

# TABLE OF CONTENTS

<b>APPROVAL PAGE .....</b>	<b>ii</b>
<b>PERNYATAAN BEBAS PLAGIASI .....</b>	<b>iii</b>
<b>FOREWORD.....</b>	<b>iv</b>
<b>TABLE OF CONTENTS .....</b>	<b>v</b>
<b>LIST OF FIGURES .....</b>	<b>viii</b>
<b>LIST OF TABLES .....</b>	<b>x</b>
<b>INTISARI .....</b>	<b>xi</b>
<b>ABSTRACT.....</b>	<b>xii</b>
<b>CHAPTER 1 INTRODUCTION.....</b>	<b>1</b>
1.1 Research Background .....	1
1.2. Research Problem .....	2
1.3. Research Objective .....	3
1.4. Research Scopes.....	3
1.5. Research Benefits.....	4
<b>CHAPTER 2 LITERATURE REVIEW.....</b>	<b>5</b>
<b>CHAPTER 3 THEORETICAL BASIS.....</b>	<b>12</b>
3.1 Authorship Identification .....	12
3.2 Recurrent Neural Network (RNN).....	13
3.3 Long Short-Term Memory (LSTM) .....	14
3.4 Gated Recurrent Unit (GRU) .....	17
3.5 Support Vector Machine (SVM).....	18
3.6 Multinomial Logistic Regression.....	18
3.7 Term Frequency-Inverse Document Frequency .....	19
3.8 Global Vectors for Word Representation (GloVe) .....	20
3.9 Confusion Matrix .....	21
3.10 Accuracy .....	22



3.11 Precision and Recall.....	23
3.12 F1 Score .....	23
<b>CHAPTER 4 RESEARCH METHODOLOGY.....</b>	<b>24</b>
4.1 Research Description .....	24
4.2 Facilities .....	26
4.3 Research Method .....	26
4.3.1 Dataset Description .....	26
4.3.2 Loading, Cleaning, and Filtering .....	29
4.3.3 Text Preprocessing.....	29
4.3.4 Global Vectors for Word Representation (GloVe) .....	29
4.3.5 Vectorization.....	30
4.3.6 GloVe Embedding vs Custom-trained Embeddings .....	34
4.3.7 Article Level vs. Sentence Level .....	35
4.3.8 Gated Recurrent Unit (GRU) Model.....	37
4.3.9 Long Short-Term Memory (LSTM) Model .....	39
4.3.10 Machine Learning Models as Baseline .....	41
4.3.11 Model Evaluation.....	41
<b>CHAPTER 5 IMPLEMENTATION.....</b>	<b>42</b>
5.1 Data Collection and Loading .....	42
5.2 Data Cleaning and Filtering .....	43
5.3 Tokenizing Each Article into Sentences .....	44
5.4 Text Preprocessing.....	46
5.5 Vectorization, Label Encoding, and Data Splitting .....	48
5.5.1 Global Vectors for Word Representation (GloVe) .....	49
5.5.2 Term Frequency-Inverse Document Frequency (TF-IDF) .....	51
5.5.3 Label Encoding and Data Splitting .....	53
5.6 Long Short-Term Memory (LSTM) Model .....	55



5.7 Gated Recurrent Unit (GRU) Model.....	57
5.8 Machine Learning Models .....	59
5.9 Model Evaluation and Visualization.....	61
<b>CHAPTER 6 RESULTS.....</b>	<b>64</b>
6.1 Long Short-Term Memory (LSTM) Model Evaluation.....	64
6.1.1 LSTM Model Accuracy .....	64
6.1.2 LSTM Model Precision, Recall, and F1 Score .....	66
6.1.3 LSTM Model Confusion Matrix .....	70
6.2 Gated Recurrent Unit (GRU) Model Evaluation .....	72
6.2.1 GRU Model Accuracy .....	73
6.2.2 GRU Model Precision, Recall, and F1 Score .....	75
6.2.3 GRU Model Confusion Matrix .....	78
6.3 Support Vector Machine (SVM) Model Evaluation.....	81
6.3.1 SVM Model Precision, Recall, and F1 Score .....	81
6.3.2 SVM Model Confusion Matrix.....	85
6.4 Logistic Regression Model Evaluation.....	87
6.4.1 Logistic Regression Model Precision, Recall, and F1 Score .....	88
6.4.2 Logistic Regression Model Confusion Matrix.....	91
6.5 Model Comparison.....	94
6.6 The Need for Sentence-Level Training.....	97
<b>CHAPTER 7 CONCLUSION.....</b>	<b>100</b>
<b>BIBLIOGRAPHY .....</b>	<b>102</b>