AN ENHANCED RELATIVE PROPORTIONAL DIFFERENCE MODEL TO ANALYZE SENTIMENT INTENSITY OF MALAYSIAN TELECOMMUNICATION OPINIONS

RONIZAM BT ISMAIL
@ CHE WAN MAHMUD

DOCTOR OF PHILOSOPHY

2025

Faculty of Information and Communications Technology

An Enhanced Relative Proportional Difference Model to Analyze Sentiment Intensity of Malaysian Telecommunication Opinions

Ronizam Bt Ismail @ Che Wan Mahmud

Doctor of Philosophy

2025

An Enhanced Relative Proportional Difference Model to Analyze Sentiment Intensity of Malaysian Telecommunication Opinions

RONIZAM BT ISMAIL @ CHE WAN MAHMUD

A thesis submitted
in fulfillment of the requirements for the degree of
Doctor of Philosophy

Faculty of Information and Communications Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2025

DECLARATION

I declare that this thesis entitled “An Enhanced Relative Proportional Difference Model to Analyze Sentiment Intensity of Malaysian Telecommunication Opinions “ is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature:

Name: RONIZAM BT ISMAL@CHE WAN MAHMUD

Date: 29 SEPTEMBER 2025

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Doctor of Philosophy.

Signature: .

Supervisor Name: PROFESSOR TS. DR. BURHANUDDIN BIN MOHD ABOOBAIDER

Date: 29 SEPTEMBER 2025

DEDICATION

To my loving and supportive parents, siblings and in laws

&

My dear husband and precious children

For their love, encouragement, and unwavering support have been my strength throughout this journey.

ABSTRACT

Sentiment intensity analysis is an advanced process within sentiment analysis that not only determines the polarity of opinions but also measures the degree of sentiment strength, offering deeper insights into public opinion. In the Malaysian telecommunications domain, where customer feedback is often expressed in English, Malay, or a mixture of both on informal and noisy social media platforms, the widely used Relative Proportional Difference (RPD) model has been applied for proportional sentiment scoring but suffers from instability, inconsistent outputs, and sensitivity to small threshold changes, leading to unreliable sentiment intensity scores. Thus, this study aims to develop an Enhanced RPD model to improve stability, sensitivity, and consistency in scoring by identifying features that enhance sentiment analysis, designing a model that ensures consistent scoring across multilingual datasets, and evaluating its performance against the baseline RPD. Customer feedback was collected from social media platform and pre-processed for sentiment analysis. The enhanced model incorporates a 0.1 smoothing factor into the original formula, mitigating threshold instability while retaining proportional scoring logic. Its performance was compared against the baseline model using accuracy, F1 score, and stability assessment across multilingual data scenarios. Based on English, Malay, and mixed-language telecom customer feedback, the Enhanced RPD model demonstrates an accuracy of 78.6%, representing a 6.5% improvement over the baseline model, and an F1 score of 78.9%, indicating a 5.7% improvement, and reduces Mean Squared Error by 20% from 0.089 to 0.071. The findings suggest that the Enhanced RPD model provides a more reliable and robust approach for sentiment intensity analysis in multilingual and noisy data environments. This contributes to more accurate sentiment-driven decision-making in the telecommunications sector and holds potential for application in other multilingual industries facing similar analytical challenges. These findings confirm that the Enhanced RPD model delivers more stable, accurate, and fine-grained sentiment intensity scores, addressing multilingual sentiment analysis gaps, providing significant and actionable insights for industry, and offering a reliable foundation for future research in complex real-world data environments.

MODEL PERBEZAAN BERKADARAN RELATIF YANG DIPERTINGKAT UNTUK MENGANALISIS INTENSITI SENTIMEN PANDANGAN TELEKOMUNIKASI DI MALAYSIA

ABSTRAK

Analisis intensiti sentimen merupakan proses lanjutan dalam analisis sentimen yang bukan sahaja menentukan polariti pendapat tetapi juga mengukur tahap kekuatan sentimen, sekaligus memberikan pemahaman yang lebih mendalam terhadap pandangan awam. Dalam domain telekomunikasi Malaysia, di mana maklum balas pelanggan sering dinyatakan dalam bahasa Melayu, bahasa Inggeris, atau gabungan kedua-duanya di platform media sosial yang tidak formal dan bising, model Perbezaan Berkadar Relatif (PBR) yang digunakan secara meluas untuk penilaian sentimen berkadar didapati menghadapi masalah ketidakstabilan, hasil yang tidak konsisten, dan kepekaan terhadap perubahan kecil pada paras penentu, yang membawa kepada skor intensiti sentimen yang kurang dipercayai. Oleh itu, kajian ini bertujuan membangunkan model PBR yang dipertingkat bagi menambahbaik kestabilan, kepekaan, dan konsisten dalam penilaian dengan mengenal pasti ciri-ciri yang dapat memperkukuh analisis sentimen, mereka bentuk model yang mengekalkan penilaian konsisten merentasi set data dwibahasa, dan menilai prestasi PBR yang dipertingkat berbanding PBR asas. Maklum balas pelanggan telah dikumpulkan daripada platform media sosial dan dipra-proses untuk analisis sentimen. Model yang dipertingkat ini menggabungkan faktor pelarasan 0.1 ke dalam formula asal, yang berupaya mengurangkan ketidakstabilan paras penentu sekaligus mengekalkan logik penilaian berkadar. Keberkesanan model telah dibandingkan dengan model asas menggunakan ketepatan, skor F1, dan penilaian kestabilan merentasi senario data berbilang bahasa. Berdasarkan pandangan pelanggan telekomunikasi dalam bahasa Inggeris, Melayu, dan bahasa campuran, model RPD dipertingkat menunjukkan ketepatan sebanyak 78.6%, iaitu peningkatan sebanyak 6.5% berbanding model asas, serta skor F1 sebanyak 78.9% yang menunjukkan peningkatan sebanyak 5.7%, serta pengurangan Ralat Kuasa Dua Min sebanyak 20% daripada 0.089 kepada 0.071. Hasil kajian menunjukkan bahawa model PBR yang dipertingkat menawarkan pendekatan yang lebih boleh dipercayai dan teguh untuk analisis intensiti sentimen dalam persekitaran data berbilang bahasa dan terdapat gangguan. Hal ini menyumbang kepada pembuatan keputusan berasaskan sentimen yang lebih tepat dalam sektor telekomunikasi serta berpotensi untuk diaplikasikan dalam industri berbilang bahasa lain yang berdepan cabaran analisis yang sama. Penemuan ini membuktikan bahawa model PBR yang dipertingkat mampu menghasilkan skor intensiti sentimen yang lebih stabil, tepat, dan terperinci, sekaligus menangani jurang dalam analisis sentimen pelbagai bahasa, menyediakan pandangan yang signifikan dan boleh dilaksanakan kepada industri, serta mewujudkan asas yang kukuh untuk penyelidikan masa hadapan dalam persekitaran data dunia sebenar yang kompleks.

ACKNOWLEDGEMENT

In the Name of Allah, the Most Gracious, the Most Merciful

First and foremost, I would like to thank and praise Allah the Almighty, my Creator, my Sustainer, for everything I received since the beginning of my life. I would like to extend my appreciation to the Universiti Teknikal Malaysia Melaka (UTeM) for providing the research platform.

Undertaking this Ph.D. research has been a truly life-changing experience for me and it wouldn't have been possible without the support and guidance that I received from many people. Foremost, my utmost appreciation goes to my main supervisor, Professor Ts. Dr. Burhanuddin Bin Mohd Aboobaider from Faculty of Artificial Intelligence and Cyber Security, Universiti Teknikal Malaysia Melaka (UTeM) for his continuous support throughout my Ph.D. journey with inspiration and aspiration through the years of my research as the road ahead. I would like to express my sincere gratitude as well to my co-supervisor, Professor Dr. Abd Samad Bin Hasan Basari from Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM) for his guidance and assistance all over the way to the completion of this study.

This thesis is lovingly dedicated to my father, Encik Ismail @ Che Wan Mahmud bin Che Wan Draman that always give me encouragement to finish my research. I am sincerely grateful to him for his support. This thesis is also dedicated to my Umi, Puan Fatimah binti Abdul Lah who always pray for my success and give affection to me. Special thanks to all my brothers (Abang Rie, Abang Yu, Abang Shac), my sisters (Kak Zum, Kak Iza, Kak As) and their caring families who continuously equipped me with their generosity, loving kindness and constant support.

I am particularly indebted to my husband, Mohd Farid bin Abdul Aziz for all his love and encouragement. He has always been there with me from day one of my academic career. My gratitude to all my children; Nurfareehah Qashrina and Fareesh Qaiser who illuminate my life with lively atmosphere and joy.

My heartfelt gratitude also goes to dearest Nazirah Hamid, Suzana Ab Rahman and Nurul Afiqah Abd Rahman for their unwavering love, constant encouragement, and the warmth they have always given me. I am equally thankful to the management and staff of Universiti Islam Melaka, whose shared knowledge, guidance, and expertise have profoundly shaped and influenced the person I am today.Without all of this supports, I doubt this research work could have been accomplished in time.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

MCMC	-	Malaysian Communications and Multimedia Commission
AI	-	Artificial Intelligent
NLP	-	Natural Language Processing
SNS	-	Social Networking Sitel
MNO	-	Mobile Network Operator
MVNO	-	Mobile Virtual Network Operator
CNN	-	Convolutional Neural Networks
RNN	-	Recurrent Neural Networks
SVM	-	Support Vector Machines
NB	-	Naïve Bayes
RPD	-	Relative Proportional Difference
MSE	-	Mean Squared Error
ML	-	Machine Learning

LIST OF APPENDICES

LIST OF PUBLICATIONS

The followings are the list of publications related to the work on this thesis:

1. Ronizam Ismail, M.A. Burhanuddin, Mokhtar Mohd Yusof and Ab. Yuseni Wahab. The Review on Investigation of Barriers in Knowledge Sharing and Partnership in Information Communication Technology Companies. International Business Management, Vol. 10, pp. 4713-4718. 2016.

2. MA Burhanuddin, Ali Abdul-Jabbar Mohammed, Ronizam Ismail, Halizah Basiron. Internet of Things Architecture: Current Challenges and Future Direction of Research. International Journal of Applied Engineering Research, Vol. 12, pp. 11055-11061. 2017

3. M. A. Burhanuddin, Ronizam Ismail, Fatin Shafiqah Rokman, M.K.A Ghani, Chairul Saleh. Mobile Network Data Analytics Assessment Tools via Tweets Using Support Vector Machine Technique. *Journal of Advance Research in Dynamical & Control Systems*, Vol. 10, 2018

4. M. A Burhanuddin, Ronizam Ismail, Nurul Izzaimah Jamaluddin, Ali Abdul-Jabbar Mohammed, Norzaimah Zainol. Analysis of Mobile Service Providers Performance Using Naive Bayes Data Mining Technique. *International Journal of Electrical and Computer Engineering*, Vol 8. 2018

CHAPTER 1

INTRODUCTION

1.1 Introduction

This introductory chapter began with the background of the study, followed by a discussion of the problem. Thereafter, research questions are formulated and used to construct the research objectives. Next, the contribution of the study as well as the scope of the study is presented. Finally, the outline of the remaining chapters of this thesis is also presented.

1.2 Background

The telecommunication industry in Malaysia has been pivotal in shaping the country's economic landscape and technological progress. As a cornerstone of Malaysia's digital economy, this sector has undergone significant transformations, reflecting global trends and local demands (Tiong et al., 2022).

The Malaysian government, through the Malaysian Communications and Multimedia Commission (MCMC) and in collaboration with various stakeholders from the telecommunications sector, drives the Jendela initiative. It represents a significant investment in the country's digital infrastructure, aiming to boost economic growth, support digital education, improve access to digital services, and ensure Malaysia remains competitive in the digital age (MCMC, 2022).

To ensure the success of Jendela Malaysia, leveraging big data and sentiment analysis can provide strategic insights and drive informed decisions. This approach can enhance network infrastructure, improve customer satisfaction, and tailor services to meet user needs more effectively. Gathering such information from the public is vital to provide insights into the real thoughts of people, and this challenge is the object of research in the discipline called "Sentiment Analysis" (Rodríguez-Ibánez et al., 2023).

Sentiment analysis, powered by Artificial Intelligence (AI), has significant potential as an analytical tool to understand public preferences. It is a specialized field of Natural Language Processing (NLP) that leverages AI techniques to automatically detect, classify, and interpret opinions expressed in text (Taherdoost & Madanchian, 2023; Muhammad et al., 2020). By integrating Machine Learning (ML) algorithms and other models, sentiment analysis can go beyond basic polarity classification (positive, neutral, negative) and analyze sentiment intensity, emotion detection, and contextual sentiment trends.

In recent years, there is an increase in interests by many organizations and companies towards the application of sentiment analysis which proved the arisen importance of this field. Furthermore, sentiment analysis has been applied to a wide variety of topics and issues as reported in previous research such as online products reviews (Fung & Belaidan, 2021), hotel reviews (Wen, Liang, & Zhu, 2023), political and financial analysis (Wen et al., 2023; Mishev et al., 2020).

At present, the current approaches used for Social Networks Sites (SNS) sentiment analysis fall into three main categories: machine learning technique, lexicon-based and hybrid methods (Ainapure et al., 2023). The machine learning approach aims to build classifiers by extracting features and algorithms from trained data (Aftab et al., 2023). The other is the lexicon-based approach, which utilizes lexical resources like sentiment lexicons

or dictionaries, to determine the polarity (Mahmood et al., 2020). Lastly, the hybrid technique combines the best of both lexicon-based and machine learning methods to improve the system's performance (Ainapure et al., 2023).

To date, a massive volume of studies has been implemented in mining the sentiment written in a single language, especially English. However, to perform sentiment analysis in the Malaysian context, two things need to be considered. First, sentiment analysis should be applied for the Malay language as Bahasa Melayu is the national language. Second, Malaysians tend to mix both Malay and English language known as *Bahasa rojak* mainly when they write on SNS. Previous sentiment analysis research is limited in fulfilling these two needs (Mountstephens et al., 2023).

This thesis employs an aggregation formulae for sentiment intensity analysis to evaluate public sentiment across five major Malaysian mobile telecommunication companies: Celcom, Digi, Maxis, UMobile, and TuneTalk. A thorough search of the relevant literature yielded that this research is among the first work to apply sentiment intensity analysis using aggregation model for telecommunication sector and, hopefully, it will be a valuable mechanism for the government or player to improve the execution of the project plan in the future.

In demonstrating the performance of the proposed approach, the experimental studies have been conducted. The findings from the research consistently demonstrate the new Malays - English Sentiment Intensity Analysis, outperforms the current leading techniques and enhances overall performance.

1.3 Research Problem

The digital landscape in Malaysia has evolved over the past few years, and it changes the way Malaysians communicate with each other, how they express their thoughts, and how they make decisions. As of 2023, there were approximately 28 million active social media users in Malaysia (Shien et al., 2023). Moreover, according to MCMC (2022) on average, Malaysians spend about 9 hours a day on online activities such as engaging on social media platforms, online shopping, streaming videos, and other similar activities.

Based on the statistic given, it can be seen that SNS has created a new way of communications. Besides, this kind of platform does facilitate real-time marketing which takes business one step ahead by enabling brands to engage with their consumers. Moreover, it allows the business to be close to the target audience, enables the companies or organizations to take direct action in satisfying their customers, produce insights that facilitate the decision-making process and engage in driving business results.

A well-known social networking service (SNS), formerly called Twitter and now rebranded as 'X, is a platform where users can create and share brief posts, limited to 280 characters known as tweets. In 2022, this platform boasts over 206 million daily active users, offering a space where individuals can share their perspectives, connect with others, communicate, and engage in wide-ranging discussions on various subjects (Qi & Shabrina, 2023).

Leveraging the benefits of SNS, telecommunication companies should utilize sentiment analysis on these platforms to gain deeper insights into customer preferences, enhance services, maintain competitiveness, manage issues effectively, detect market trends, and offer personalized experiences. The Malaysian Communications and Multimedia Commission (MCMC) has reported a significant number of complaints related to

telecommunication services. For instance, between January and June 2022, there were 40,863 network-related complaints, underscoring the need for telecommunication companies to enhance their customer service strategies (MCMC 2022b). This significant volume of complaints reflects increasing consumer frustration over network performance, reliability, and service responsiveness. Traditional customer service methods, such as call centers and manual complaint processing, are often reactive and inefficient in addressing such a high volume of grievances (Pi. S. et al., 2024)

To address these challenges, adopting sentiment analysis through social media can be instrumental. A study by Mat Zain et al. (2022) explores the application of sentiment analysis in understanding customer feedback in the telecommunications industry. The researchers employed machine learning techniques to analyze sentiment from Twitter, utilizing Support Vector Machine (SVM), Random Forest, and Naïve Bayes algorithms. Their findings indicated that SVM demonstrated the highest accuracy in detecting and classifying customer sentiments, highlighting its effectiveness in extracting actionable insights to enhance customer service strategies.

Another relevant study by Rahim et al. (2021) investigates public sentiment on Twitter towards major Malaysian mobile telecommunication companies, including Celcom, Digi, and Maxis. The researchers employed machine learning models, such as Support Vector Machine (SVM), Naïve Bayes, and Deep Learning, to classify sentiments. Their findings demonstrated that Deep Learning classifiers achieved the highest performance in sentiment classification, underscoring their effectiveness in analyzing customer opinions to enhance service quality.

These studies highlight the effectiveness of sentiment analysis in understanding and addressing customer complaints in the Malaysian telecommunication sector. By leveraging

such analytical tools, companies can gain valuable insights into customer perceptions and improve their services accordingly.

In this thesis, there are two issues have been addressed which are:

1. The needs for social media sentiment intensity analysis to understand public

In recent years, sentiment analysis has become an indispensable tool for organizations seeking to understand and respond to customer opinions, particularly in highly competitive industries such as telecommunications (Smith & Lee, 2020). In the Malaysian context, telecommunication service providers face mounting challenges in maintaining customer satisfaction and loyalty, as customers often voice their opinions and grievances through digital platforms such as social media and online forums (Ahmad & Lim, 2019). Malaysia is a linguistically and culturally diverse country, with multiple languages spoken across different regions. Opinions expressed in different languages may vary in tone, context, and sentiment, making it challenging to accurately interpret and analyze sentiments.

Although sentiment analysis has been widely applied in the telecommunications industry to understand customer perceptions, most studies have focused primarily on sentiment polarity classification (positive, negative, neutral) without addressing sentiment intensity in detail (Benoit et al., 2018). Existing techniques for sentiment intensity analysis, such as the Relative Proportional Difference (RPD) model, offer an interpretable approach to assess sentiment strength proportionally, but suffer from several critical shortcomings. The baseline RPD model demonstrates inconsistency when handling sparse data or when minor changes occur in threshold values, resulting in unstable and unreliable sentiment intensity scores (Agarwal & Mittal, 2012; Nayak & Sinha, 2018).

To address these limitations, this study focuses on developing and evaluating a more stable and consistent sentiment intensity analysis technique through an enhanced version of