UNIVERSITAS
GADJAH MADA
Comparative Analysis of Sentiment Analysis Using Naive Bayes with TF-IDF, Bag of Word, and Word2Vec
(Case Study: Lazada Electronic Product Reviews)
MUHAMMAD DZAKI P, Anny Kartika Sari, S.Si., M.Sc., Ph.D
Universitas Gadjah Mada, 2021 | Diunduh dari http://etd.repository.ugm.ac.id/

TABLE OF CONTENTS

APPROVAL PAGE ii
TABLE OF CONTENTS iii
TABLE OF FIGURES vi
TABLE OF TABLES viii
TABLE OF EQUATIONS ix
ABSTRACT x
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 Advantage 4
1. 6 Research Methodology 4
1. 6.1 Literature Review 4
1. 6.2 Gathering Requirements 4
1. 6.3 Implementation 5
1. 6.4 Testing 5
1. 6.5 Evaluation 5
1. 6.6 Thesis Organization 6
CHAPTER II: LITERATURE REVIEW 7
CHAPTER III: THEORETICAL BASIS 13
3. 1 Data Mining 13
3. 2 Sentiment Analysis 14
3. 3 Preprocessing 15
3. 3. 1 Data Cleaning 16
3. 3. 2 Tokenization 16

UNIVERSITAS
GADJAH MADA
Comparative Analysis of Sentiment Analysis Using Naive Bayes with TF-IDF, Bag of Word, and Word2Vec
(Case Study: Lazada Electronic Product Reviews)
MUHAMMAD DZAKI P, Anny Kartika Sari, S.Si., M.Sc., Ph.D
Universitas Gadjah Mada, 2021 | Diunduh dari http://etd.repository.ugm.ac.id/

3. 3. 3 Stop Word Removal 16
3. 3. 4 Stemming 17
3. 4 Feature Extraction 18
3. 4. 1 TF-IDF 19
3. 4. 2 Bag of Words 19
3. 4. 3 Word2Vec 20
3. 5 Classification 21
3. 5. 1 Naive Bayes Classification 21
3. 5. 2 Confusion Matrix 25
CHAPTER IV: ANALYSIS AND DESIGN 26
4. 1 Research description 26
4. 2 Research Workflow 26
4. 3 Data Acquisition 28
4. 4 Dataset Analysis 29
4. 5 Preprocessing 29
4. 5. 1 Tokenization 30
4. 5. 2 Stop Word Removal 31
4. 5. 3 Stemming 31
4. 6 Feature Extraction 34
4. 6. 1 TF-IDF 34
4. 6. 2 Word of Bags 35
4. 6. 3 Word2Vec 36
4. 7 Naive Bayes 36
4. 8 Implementation 40
4. 9 Evaluation 40
CHAPTER V: IMPLEMENTATION 43
5. 1 Implementation of Data Acquisition 43
5. 2 Implementation of Preprocessing 43
5. 2. 1 Case Folding 43
5. 2. 2 Tokenization 44
5. 2. 3 Stop Word Removal 45

UNIVERSITAS
GADJAH MADA
Comparative Analysis of Sentiment Analysis Using Naive Bayes with TF-IDF, Bag of Word, and Word2Vec
(Case Study: Lazada Electronic Product Reviews)
MUHAMMAD DZAKI P, Anny Kartika Sari, S.Si., M.Sc., Ph.D
Universitas Gadjah Mada, 2021 | Diunduh dari http://etd.repository.ugm.ac.id/

5. 2. 4 Stemming 46
5. 3 Implementation of Feature Extractions 47
5. 3. 1 TF-IDF on Naïve Bayes 47
5. 3. 2 Bag of Word on Naïve Bayes 48
5. 3. 3 Word2Vec on Naïve Bayes 49
5. 4 Evaluation 53
CHAPTER VI: RESULTS AND DISCUSSION 54
6. 1 Result of Preprocessing 54
6. 2 Evaluation Result 55
6. 2.1 TF-IDF on Naïve Bayes 55
6. 2.2 Bag of Words on Naïve Bayes 56
6. 2.3 Word2Vec on Naïve Bayes 56
6. 3 Feature Extraction Comparison 57
6. 4 Previous Work Comparison 58
CHAPTER VII: CONCLUSIONS AND FUTURE WORK 59
7. 1 Conclusions 59
7. 2 Future Work 59
REFERENCES 60