IDENTIFYING HARMFUL COMMENTS FOR TAMIL LANGUAGE ON SOCIAL MEDIA

Prepared by

Ms. Disne Sivalingam

198773R

Dissertation submitted to the Faculty of Information Technology, University of Moratuwa, Sri Lanka for the partial fulfillment of the requirements of the Degree of Master of Science in Information Technology

July 2022

DECLARATION

We declare that this thesis titled “Identifying Harmful Comments for Tamil Language on Social Media” is my own work. Where any part of this thesis has not been submitted in any form for another degree or diploma at this university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged. Due references have been provided on all supporting works of literature and resources.

Name of the Student: Signature of Student:

Ms. Disne Sivalingam *UOM Verified Signature*

Date: 30/07/2022

UOM Verified Signature

Supervised By: Signature of Supervisor:

Dr. S. C. Premaratne
Senior Lecturer
Faculty of Information Technology Date: 30/07/2022
University of Moratuwa

DEDICATION

I dedicate this research report affectionately to the following:

My Research Supervisor, Dr. S. C Premaratne for dedicating his valuable time for guiding me throughout this.

My Husband and Parents for providing me with continuous encouragement throughout my years of study.

Finally for my friends who helped me on giving ideas and support.

Thank you.

ACKNOWLEDGEMENT

I would like to acknowledge and convey my appreciation for those efforts and support in one way or another contributed to the successful completion of this research.

First, I am deeply grateful for the supervision throughout the one year of my research and the help received from my supervisor Dr. S. C. Premaratne, Department of Information Technology, Faculty of Information Technology, University of Moratuwa. I have learned so much from our project discussions. His willingness to motivate me contributed tremendously to my job.

Besides, I would like to thank all academic staff from the Faculty of Information and technology, who shared their vast knowledge throughout these two years by providing me with a good environment which influenced a lot to achieve this goal.

It is with great pleasure to thank the University of Moratuwa, Sri Lanka, for all the efforts and facilities that it has taken to contributions this postgraduate programme. Especially, my thanks and gratitude to my colleagues for their help and support to complete this research in many ways.

I am as ever, especially indebted to my parents, brother, and husband for their love and support throughout my life to improve my career.

ABSTRACT

The era of social media, such as YouTube, Facebook, and Twitter adding comments to posts are being fun in the daily life of people. But this is also used to spread hate speech and organize hate based activities increasingly nowadays. Harmful and offensive text identification on social media platforms is being a trending research area over the last few years. In a country like Sri Lanka with multiple native languages, people like to comment on social media mostly in their native language. Tamil is one of the Languages commonly used and spoken in the North and East part of Sri Lanka. In recent years people like to comment not only in their native language they also comment in more than one language. In Sri Lanka, people use Singlish (Sinhala + English) or Tanglish (Tamil + English).

Because of the rapid growth of hateful content on social media, there is an immediate need for an efficient and effective method to identify harmful content. A huge number of researches have been done and are being done for automated harmful content detection online. The complication of the Natural Language constructs builds this task very challenging.

A maximum of the research are done in the English Language. This research work aims to classify the code-mixed Tamil comments on social media by categorizing them as harmful and non-harmful by using machine learning models.

TABLE OF CONTENTS

ABBREVIATIONS

TF-IDF	Term Frequency-Inverse Document Frequency
WEKA	Waikato Environment for Knowledge Analysis
NB	Naïve Bayes
SVM	Support Vector Machine
LR	Logistic Regression
TN	True Negative
FP	False Positive
FN	False Negative
TP	True Positive

LIST OF FIGURES

LIST OF TABLES