

**UNDERSTANDING THE POLITICAL OPINION OF
SRI LANKANS THROUGH DEEP LEARNING BASED
SOCIAL MEDIA DATA ANALYSIS**

A.A. Ameen
(209309F)

MSc in Computer Science Specializing in
Data Science Engineering and Analytics

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2022

**UNDERSTANDING THE POLITICAL OPINION OF
SRI LANKANS THROUGH DEEP LEARNING BASED
SOCIAL MEDIA DATA ANALYSIS**

A.A. Ameen
(209309F)

This dissertation submitted in partial fulfillment of the requirements for the Degree of
MSc in Computer Science specializing in Data Science Engineering and Analytics

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2022

Declaration

I declare that this is my own work and this dissertation does not incorporate without acknowledgment any material previously submitted for degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief, it does not contain any material previously published or written by another person except where the acknowledgment is made in the text.

Also, I hereby grant to the University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other media. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature:

Date:

Name: A.A. Ameen

The supervisor/s should certify the thesis/dissertation with the following declaration.

I certify that the declaration above by the candidate is true to the best of my knowledge and that this report is acceptable for evaluation for the CS5999 PG Diploma Project.

Signature of the supervisor:

Date:

Name: Dr Surangika Ranathunga

Acknowledgment

I would like to express profound gratitude to my advisor, Dr. Surangika Ranathunga, for her invaluable support in guiding the selection, scoping, and completion of this research study.

Further, I would like to thank all my colleagues for their help in finding relevant research material, for sharing knowledge and experience, and for their encouragement. Big gratitude goes to University of Moratuwa Codestars team and dear friend Sanjeth Kathir for supporting me to complete the thesis. I am as ever, especially indebted to my family for their love and support throughout my life. Finally, I wish to express my gratitude to all my colleagues at work, for the support given to me to manage my MSc research work.

Abstract

With the rising popularity of social media usage, the data generated through it has significantly increased. By focusing critically on these data, certain patterns, traits and opinions can be obtained which can be used for the betterment of the society. This research focuses on finding out the possibility of understanding the political opinion of the country based on these social media data. Understanding the impact of these data will prompt the public to use these platforms to a greater effect thus guiding the decision-makers to make better and informed decisions.

To achieve the above objective English and Sinhala comments from 60 posts from six prominent politicians in Sri Lanka were collected from November 2019 to December 2021. These data was then annotated as positive, negative or neutral sentiments.

6591 annotated comments were used to fine-tune the XLM-RoBERTa (XLM-R) pre-trained model for a text classification task. XLM-R is the new state-of-the-art multilingual masked language model which performs exceptionally well in cross-lingual understanding. To improve the performance of the baseline model a novel approach of adding the context of the comment as a feature in the comment is proposed in the thesis. The XLM-R baseline model achieved an F1 score of 79% while the model using politicians' representation in parliament as a context obtained an F1 score of 91%. The models performed exceptionally well for unseen data as well, when tested with data related to politicians not considered in the training data, the model reported an F1 score of 86%.

Predicting the sentiments using the best model for the latest posts of the six main politicians in the study, the current opinion of the people was derived. Based on it, the Government representatives President Gotabaya Rajapaksha, Prime Minister Mahinda Rajapaksha, and Former President Maithripala Sirisena obtained negative sentiments while Opposition MPs Sajith Premadasa and Anura Kumara obtained positive sentiments.

TABLE OF CONTENTS

Declaration	i
Acknowledgment	ii
Abstract	iii
Chapter 1 INTRODUCTION.....	1
1.1 Research Problem	2
1.2 Research Objectives	2
1.3 Organization of the report	3
Chapter 2 LITERATURE REVIEW.....	4
2.1 Understanding public opinion using Sentiment Analysis Techniques	4
2.1.1 Sentiment Analysis.....	4
2.1.2 Sentiment Analysis to understand political opinion.....	7
2.2 Understanding public opinion using Structural Feature Analysis	10
2.3 Hybrid Approach: Using both Sentiment Analysis and Structural Analysis.....	11
2.4 Evaluating Performance of Text classification tasks.....	12
Chapter 3 METHODOLOGY	14
3.1 Data Collection	15
3.1.1 Data Collection Period	15
3.2 Data Cleaning	16
3.3 Data Annotation.....	16
3.4 Model Building.....	17
3.4.1 Using XLM-R to model the sentiments of the people	18
3.4.2 Model Performance and Evaluation.....	21
3.5 Predicting sentiments for New Politicians.....	21
3.6 Predicting the sentiment for current data.....	22
Chapter 4 ANALYSIS AND RESULTS	23
4.1 Performance evaluation of the models built	25

4.2	Model Validation with new data	26
4.2.1	Results of the Validation data	26
4.3	Predicting the sentiment for the current data.....	27
4.3.1	Understanding the political opinion of people toward the politicians	28
4.3.2	Understanding the overall political opinion of the people	29
Chapter 5	CONCLUSIONS	30
5.1	Conclusions	30
5.2	Limitations and future work	31
Chapter 6	REFERENCES.....	32

LIST OF FIGURES

Figure	Description	Page
Figure 2.1	Sentiment analysis process	5
Figure 2.2	Sentiment classification techniques.....	5
Figure 2.3	Lexicon based sentiment analysis framework	8
Figure 2.4	Machine Learning framework	9
Figure 3.1	Summary of the methodology undertaken in the study	14
Figure 3.2	Summarized view of the data gathering period	15
Figure 3.3	Text classification architecture with XLM-R.....	17
Figure 3.4	Illustration of the models built in the study	18
Figure 4.1	Distribution of the sentiments.....	23
Figure 4.2	Sentiment score of the politicians over the period based on human annotation	24
Figure 4.3	Predicted sentiment score for the current data.....	27
Figure 4.4	Sentiment score of the politicians.....	28
Figure 4.5	Overall sentiment score of the Government and Opposition over the years ..	29

LIST OF TABLES

Table	Description	Page
Table 2.1	Fleiss' Kappa value interpretation	13
Table 3.1	Sample comments and Labels used for the baseline model.....	18
Table 3.2	Sample comments and Labels used for the Language Predictor.....	19
Table 3.3	Sample comments and Labels used for the language tag model.....	19
Table 3.4	Page tag corresponding to the politician	20
Table 3.5	Sample comments and Labels used for the page name as the tag model.....	20
Table 3.6	Tag corresponding to the politician.....	20
Table 3.7	Sample comments and Labels used for the Government tag model	21
Table 4.1	Distribution of comments obtained for the politicians.....	24
Table 4.2	Summary of the sentiment analysis model performance	25
Table 4.3	Summary of the Language predictor model performance.....	25
Table 4.4	The distribution of comments in the validation dataset	26
Table 4.5	Representation in the parliament.....	26
Table 4.6	Performance of the models for the new data.....	26
Table 4.7	Predicted sentiments for the current data	27