

## TABLE OF CONTENTS

<b>APPROVAL PAGE .....</b>	<b>ii</b>
<b>STATEMENT.....</b>	<b>iii</b>
<b>PREFACE.....</b>	<b>iv</b>
<b>TABLE OF CONTENTS.....</b>	<b>v</b>
<b>LIST OF TABLES .....</b>	<b>ix</b>
<b>LIST OF FIGURES .....</b>	<b>x</b>
<b>ABSTRACT.....</b>	<b>xii</b>
<b>CHAPTER I INTRODUCTION.....</b>	<b>1</b>
<b>1.1 Research Background.....</b>	<b>1</b>
<b>1.2 Research Problem .....</b>	<b>3</b>
<b>1.3 Research Scope .....</b>	<b>3</b>
<b>1.4 Research Objective .....</b>	<b>4</b>
<b>1.5 Research Benefit.....</b>	<b>4</b>
<b>1.6 Research Writing System .....</b>	<b>4</b>
<b>CHAPTER II LITERATURE REVIEW .....</b>	<b>6</b>
<b>CHAPTER III THEORETICAL BASIS .....</b>	<b>13</b>
<b>3.1 Human-Generated and Machine-Generated Text .....</b>	<b>13</b>
<b>3.2 Stylometric Features .....</b>	<b>13</b>
3.2.1 Lexical Features .....	15
3.2.2 Syntactic Features .....	15
3.2.3 Character and N-gram Features .....	16
3.2.4 Limitation of Stylometry Analysis.....	16

<b>3.3</b>	<b>Sentiment Polarity</b> .....	<b>17</b>
<b>3.4</b>	<b>Readability Score</b> .....	<b>17</b>
<b>3.5</b>	<b>Machine Learning Model</b> .....	<b>18</b>
3.5.1	SVM .....	19
3.5.2	Logistic Regression .....	20
3.5.3	Random Forest .....	22
<b>3.6</b>	<b>Model Evaluation</b> .....	<b>23</b>
3.6.1	Accuracy .....	23
3.6.2	Precision .....	23
3.6.3	Recall .....	24
3.6.4	F-1 Score .....	24
<b>CHAPTER IV RESEARCH METHODOLOGY</b> .....		<b>26</b>
<b>4.1</b>	<b>Research Description</b> .....	<b>26</b>
<b>4.2</b>	<b>Dataset</b> .....	<b>27</b>
<b>4.3</b>	<b>Stylometric Features</b> .....	<b>29</b>
<b>4.4</b>	<b>Content-Based Features</b> .....	<b>30</b>
<b>4.5</b>	<b>Model Training and Testing Plan</b> .....	<b>30</b>
<b>4.6</b>	<b>Model Evaluation and Visualization</b> .....	<b>31</b>
<b>4.7</b>	<b>Feature Importance</b> .....	<b>31</b>
<b>CHAPTER V IMPLEMENTATION</b> .....		<b>33</b>
<b>5.1</b>	<b>Library Installation</b> .....	<b>33</b>
<b>5.2</b>	<b>Stylometric Feature Extraction</b> .....	<b>34</b>
5.2.1	Word Count.....	34
5.2.2	Unique Word Count .....	35
5.2.3	Character Count .....	35
5.2.4	Average Word Length.....	36

5.2.5	Type Token Ratio.....	37
5.2.6	Hapax Legomenon Rate.....	37
5.2.7	Word Unigram.....	39
5.2.8	Word Bigram.....	41
5.2.9	Word Trigram.....	43
<b>5.3</b>	<b>Content Based Features Extraction.....</b>	<b>45</b>
5.3.1	TF-IDF with Stopword Removal.....	45
5.3.2	Sentiment Polarity Score.....	48
5.3.3	Readability Score.....	48
<b>5.4</b>	<b>Feature Engineering.....</b>	<b>48</b>
<b>5.5</b>	<b>Model Training and Evaluation.....</b>	<b>49</b>
5.5.1	Data Splitting.....	49
5.5.2	Support Vector Machine (SVM) Model.....	49
5.5.3	Logistic Regression Model.....	51
5.5.4	Random Forest Model.....	52
5.5.5	Metrics Calculation.....	54
<b>5.6</b>	<b>Feature Importance.....</b>	<b>54</b>
<b>5.7</b>	<b>Refit Model.....</b>	<b>55</b>
<b>CHAPTER VI RESULT AND DISCUSSION.....</b>		<b>56</b>
<b>6.1</b>	<b>Classification Model Performances.....</b>	<b>56</b>
<b>6.2</b>	<b>Stylometric Features Result.....</b>	<b>57</b>
<b>6.3</b>	<b>Content-Based Features Result.....</b>	<b>64</b>
<b>6.4</b>	<b>Feature Importance and Analysis.....</b>	<b>66</b>
<b>CHAPTER VII CONCLUSION.....</b>		<b>72</b>
<b>7.1</b>	<b>Conclusion.....</b>	<b>72</b>
<b>7.2</b>	<b>Future Work.....</b>	<b>73</b>



UNIVERSITAS  
GADJAH MADA

**IDENTIFICATION OF MACHINE-GENERATED TEXT USING STYLOMETRIC AND CONTENT-BASED FEATURES**

PUTU ANNAND KURNIA ISWARI, Yunita Sari, S.Kom., M.Sc., Ph.D

Universitas Gadjah Mada, 2024 | Diunduh dari <http://etd.repository.ugm.ac.id/>

**BIBLIOGRAPHY ..... X**