

CONTENTS

Title Page	ii
Approval Page	iii
Declaration Page	iv
PREFACE	v
ABSTRACT	xii
I INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement	3
1.3 Research Scope	4
1.4 Research Goal	4
1.5 Research Benefits	5
II LITERATURE REVIEW	6
2.1 Sentiment Analysis	6
2.2 Topic Modeling	12
2.3 User Retention	17
2.4 Framework	18
III THEORETICAL FOUNDATIONS	19
3.1 Sentiment Lexicons	19
3.2 Natural Language Processing	19
3.3 Aspect-Based Sentiment Analysis	20
3.4 Bidirectional Encoder Representations from Transformers	21
3.5 Latent Dirichlet Allocation	22
3.6 Web Scraping	25
3.7 Pandas DataFrame	25
3.8 Part-of-Speech Tagging	26

3.9	Bag of N-Grams	26
3.10	Named Entity Recognition	28
3.11	Evaluation Metrics	28
IV	RESEARCH METHODOLOGY	31
4.1	Problem Description	31
4.2	Research Steps	31
4.3	Data Collection	33
4.4	Data Preprocessing	33
4.5	Sentiment Analysis	34
4.6	Topic Modeling	37
4.7	Evaluation	39
4.8	User Retention Calculation	40
V	IMPLEMENTATION	41
5.1	Data Preparation	41
5.2	Aspect-Based Sentiment Analysis	45
5.3	Part-of-Speech Tagging	48
5.4	Bag-of-N-Grams	50
5.5	Topic Modeling	51
VI	RESULT AND DISCUSSION	55
6.1	Interpretation of Findings	55
6.1.1	Data Preprocessing	56
6.1.2	Results of Named Entity Recognition	57
6.1.3	Aspect Based Sentiment Analysis	59
6.1.4	Part-of-Speech Tagging	61
6.1.5	Bag-of-N-grams	63
6.1.6	Latent Dirichlet Allocation	65
6.2	Method Evaluation	72
6.2.1	ABSA Evaluation	72
6.2.2	LDA Evaluation	73
6.3	Bukalapak User Retention	74



VI	CONCLUSION	77
7.1	Summary of Key Findings	77
7.2	Limitations	78
7.3	Suggestion for Future Work	79

LIST OF TABLES

2.1	Literature Review of Sentiment Analysis	8
2.2	Literature Review of Topic Modeling	14
3.1	Example of a Pandas DataFrame	26
3.2	Example of Bag of N-Grams	27
5.1	Tokenizer Results	47
6.1	Sample of Bukalapak Data	55
6.2	Number of reviews per version	56
6.3	NER Results	58
6.4	ABSA Sentiment Prediction	59
6.5	Sentiment Distribution Amount	61
6.6	Sample of POS Tagged Review	62
6.7	Sample of BoNG	63
6.8	Top 5 Most Frequent Words	64
6.9	Topic and 10 Most Keywords Associated (NumTop = 5)	66
6.10	Evaluation of ABSA	73
6.11	Evaluation of LDA	74

LIST OF FIGURES

3.1	Example of ABSA (Chiusano, 2022)	20
3.2	BERT Process	21
3.3	NER Process	28
4.1	Research Flow	32
4.2	Research Preprocessing	34
4.3	BERT Process	35
4.4	LDA Process	38
5.1	Install Library	42
5.2	Cleaning Data	42
5.3	Labeling Data using Label Studio	43
5.4	NER Process	44
5.5	Aspect-Based Sentiment Analysis	46
5.6	POS Tagging	49
5.7	Implementation of Bag-of-N grams	50
5.8	Topic Modeling	51
5.9	Topic Distribution	53
6.1	Sentiment Distribution Chart	60
6.2	Topic Distribution	67
6.3	Topic Distribution per Version	69
6.4	Topic Distribution per 4 Month	71
6.5	Bukalapak User Retention	75