

UNIVERSITI TEKNOLOGI MARA

**HYB-ABSAFAKE: A HYBRID
APPROACH INTEGRATING
IMPLICIT ASPECT-BASED
SENTIMENT ANALYSIS WITH
IMBALANCED DATASET
HANDLING FOR ENHANCING
FAKE REVIEW DETECTION**

**LEENA ARDINI BINTI ABDUL
RAHIM**

MSc

October 2025

ABSTRACT

Online shopping has gained popularity due to its convenience and vast product selection, leading consumers to share reviews online. However, the rise of fake reviews undermines consumer trust and misleads potential buyers. Existing detection models often analyze full review text but fail to capture subtle cues such as exaggerated sentiment, duplicated review, or lack of specificity. Moreover, the significant class imbalance, where genuine reviews far outnumber fake ones, leads to biased models that struggle to detect deceptive content effectively. To address these limitations, this research introduces a novel hybrid approach, named HYB-ABSAFAKE, that integrates Bidirectional Encoder Representations from Transformers (BERT) for Implicit Aspect-Based Sentiment Analysis (ABSA), rule-based indicators of fake reviews and Synthetic Minority Over-sampling Technique (SMOTE) to handle imbalanced data. A Support Vector Machine (SVM) is used for classification, and the model is evaluated using k -fold cross-validation. The dataset, obtained from Kaggle's Amazon Reviews Dataset, includes 2,582 reviews from four categories, including foods, home care, personal care, and refreshments. This is the first study to combine implicit ABSA, rule-based indicators, and SMOTE for fake review detection. The hybrid approach is compared with two feature approaches, which are SVM baseline model and a BERT + Rule-based + SVM without SMOTE. It achieved 96% accuracy, 60% precision, 100% recall, and a 75% F1 score, demonstrating improved detection performance, particularly in identifying all fake reviews. The study contributes to fake review research by highlighting how implicit aspects and class imbalance influence deceptive patterns, thereby expanding the theoretical understanding of subtle linguistic cues. Methodologically, it presents a novel sequencing strategy and layered detection process that enhance both model interpretability and the detection of rare cases. Practically, the developed hybrid approach is scalable, explainable, and adaptable to real-world review platforms, supporting greater trust and more informed decision-making. Future work will explore larger annotated datasets, advanced resampling techniques such as Edited Nearest Neighbors, and adaptation of the hybrid approach to domain-specific areas like healthcare or education to improve practical relevance.

ACKNOWLEDGEMENT

Alhamdulillah, praise and thanks to Allah and His Almighty and His utmost blessings, I could finish this research within the given time.

First and foremost, I would like to express my deepest gratitude to my supervisor, Associate Professor Ts. Dr. Khyrina Airin Fariza Binti Abu Samah, for her exceptional guidance, steadfast support, and continuous encouragement throughout this research journey. I am also sincerely grateful to my co-supervisor, Dr. Raiyah Binti Aminuddin, for her insightful advice and dedicated support in refining my research.

A special thanks goes to my beloved father, Abdul Rahim bin Hj. Salam, for his endless prayers, love, support, and encouragement throughout my academic journey. I also wish to express my heartfelt remembrance and love for my late mother,

whose unwavering belief in me continues to be a source of inspiration and strength.

I would also like to thank the Institute of Graduate Studies, Universiti Teknologi MARA, for the Conference Support Fund and Journal Support Fund, which greatly facilitated the smooth progress of my research.

Thanks to all my panels for the Defence Research Proposal, Mock Viva, and Viva Voce for their invaluable feedback, critical insights, and constructive suggestions that have contributed significantly to the improvement of this work. Their expertise and assessments have helped refine my research direction and strengthen the quality of this thesis. Lastly, I would like to express my gratitude to all the speakers from the webinars I attended, whose insightful ideas and guidance have significantly contributed to my research. I am truly fortunate to have had such an incredible support system, providing me with the strength and confidence to complete my thesis.

Thank you very much.

Leena Ardini Binti Abdul Rahim
Master of Science (Computer Science) (CDCS750)

TABLE OF CONTENTS

| | Page |
|--|-------------|
| CONFIRMATION BY PANEL OF EXAMINERS | ii |
| AUTHOR'S DECLARATION | iii |
| ABSTRACT | iv |
| ACKNOWLEDGEMENT | v |
| TABLE OF CONTENTS | vi |
| LIST OF TABLES | x |
| LIST OF FIGURES | xii |
| LIST OF SYMBOLS | xv |
| LIST OF ABBREVIATIONS | xvi |
| | |
| CHAPTER 1 INTRODUCTION | 1 |
| 1.1 Research Background | 1 |
| 1.2 Problem Statements | 7 |
| 1.3 Research Questions | 8 |
| 1.4 Research Objectives | 9 |
| 1.5 Research Motivation | 10 |
| 1.6 Research Scopes | 12 |
| 1.7 Research Contribution | 13 |
| 1.8 Thesis Outline | 13 |
| 1.9 Summary | 14 |
| | |
| CHAPTER 2 LITERATURE REVIEW | 15 |
| 2.1 Introduction | 15 |
| 2.2 Fake Review Detection | 16 |
| 2.2.1 Types of Fake Reviews | 17 |
| 2.2.2 Criteria of Fake Reviews | 19 |
| 2.2.3 Impacts of Fake Reviews on Consumers and Business | 20 |
| 2.2.4 Previous Research Study on Methods Used for Fake Review Detection | 21 |

CHAPTER 1

INTRODUCTION

This chapter examines the research background, problem statements, objectives, scopes, and significance of the research entitled “HYB-ABSAFAKE: A Hybrid Approach Integrating Implicit Aspect-Based Sentiment Analysis with Imbalanced Dataset Handling for Enhancing Fake Review Detection”.

1.1 Research Background

Nowadays, online shopping has become increasingly popular due to its convenience and the wide variety of products available to customers. Additionally, it enabled tremendous growth in business information sharing, allowing customers to share their experiences and feedback on specific products or services online (Duma et al., 2023). The feedback or reviews that are given have become an indispensable tool in determining the success of a product. Not only do they provide valuable feedback to the manufacturer, but they can also influence customer purchasing decisions. Consumers rely heavily on the experiences of others when making purchasing decisions, and compelling reviews are used as a deciding factor in whether a product is added to their cart. Therefore, businesses prioritising customer satisfaction and encouraging customers to leave reviews will likely increase sales and customer loyalty.

Consumer reviews are an essential resource for people since they can lead to significant financial benefits for businesses. In contrast, negative reviews can sometimes have a detrimental effect on the economy (Deshai & Rao, 2022). As a result, reviews posted on online platforms can be subject to various forms of impropriety, including instances of self-promotion or even deliberate fraudulent activity, such as the use of fake reviews, which may be driven by the potential for monetary gain (Jakupov et al., 2022). A study by Duma et al. (2023) revealed that fake reviews are deceptive evaluations made to mislead customers rather than reflect genuine customer experiences. These reviews are often produced by individuals, such as family, friends, bots, or employees who lack first-hand experience with the product or service.

There are many reasons to write fake reviews, such as influencing product