curves for the 60:20:20 Split Ratio	111
Figure 4.30 Confusion Matrix for the 60:20:20 Split Ratio.....	111
Figure 4.31 Classification Report for the 60:20:20 Split Ratio	112
Figure 4.32 Accuracy and Loss Curves for the 70:20:10 Split Ratio	112
Figure 4.33 Confusion Matrix for the 70:20:10 Split Ratio.....	113
Figure 4.34 Classification Report for the 70:20:10 Split Ratio	113
Figure 4.35 Accuracy and Loss Curves for the 80:10:10 Split Ratio	114
Figure 4.36 Confusion Matrix for the 80:10:10 Split Ratio.....	115
Figure 4.37 Classification Report for the 80:10:10 Split Ratio	115
Figure 4.38 Accuracy and Loss Curves for Batch Size 16	117

Figure 4.39 Confusion Matrix for Batch Size 16.....	117
Figure 4.40 Classification Report for Batch Size 16.....	118
Figure 4.41 Accuracy and Loss Curves for Batch Size 32	118
Figure 4.42 Confusion Matrix for Batch Size 32.....	119
Figure 4.43 Classification Report for Batch Size 32.....	119
Figure 4.44 Accuracy and Loss Curves for Batch Size 64	120
Figure 4.45 Confusion Matrix for Batch Size 64.....	121
Figure 4.46 Classification Report for Batch Size 64.....	121
Figure 4.47 Accuracy and Loss Curves for Learning Rate 0.01	123
Figure 4.48 Confusion Matrix for Learning Rate 0.01	123
Figure 4.49 Classification Report for Learning Rate 0.01	124
Figure 4.50 Accuracy and Loss Curves for Learning Rate 0.001	124
Figure 4.51 Confusion Matrix for Learning Rate 0.001	125
Figure 4.52 Classification Report for Learning Rate 0.001	125
Figure 4.53 Accuracy and Loss Curves for Learning Rate 0.0001	126
Figure 4.54 Confusion Matrix for Learning Rate 0.0001	127
Figure 4.55 Classification Report for Learning Rate 0.0001	127
Figure 4.56 Accuracy and Loss Curves for LSTM Hidden Unit 32	129
Figure 4.57 Confusion Matrix for LSTM Hidden Unit 32	129
Figure 4.58 Classification Report for LSTM Hidden Unit 32	130
Figure 4.59 Accuracy and Loss Curves for LSTM Hidden Unit 64	131
Figure 4.60 Confusion Matrix for LSTM Hidden Unit 64	131
Figure 4.61 Classification Report for LSTM Hidden Unit 64	132
Figure 4.62 Accuracy and Loss Curves for LSTM Hidden Unit 128.....	132
Figure 4.63 Confusion Matrix for LSTM Hidden Unit 128	133
Figure 4.64 Classification Report for LSTM Hidden Unit 128	133
Figure 4.65 Accuracy and Loss Curves for Dropout Rate 0.2.....	135
Figure 4.66 Confusion Matrix for Dropout Rate 0.2	135
Figure 4.67 Classification Report for Dropout Rate 0.2	136
Figure 4.68 Accuracy and Loss Curves for Dropout Rate 0.3	137
Figure 4.69 Confusion Matrix for Dropout Rate 0.3	137
Figure 4.70 Classification Report for Dropout Rate 0.3	138
Figure 4.71 Accuracy and Loss Curves for Dropout Rate 0.4.....	138
Figure 4.72 Confusion Matrix for Dropout Rate 0.4	139

Figure 4.73 Classification Report for Dropout Rate 0.4	139
Figure 4.74 Accuracy and Loss Curves for Threshold 0.3	141
Figure 4.75 Confusion Matrix for Threshold 0.3	141
Figure 4.76 Classification Report for Threshold 0.3.....	142
Figure 4.77 Accuracy and Loss Curves for Threshold 0.5	143
Figure 4.78 Confusion Matrix for Threshold 0.5	143
Figure 4.79 Classification Report for Threshold 0.5.....	144
Figure 4.80 Accuracy and Loss Curves for Threshold 0.7	144
Figure 4.81 Confusion Matrix for Threshold 0.7.....	145
Figure 4.82 Classification Report for Threshold 0.7.....	145
Figure 4.83 LIME Implementation Results	150
Figure 4.84 LIME Interpretation Results for Normal Class Comments (a) First Sample, (b) Second Sample	151
Figure 4.85 LIME Interpretation Results for Bullying Class Comments (a) Third Sample, (b) Fourth Sample	151
Figure 4.86 LIME Interpretation Results for Misclassified Comments (a) Fifth Sample, (b) Sixth Sample.....	152
Figure 4.87 First Endpoint Response	154
Figure 4.88 Second Endpoint Response	157
Figure 4.89 Extension Control Interface Implementaion (a) Inactive Status, (b) Active Status	158
Figure 4.90 Additional Analysis Elements Implementation (a) Loading Explanation, (b) Explanation Unavailable	159
Figure 4.91 Analysis Page Implementation (a) Normal Class, (b) Bullying Class..	160
Figure 4.92 Filtering Feature Implementation Results 1 (a) Direct Link, (b) Profile Page	161
Figure 4.93 Filtering Feature Implementation Results 2 (a) Homepage, (b) Reels Page	162
Figure 4.94 Filter Feature Implementation	163

LIST OF TABLES

Table 2.1 Example of Sequence Length Adjustment.....	29
Table 3.1 Research Dataset Sources	45
Table 3.2 Sample Data from Dataset A	46
Table 3.3 Sample Data from Dataset B.....	46
Table 3.4 Sample Data from Dataset C.....	47
Table 3.5 Sample Data from Dataset D	47
Table 3.6 Sample Data from Dataset E.....	48
Table 3.7 Sample Data from Dataset F	48
Table 3.8 Sample Data from Dataset G	49
Table 3.9 Research Dataset After Processing	51
Table 3.10 Balanced Research Dataset	52
Table 3.11 Final Dataset Composition.....	52
Table 3.12 Example Sentences	53
Table 3.13 Case Folding Results.....	54
Table 3.14 Text Cleaning Steps	54
Table 3.15 Text Cleaning Results	55
Table 3.16 Tokenizing Results.....	56
Table 3.17 Sources of the Normalization Dictionary.....	57
Table 3.18 Normalization Results.....	57
Table 3.19 Stopword Removal Results.....	58
Table 3.20 Lemmatization Results.....	59
Table 3.21 Example of Initial FastText Word Vectors	61
Table 3.22 Example of Word Embedding Representations	62
Table 3.23 Example of Spatial Dropout 1D Output.....	64
Table 3.24 Assumed Weight and Bias Values	65
Table 3.25 Forward Pass Calculation Results.....	66
Table 3.26 Backward Pass Calculation Results	67
Table 3.27 Assumed Prediction Results.....	68
Table 3.28 Model Testing Hyperparameters.....	70
Table 3.29 Assumed Perturbed Samples and Their Prediction Probabilities	71
Table 3.30 Distance and Local Weight Calculation Results.....	72
Table 4.1 Dataset File Name Details.....	79

Table 4.2 Dataset Composition After Preprocessing	98
Table 4.3 Sequence Length Evaluation Results.....	110
Table 4.4 Data Split Ratio Evaluation Results.....	116
Table 4.5 Batch Size Evaluation Results	122
Table 4.6 Learning Rate Evaluation Results.....	128
Table 4.7 LSTM Hidden Unit Evaluation Results	134
Table 4.8 Dropout Rate Evaluation Results.....	140
Table 4.9 Threshold Evaluation Results	146
Table 4.10 Model Evaluation Results	146
Table 4.11 Test Data Samples for LIME Interpretation	150