

UNIVERSITI TEKNOLOGI MARA

**NATURAL LANGUAGE PROCESSING FOR
DETECTING CYBERBULLYING IN SOCIAL MEDIA**

MOHAMAD RADZI BIN RAMLAN

BACHELOR OF COMPUTER SCIENCE (HONS.)

JULY 2026

SUPERVISOR APPROVAL

Natural Language Processing for Detecting Cyberbullying in Social Media

By

MOHAMAD RADZI BIN RAMLAN

2024227874

This thesis was prepared under the supervision of the project supervisor, Sulastris Binti Putit. It was submitted to the Faculty of Computer and Mathematical Sciences and was accepted in partial fulfilment of the requirements for the degree of Bachelor of Computer Science (Hons.).

Approved by

.....

SULASTRI BINTI PUTIT

Project Supervisor

AUGUST 9, 2026

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to Allah S.W.T. for granting me the strength, patience, and perseverance to successfully complete this Final Year Project. I would like to extend my sincere appreciation to my esteemed supervisor, Sulastris Binti Putit, and my lecturer, Dr Liew Chin Ying, for their continuous guidance, invaluable feedback, and unwavering support throughout the entire duration of this research. Their expertise and encouragement have been instrumental in shaping this study and keeping me on the right track. My gratitude also goes to the panels and reviewers, especially for their constructive critiques that significantly improved the quality and academic rigour of this thesis. I would also like to thank the faculty and staff of the Faculty of Computer Science and Mathematics at Universiti Teknologi MARA for providing the necessary resources and environment to conduct this research. Furthermore, I am deeply grateful to my parents and family members for their endless love, prayers, and moral support. Their belief in me has been my greatest source of motivation during challenging times. Lastly, I would like to thank my friends and classmates for their shared insights, encouragement, and camaraderie. To everyone who has contributed directly or indirectly to the success of this project thank you.

Abstract

Cyberbullying in Malaysia is a critical issue exacerbated by the widespread use of code-mixed Manglish on social media. Traditional keyword-based moderation systems failed to interpret the local slang and complex informal grammar, leading to high rates of undetected harassment. To address this technical gap, this study developed a localised detection framework by integrating a custom Manglish slang normalisation pipeline with a fine-tuned Bidirectional Encoder Representations from Transformers model. The model was trained on a dataset comprising 26,985 social media comments and deployed within an interactive web prototype equipped with explainable artificial intelligence features. The empirical evaluation demonstrated that the fine-tuned transformer model achieved an accuracy of 96.4% and an F1-score of 0.950. This performance significantly outperformed the traditional Support Vector Machine baseline, which only attained an 81.5% accuracy rate. Furthermore, the slang normalisation pipeline successfully reduced subword fragmentation and preserved contextual semantics before classification. The resulting web-based prototype processed inferences efficiently, providing transparent insights for human moderators. The findings established that combining rule-based lexical normalisation with advanced transformer architectures provided a demonstrated scalability potential and accurate solution for automated content moderation in code-mixed linguistic environments.

TABLE OF CONTENTS

CONTENT	PAGE
SUPERVISOR APPROVAL	ii
STUDENT DECLARATION	iii
ACKNOWLEDGEMENT	iv
ABSTRACT	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	xi
LIST OF TABLES	xiii
CHAPTER 1: INTRODUCTION	1
1.1 Introduction	1
1.2 Background of Study	1
1.3 Problem Statement	2
1.4 Aim and Objectives of the Project	3
1.5 Research Questions	4
1.6 Scope of Study	4
1.7 Significance of the Study	5
1.8 Chapter Summary	6
CHAPTER 2: LITERATURE REVIEW	8
2.1 Introduction	8
2.2 Overview of Cyberbullying	8
2.2.1 Definitions and Classification	9
2.2.2 Prevalence in Malaysia	9