

TABLE OF CONTENTS

APPROVAL PAGE	iv
TABLE OF CONTENTS	viii
LIST OF TABLES	xi
LIST OF FIGURES	xiii
LIST OF LISTINGS	xiv
ABSTRACT	xv
CHAPTER I INTRODUCTION	1
1. 1 Research Background	1
1. 2 Research Problem	3
1. 3 Research Scope	3
1. 4 Research Objective	4
1. 5 Research Benefits	4
1. 6 Research Methodology	4
CHAPTER II LITERATURE REVIEW	7
CHAPTER III THEORETICAL BASIS	14
3.1 Readability	14
3.2 Text Classification	14
3.3 Buku Sekolah Elektronik Kemendikbud	14
3.4 Flesch Kincaid	15
3.4.1 Flesch Reading Ease (1948)	15
3.4.2 Flesch Kincaid Grade Levels (1975)	16
3.5 Data Preprocessing	17
3.5.1 Case Folding	17
3.5.2 Tokenization	17
3.5.3 Syllabelizer	17
3.5.4 Stop Word Removal	17
3.5.5 Stemming	17
3.5.6 Sastrawi	18
3.6 Feature Extraction	18
3.6.1 TF-IDF	18
3.7 Machine Learning	19
3.7.1 Supervised Learning	20
3.7.2 Naïve Bayes	20

3.7.3 Multinomial Naïve Bayes	21
3.7.4 Gaussian Naïve Bayes	23
3.8 Gold-Standard Data set	24
3.9 K-Fold Cross Validation	24
3.10 Performance Measurements	24
3.10.2 Precision	26
3.10.3 Recall	26
3.10.4 F1-Score	26
3.11 Hyperparameter Tuning	26
CHAPTER IV ANALYSIS AND SYSTEM DESIGN	27
4.1 Research Description	27
4.2 Tools and Materials	27
4.2.1 Tools	27
4.2.2 Materials	27
4.3 Research Phase	29
4.4 Dataset Collection	32
4.5 Preprocessing	32
4.6 Feature Extraction	34
4.6.1 TF-IDF	34
4.7 Data Labeling	39
4.7.1 Data Labeling Based on Flesch Kincaid Table (1 st Model)	39
4.7.2 Data Labeling Based on Average Score (2 nd Model)	39
4.8 Naïve Bayes Classifier	40
4.9 Evaluation Design	41
CHAPTER V IMPLEMENTATION	43
5.1 Dataset Collection	43
5.2 Data Preprocessed	44
5.2.1 Data Cleaning	44
5.2.2 Stop Word Removal	45
5.2.3 Stemming	46
5.2.4 Syllabelizer	46
5.3 Feature Extraction	46
5.3.1 TF-IDF	46
5.4 Implementation of Flesch Kincaid Formula	47

5.4.1 Flesch Kincaid Reading Ease	47
5.4.2 Flesch Kincaid Grade Level	48
5.5 Data Labeling	48
5.5.1 Data Labeling Based on Flesch Kincaid Table (1st Model)	48
5.5.2 Data Labeling Based on Average Score (2 nd Model)	48
5.6 Model Implementation	49
5.6.1 Gaussian Naïve Bayes (1 st Model)	50
5.6.2 Gaussian Naïve Bayes (2 nd Model)	50
5.6.3 Multinomial Naive Bayes (1 st Model)	50
5.6.4 Multinomial Naïve Bayes (2 nd Model)	50
5.7 Training and Testing	51
5.7.1 K-Fold Cross Validation	52
5.8 Hyperparameter Tuning	53
CHAPTER VI RESULT AND DISCUSSION	54
6.1 Result	54
6.1.1 Result from Data Pre-processed	54
6.1.2 Result Score from Flesch Kincaid Formula	55
6.1.3 Result from Data Labeling	56
6.1.4 Result from TF-IDF	58
6.1.5 Result from Confusion Matrix	59
6.1.6 Result from Machine Learning	60
6.1.7 Result from Hyperparameter Tuning	63
6.2 Model Comparison	66
6.3 Discussion	69
CHAPTER VII CONCLUSION	73
7.1 Conclusion	73
7.2 Future work	74
REFERENCES	75