

## TABLE OF CONTENTS

|                                                                     |             |
|---------------------------------------------------------------------|-------------|
| <b>PREFACE.....</b>                                                 | <b>ii</b>   |
| <b>TABLE OF CONTENTS.....</b>                                       | <b>iv</b>   |
| <b>LIST OF TABLES.....</b>                                          | <b>vi</b>   |
| <b>LIST OF IMAGES.....</b>                                          | <b>vii</b>  |
| <b>ABSTRACT.....</b>                                                | <b>viii</b> |
| <b>CHAPTER 1</b>                                                    |             |
| <b>INTRODUCTION.....</b>                                            | <b>1</b>    |
| 1.1. Background.....                                                | 1           |
| 1.2. Problem Statement.....                                         | 4           |
| 1.3. Scope of Research.....                                         | 4           |
| 1.4. Goal of Research.....                                          | 5           |
| <b>CHAPTER 2</b>                                                    |             |
| <b>LITERATURE REVIEW.....</b>                                       | <b>6</b>    |
| 2.1. Sentiment Analysis.....                                        | 6           |
| 2.2. Topic Modeling.....                                            | 7           |
| 2.3. Mixed-Method Approaches.....                                   | 8           |
| <b>CHAPTER 3</b>                                                    |             |
| <b>THEORETICAL FOUNDATION.....</b>                                  | <b>14</b>   |
| 3.1. Sentiment Lexicons.....                                        | 14          |
| 3.1.1. VADER (Valence Aware Dictionary and sEntiment Reasoner)..... | 14          |
| 3.2. Matrix Factorization.....                                      | 16          |
| 3.2.1 Non-negative Matrix Factorization (NMF).....                  | 16          |
| 3.3. Web Scraping.....                                              | 19          |
| 3.4. Bigram and Trigram Analysis.....                               | 19          |
| 3.5. Likert Scale.....                                              | 20          |
| 3.6. Intra-Class Correlation (ICC).....                             | 21          |
| 3.7. Cronbach's alpha.....                                          | 21          |
| 3.8. Pointwise Mutual Information (PMI).....                        | 22          |
| 3.9. Natural Language Toolkit (NLTK).....                           | 23          |
| 3.10. Pandas DataFrame.....                                         | 23          |
| 3.11. TfidfVectorizer.....                                          | 24          |
| <b>CHAPTER 4</b>                                                    |             |
| <b>RESEARCH METHODOLOGY.....</b>                                    | <b>26</b>   |
| 4.1. Problem Description.....                                       | 26          |
| 4.2. Research Design.....                                           | 26          |
| 4.3. Research Steps.....                                            | 27          |

|                                              |           |
|----------------------------------------------|-----------|
| 4.4. Data Collection.....                    | 28        |
| 4.5. Data Preprocessing.....                 | 29        |
| 4.6. Sentiment Analysis.....                 | 30        |
| 4.7. Topic Modeling.....                     | 31        |
| 4.8. Bigram and Trigram Analysis.....        | 32        |
| 4.9. Evaluation.....                         | 33        |
| <b>CHAPTER 5</b>                             |           |
| <b>IMPLEMENTATION.....</b>                   | <b>35</b> |
| 5.1. Data Preparation.....                   | 35        |
| 5.2. Sentiment Analysis.....                 | 38        |
| 5.3. Topic Modeling.....                     | 39        |
| 5.4. Bigram and Trigram Analysis.....        | 41        |
| <b>CHAPTER 6</b>                             |           |
| <b>RESULT AND DISCUSSION.....</b>            | <b>45</b> |
| 6.1. Interpretation of Findings.....         | 45        |
| 6.1.1 Data Preprocessing.....                | 46        |
| 6.1.2. Result of Sentiment Analysis.....     | 46        |
| 6.1.3. Result of Topic Modeling.....         | 49        |
| 6.1.4. Result of Bigram and Trigrams.....    | 53        |
| 6.1.5. Result of External Reviewer.....      | 59        |
| 6.2. Method Evaluation.....                  | 61        |
| 6.2.1. Evaluation of VADER.....              | 61        |
| 6.2.2. Evaluation of NMF.....                | 63        |
| 6.2.3. Evaluation of External Reviewers..... | 64        |
| <b>CHAPTER 7</b>                             |           |
| <b>CONCLUSION.....</b>                       | <b>66</b> |
| 7.1. Summary of Key Findings.....            | 66        |
| 7.2. Suggestions for Future Work.....        | 66        |
| <b>REFERENCES.....</b>                       | <b>67</b> |
| <b>APPENDIX A</b>                            |           |
| <b>EXTERNAL REVIEWERS SURVEY.....</b>        | <b>71</b> |

## LIST OF TABLES

|                                                                                            |    |
|--------------------------------------------------------------------------------------------|----|
| Table 2.1. Table of Literature Review Summary.....                                         | 9  |
| Table 6.1. Sample Data Obtained from Web Scraper.....                                      | 45 |
| Table 6.2. Frequency of Reviews per Version.....                                           | 46 |
| Table 6.3. Amount of Negative Review per Language.....                                     | 48 |
| Table 6.4. Topic and 10 Most Keywords Associated (NumTop = 10).....                        | 49 |
| Table 6.5. Topic and 10 Most Keywords Associated (NumTop = 5).....                         | 50 |
| Table 6.6. Topic and 10 Most Keywords Associated (NumTop = 7).....                         | 51 |
| Table 6.7. Topic and 10 Most Keywords Associated (NumTop = 6).....                         | 52 |
| Table 6.8. Most Frequent Bigram and Trigram in Topic 5.....                                | 54 |
| Table 6.9. Highest PMI Scoring Bigram and Trigram in Topic 5.....                          | 55 |
| Table 6.10. Highest PMI Scoring Bigram and Trigram in Topic 5 Without Duplicate Words..... | 57 |
| Table 6.11. Expert Approval.....                                                           | 60 |
| Table 6.12. PMI Score for Each Topic in the Corpus.....                                    | 63 |
| Table 6.13. ICC and Crobach's Alpha Values per Category.....                               | 64 |
| Table A.1. Likert Scale Scores per Sample per Category.....                                | 71 |

## LIST OF IMAGES

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| Fig 3.1. Example of VADER Usage (Geeksforgeeks, 2021).....                           | 15 |
| Fig 3.2. Diagram of NMF (Saxena and Mueller, 2018).....                              | 18 |
| Fig 3.3. Example of Unigram, Bigram, and Trigram Representation (Yngie, 2020).<br>20 |    |
| Fig 3.4. Example of Pandas DataFrame (Pandas, 2023).....                             | 24 |
| Fig 5.1. Used Libraries and Modules.....                                             | 35 |
| Fig 5.2. Web Scraping Process.....                                                   | 36 |
| Fig 5.3. Data Cleaning Process.....                                                  | 37 |
| Fig 5.4. VADER Sentiment Analysis Process.....                                       | 38 |
| Fig 5.5. Review Concatenation Process.....                                           | 39 |
| Fig 5.6. Additional Stopwords Initialization.....                                    | 39 |
| Fig 5.7. NMF Topic Modeling Process.....                                             | 39 |
| Fig 5.8. Topic Designation Process.....                                              | 40 |
| Fig 5.9. Grouping Documents According to Topics.....                                 | 41 |
| Fig 5.10. Tokenizing Process.....                                                    | 42 |
| Fig 5.11. Stopword and Duplicate Word Filtering.....                                 | 42 |
| Fig 5.12. Bigram and Trigram Function Definition.....                                | 42 |
| Fig 5.13. Bigram and Trigram Analysis Based on PMI.....                              | 43 |
| Fig 5.14. Bigram and Trigram Analysis Based on Raw Frequency.....                    | 43 |
| Fig 6.1. English Review Sentiment Distribution Chart.....                            | 47 |
| Fig 6.2. Indonesian Review Sentiment Distribution Chart.....                         | 47 |
| Fig 6.3. Topic Frequency Chart.....                                                  | 53 |
| Fig. 6.4. Sample Pain Points Given to External Reviewer.....                         | 59 |