Multi-Aspect Sentiment Analysis of Indonesian Hotel Reviews Using Hybrid Classifier Based on SVM, NB, RF, and K-NN

FINAL THESIS

FAZRAH RAHMAWATI MEWAR
1212001002

INFORMATICS STUDY PROGRAM
FACULTY OF ENGINEERING AND COMPUTER SCIENCE
BAKRIE UNIVERSITY
JAKARTA
2025

Statement of Originality

This Final Project is my own work, and all sources, whether quoted or referenced, have been properly acknowledged and stated truthfully.

Name : Fazrah Rahmawati Mewar

Student ID : 1212001002

Signature : 

Tanggal : 2 September 2025

Statement of Approval

This final thesis is prepared and submitted by :

Name	: Fazrah Rahmawati Mewar
NIM	: 1212001002
Departement	: Informatika
Faculty	: Engineering and Computer Science
Title	: Multi-Aspect Sentiment Analysis of Indonesian Hotel Reviews Using Hybrid Classifier Based on SVM, NB, RF, and K-NN with TF-IDF Representation

has been approved by the Board of Examiners and accepted as partial fulfillment of the requirements to obtain a Bachelor degree in Informatics Departement, Faculty of Engineering and Computer Science, Bakrie University.

Jakarta, 2 September 2025

Primary Supervisor	Irwan Prasetya Gunawan, S.T., M.Eng., Ph.D	(.....20250903)
Co-Supervisor	Guson P. Kuntarto, S.T., M.Sc.	(.....03/09/2025)
Examiner 1	Dewi Fatmawati Surianto, S.Kom., M.Kom.	(.....3/9/2025)
Examiner 2	Berkah Iman Santoso, S.T., M.T.I	(.....) 03/09/2025 16:38 Appr.TA for FRM

Acknowledgements

All praise and gratitude be to God Almighty for His grace and blessings, which have enabled the author to complete this thesis entitled, ”Multi-Aspect Sentiment Analysis of Indonesian Hotel Reviews Using Hybrid Classifier Based on SVM, NB, RF, and K-NN with TF-IDF Representation”. This thesis is submitted as a partial fulfillment of the requirements for the degree of Bachelor of Computer Science in the Informatics Study Program, Faculty of Engineering and Computer Science, Bakrie University. In the process of writing this thesis, the author received a great deal of help, guidance, and support from various parties. Therefore, the author would like to express sincere gratitude to:

1. Prof. Ir. Sofia W. Alisjahbana, M.Sc., Ph.D., IPU., as the Rector of Bakrie University.
2. Ir. Esa Haruman Wiraatmadja, M.Sc., Ph.D. as the Dean of the Faculty of Engineering and Computer Science at Bakrie University.
3. Dr. Iwan Adhicandra as the Head of the Informatics Study Program at Bakrie University.
4. Dr. Irwan Prasetya Gunawan and Mr. Guson P. Kuntarto, S.T., M.Sc., as the author’s supervisors, for the guidance, direction, and support provided throughout the research and writing process of this thesis.
5. Parents and family, for their endless prayers, moral support, and encouragement that have always strengthened the author in completing this study.
6. Friends who provided assistance, motivation, and technical and moral support during the completion of this thesis.

Finally, the author hopes that God Almighty will repay the kindness of all of those who have helped. May this thesis be beneficial for the advancement of knowledge.

Jakarta, 2 September 2025

Fazrah Rahmawati Mewar

Statement of Publication Approval

As a member of the academic community of Bakrie University, I, the undersigned below:

Name	: Fazrah Rahmawati Mewar
NIM	: 1212001002
Department	: Informatika
Faculty	: Engineering and Computer Science
Title	: Multi-Aspect Sentiment Analysis of Indonesian Hotel Reviews Using Hybrid Classifier Based on SVM, NB, RF, and K-NN

For the development of science, I agree to grant Bakrie University the **Non-exclusive Royalty-Free Right** over my scientific work entitled:

Multi-Aspect Sentiment Analysis of Indonesian Hotel Reviews Using Hybrid Classifier Based on SVM, NB, RF, and K-NN

along with any accompanying devices (if necessary). With this Non-exclusive Royalty-Free Right, Bakrie University is entitled to store, transfer/convert, manage in the form of a *database*, maintain, and publish my thesis as long as my name is still listed as the author/creator and as the copyright holder for academic purposes.

This statement is hereby made truthfully.

Jakarta, 2 September 2025

The Declarant,

(Fazrah Rahmawati Mewar)

Universitas Bakrie

Multi-Aspect Sentiment Analysis of Indonesian Hotel Reviews Using Hybrid Classifier Based on SVM, NB, RF, and K-NN

Fazrah Rahmawati Mewar
mewarfazrahrahmawati@gmail.com

Abstract

Multi-aspect sentiment analysis is a crucial task for understanding detailed user opinions on various facets of a product or service. This study aims to develop and evaluate a robust multi-aspect sentiment classification model for Indonesian hotel reviews. Four individual machine learning algorithms, namely *Support Vector Machine* (SVM), *Naive Bayes* (NB), *K-Nearest Neighbors* (K-NN), and *Random Forest* (RF), are implemented and compared. The models are trained using three different feature representation techniques: *Bag-of-Words* (BoW), *Term Frequency-Inverse Document Frequency* (TF-IDF), and *Word2Vec*. Furthermore, a *Hybrid Classifier* using a stacking methodology is proposed to combine the strengths of the individual models. The experiments are conducted on the HoASA dataset from the IndoNLU benchmark. The experimental results demonstrate that the proposed hybrid stacking model achieves a peak accuracy of 93.40%, which was obtained when using Bag-of-Words (BoW) and Term Frequency-Inverse Document Frequency (TF-IDF) feature representations. This figure surpasses the performance of the best individual classifier, which was SVM with TF-IDF features, recording an accuracy of 93.10%. Interestingly, in the tests using Word2Vec features, the Random Forest model showed slightly superior performance with an accuracy of 86.10%. The conclusion of this study highlights the effectiveness of the hybrid approach, particularly when paired with classic feature representations like BoW and TF-IDF, in improving the accuracy of multi-aspect sentiment classification.

Keywords: Sentiment Analysis, Multi-Aspect, SVM, Naive Bayes, Random Forest, K-NN, Stacking, TF-IDF, BoW, Word2Vec

Universitas Bakrie

Analisis Sentimen Multi-Aspek pada Ulasan Hotel di Indonesia Menggunakan Klasifier Hibrida Berbasis SVM, NB, RF, dan K-NN

Fazrah Rahmawati Mewar
mewarfazrahrahmawati@gmail.com

Abstrak

Multi-aspect sentiment analysis merupakan tugas krusial untuk memahami opini pengguna secara terperinci pada berbagai faset produk atau layanan. Penelitian ini bertujuan untuk membangun dan mengevaluasi model klasifikasi sentimen multi-aspek yang kuat untuk ulasan hotel berbahasa Indonesia. Empat algoritma machine learning individual, yaitu *Support Vector Machine* (SVM), *Naive Bayes* (NB), *K-Nearest Neighbors* (K-NN), dan *Random Forest* (RF), diimplementasikan dan dibandingkan. Model dilatih menggunakan tiga teknik representasi fitur yang berbeda: *Bag-of-Words* (BoW), *Term Frequency-Inverse Document Frequency* (TF-IDF), dan *Word2Vec*. Selanjutnya, sebuah *Hybrid Classifier* menggunakan metodologi stacking diusulkan untuk menggabungkan kekuatan dari model-model individual. Eksperimen dilakukan pada dataset HoASA dari benchmark IndoNLU. Hasil eksperimen menunjukkan bahwa model hybrid stacking yang diusulkan mencapai akurasi puncak sebesar 93.40%, yang diperoleh saat menggunakan representasi fitur Bag-of-Words (BoW) dan Term Frequency-Inverse Document Frequency (TF-IDF). Angka ini melampaui performa classifier individual terbaik, yaitu SVM dengan fitur TF-IDF yang mencatatkan akurasi sebesar 93.10%. Menariknya, pada pengujian dengan fitur Word2Vec, model Random Forest menunjukkan kinerja sedikit lebih unggul dengan akurasi 86.10%. Kesimpulan dari studi ini menyoroti efektivitas pendekatan hibrida, terutama ketika dipasangkan dengan representasi fitur klasik seperti BoW dan TF-IDF, dalam meningkatkan akurasi klasifikasi sentimen multi-aspek.

Keywords: Analisis Sentimen, Multi-Aspek, SVM, Naive Bayes, Random Forest, K-NN, Stacking, TF-IDF, BoW, Word2Vec.

Contents

List of Figures

List of Tables

List Of Abbreviations

SVM	: Support Vector Machine
NBC	: Naive Bayes Classifier
K-NN	: The K-Nearest Neighbors
RF	: Random Forest
TF-IDF	: Term Frequency-Inverse Document Frequency
BoW	: Bag of Words
NBSVM	: Naive Bayes Support Vector Machine
ABSA	: Aspect-Based Sentiment Analysis
NLP	: Natural Language Processing