



**DEPRESSION DETECTION MODEL BASED ON SOCIAL MEDIA POST
USING SENTIMENT ANALYSIS IN BENGALI LANGUAGE**

By

HASSAN MD HASIBUL

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfilment of the Requirements for the Degree of Master of Science**

August 2022

FSKTM 2022 21

COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs, and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia



Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

**DEPRESSION DETECTION MODEL BASED ON SOCIAL MEDIA POST
USING SENTIMENT ANALYSIS IN BENGALI LANGUAGE**

By

HASSAN MD HASIBUL

August 2022

Chairman : Azrina Kamaruddin, PhD
Faculty : Computer Science and Information Technology

Sadness is a basic human emotion. In our interview with four experts in psychology, they agreed that a person feeling sad for 14 days or longer can be said to be suffering from chronic sadness. Chronic sadness is a major symptom of depression. Numerous research studies have previously been conducted on the detection of depression. The technique adopted for analysis in those studies is Lexicon-based (text) Analysis, which is a popular type of Sentiment Analysis. Previous research on depression did not address symptom identification. If we were to map text polarity in those studies with depressive symptoms based on expert comments, then the symptoms identified in those research studies would be mapped with sadness. However, there are several other potential symptoms of depression that need to be considered, such as insomnia. To identify this symptom, we propose the new DepSymdetect model. The technique used to develop our model is Lexicon-based Analysis with a linguistic dictionary. Our model is focused on the Bengali language. The main reason for choosing Bengali is that it is one of the most widely spoken languages in the world. Another reason is that no such study has ever been conducted in Bengali to identify depressive symptoms using a linguistic dictionary approach. We enhanced the existing linguistic dictionary that was compiled through an opinion review from a past research study in the Bengali language, resulting in the addition of more than 300 new negative sentiment words and 45-degree modifiers. This expanded linguistic dictionary will enable us to identify more negative words for depression analysis. For data collection, we first obtained the consent of ten participants. Then, we manually crawled their Facebook profiles and groups for posts. In our improved DepSymdetect model, we used a number of new features such as text polarity, statement polarity and time of post. With the help of these features, we successfully identified two depressive symptoms, which are chronic sadness and insomnia. We also sought experts to verify the model and the dataset. The experts provided help by validating the newly added negative sentiment words and the degree modifiers. We measured the accuracy of DepSymdetect model that was 93.33%.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

**MODEL PENGESAN KEMURUNGAN BERDASARKAN SIARAN MEDIA
SOSIAL MENGGUNAKAN ANALISIS SENTIMEN DALAM BAHASA
BENGALI**

Oleh

HASSAN MD HASIBUL

Ogos 2022

Pengerusi : Azrina Kamaruddin, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

Perasaan sedih ialah salah satu emosi asas manusia. Dalam temu ramah kami dengan empat orang pakar psikologi, mereka bersetuju bahawa seseorang yang berasa sedih selama 14 hari atau lebih boleh dikatakan sedang mengalami kesedihan kronik. Kesedihan kronik ialah salah satu simptom utama kemurungan. Berbilang kajian penyelidikan telah dijalankan untuk mengesan kemurungan. Teknik yang digunakan untuk menganalisis dalam kajian tersebut ialah Analisis Leksikon (teks), iaitu suatu bentuk Analisis Sentimen yang popular. Penyelidikan yang lepas mengenai kemurungan tidak mengkaji cara mengenal pasti simptom. Andai kata kita memetakan kekutuban teks dalam analisis tersebut dengan simptom kemurungan berdasarkan komen pakar, maka simptom yang dikenal pasti dalam kajian penyelidikan yang lepas itu akan dipetakan dengan kesedihan. Walau bagaimanapun, terdapat beberapa simptom lain, yang mungkin berupa simptom kemurungan, yang perlu dipertimbangkan, misalnya insomnia. Untuk mengenal pasti simptom ini, kami cadangkan model DepSympdetect baharu. Teknik yang digunakan untuk menghasilkan model kami ini ialah Analisis Leksikon berserta kamus linguistik. Model kami tertumpu pada bahasa Bengali. Sebab utama bahasa Bengali dipilih adalah kerana inilah salah satu bahasa yang paling luas dipertuturkan di dunia. Sebab lain adalah kerana kajian belum pernah dilakukan dalam bahasa Bengali untuk mengenal pasti simptom kemurungan menggunakan pendekatan kamus linguistik. Kami meluaskan kamus linguistik yang sedia ada, yang telah disusun melalui penyemakan pendapat daripada kajian penyelidikan yang lepas dalam bahasa Bengali, dan berjaya menambah lebih daripada 300 perkataan sentimen negatif baharu dan 45 darjah pengubah suai. Kamus linguistik yang diperluas ini dapat membantu kami mengenal pasti lebih banyak perkataan sentimen negatif untuk tujuan analisis kemurungan. Untuk mengutip data, kami memperoleh persetujuan sepuluh peserta dahulu. Kemudian, kami merangkak profil dan kumpulan Facebook mereka secara manual untuk mencari bahan siaran. Dalam model DepSympdetect kami yang dipertingkatkan, kami menggunakan beberapa ciri baharu seperti kekutuban teks, kekutuban pernyataan dan masa siaran. Dengan bantuan ciri-ciri ini, kami berjaya

mengenal pasti dua simptom kemurungan, iaitu kesedihan kronik dan insomnia. Kami juga mendapatkan bantuan pakar untuk mengesahkan model dan set data. Para pakar itu memberikan membantu mengesahkan perkataan-perkataan sentimen negatif yang baru ditambah dan darjah pengubah suai. Kami mengukur ketepatan model DepSymdetect iaitu 93.33%.



ACKNOWLEDGEMENTS

In the name of Allah, the Most Gracious and the Most Merciful, Alhamdulillah, I must acknowledge my limitless thanks to Allah, the Ever-Magnificent; the Ever-Thankful, for His blessing in completing this Thesis.

Noteworthy appreciation goes to my supervisor, Senior Lecturer **Dr. AZRINA KAMARUDDIN**, for her patience, motivation, enthusiasm, immense knowledge, and constant support. Her invaluable help of constructive comments and suggestions throughout the thesis work have contributed to the success of this thesis.

I am also thankful to my co-supervisor, Associate Professor **Dr. Masrah Azrifah Azmi Murad**, for her incredible support, guidance, and motivation.

I am extremely grateful to my father (**Md. Sazzadul Islam**) for his generous support, both financial and moral that could not have gone through the master program overseas without his financial support.

Most importantly, I would like to express my wholehearted thanks to my lovely Mother (**Hafiza Akhter**) and the rest of my family for the generous support they provided me throughout my entire life and particularly through the process of pursuing the master's degree. Due to their unconditional love and prayers, I have the chance to complete this degree.

I also extend my thanks to the people who participated in my study for their responses and cooperation. Your response to my questionnaire was important to the success of my project. Finally, Sincere thanks to all my friends for their kindness and moral support during my study. Thanks for the friendship and memories.

Last but not the least I like to thank all my lab members and the staffs of my faculty for their support and cooperation.

THANK YOU VERY MUCH

This thesis was submitted to the Senate of the Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Azrina Kamaruddin, PhD

Senior Lecturer

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Chairman)

Masrah Azrifah Azmi Murad, PhD

Associate Professor

Faculty of Computer Science and Information Technology

Universiti Putra Malaysia

(Member)

ZALILAH MOHD SHARIFF, PhD

Professor and Dean

School of Graduate Studies

Universiti Putra Malaysia

Date: 8 June 2023

Declaration by the Graduate Student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature: _____

Date: _____

Name and Matric No: MD. Hasibul Hassan

Declaration by Members of the Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____

Name of Chairman
of Supervisory

Committee:

Dr. Azrina Kamaruddin

Signature: _____

Name of Member
of Supervisory

Committee:

Associate Professor Dr. Masrah Azrifah Azmi Murad

TABLE OF CONTENTS

	Page
ABSTRACT	i
ABSTRAK	ii
ACKNOWLEDGEMENTS	iv
APPROVAL	v
DECLARATION	vii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
LIST OF APPENDICES	xv
LIST OF ABBREVIATIONS	xvi
CHAPTER	
1 INTRODUCTION	1
1.1 Introduction	1
1.2 Research Motivation	2
1.3 Problem Statement	3
1.4 Research Objectives	4
1.5 Research Questions	5
1.6 Research Scope	5
1.7 Summary	5
2 LITERATURE REVIEW	6
2.1 Introduction	6
2.2 Emotions	6
2.2.1 Depression	7
2.2.2 Identification of Depression	8
2.2.3 Detecting Depression from Social Media	9
2.2.4 Detecting Depression by Analyzing Bengali text	10
2.3 Social Media	10
2.4 Facebook and Depression	11
2.5 Sentiment Analysis	11
2.6 Approaches of Sentiment Analysis	11
2.6.1 Features of Sentiment Analysis	12
2.6.2 Lexicon-based Analysis	13
2.7 Sentiment Analysis in Bengali	17
2.8 Data Extraction	18
2.9 Importance of Expert Validation in HCI	19
2.10 Literature Gap Analysis	19
2.11 Summary	21
3 RESEARCH METHODOLOGY	22
3.1 Introduction	22
3.2 Research design	22
3.2.1 Qualitative analysis	22
3.2.2 Quantitative analysis	22

3.3	Ethics	22
3.4	Participants	23
3.5	Research Phases	24
3.5.1	Phase 1: Analysis of LR	24
3.5.2	Phase 2: Dataset Preparation	25
3.5.3	Phase 3: Algorithm Preparation	26
3.5.4	Phase 4: Implementation	27
3.5.5	Phase 5: Expert Validation	27
3.6	Summary	32
4	DEPSYMPDETECT MODEL IMPLEMENTATION	33
4.1	Introduction	33
4.2	The DepSympdetect Model	33
4.3	DepSympdetect Model Flow	34
4.4	Segment 1: Data Collection	36
4.5	Segment 2: Text Preprocessing	37
4.5.1	Punctuation Removal	37
4.5.2	Pronouns Removal	38
4.5.3	Emoji Removal	40
4.5.4	Tokenization	42
4.6	Segment 3: Sentiment Engine	43
4.6.1	Lexical Analysis	44
4.6.2	Sentiment Analysis	44
4.6.3	Linguistic Rules	44
4.6.4	Sentiment Calculation	46
4.6.5	Depression Analysis	48
4.6.6	Statement Labelling	50
4.7	Segment 4: Enhancement of Linguistic Dictionary	50
4.8	Segment 5: Symptom Detection	51
4.9	Segment 6: User Interface	51
4.10	Implementation	52
4.10.1	Dataset Extraction	52
4.10.2	Accuracy Metrics	52
4.10.3	Experiment Design	54
4.11	Discussion	54
4.12	Summary	54
5	RESULT INTERPRETATION	55
5.1	Introduction	55
5.2	Results from Expert Validation	55
5.3	Dataset Validation by the Experts	57
5.4	Results of Dataset Validation by the Expert	58
5.5	Comparison between Classification by Expert and DepSympdetect	66
5.6	Accuracy Comparison of SA and DA Models in Bengali	67
5.7	Discussion	69
5.8	Evaluation of Study Achievements	71
5.9	Summary	72

6	CONCLUSION AND FUTURE WORK	73
6.1	Introduction	73
6.2	Research Findings	73
6.3	Limitation of the Research	74
6.4	Future Work	74
6.5	Contributions of the study	74
	REFERENCES	76
	APPENDICES	83
	BIODATA OF STUDENT	175
	LIST OF PUBLICATIONS	176



LIST OF TABLES

Table		Page
2.1	Literature Review on the Lexicon-based Approach	14
2.2	Literature Review for Sentiment Analysis in Bengali	18
3.1	Tools for implementation	27
3.2	Profile of Experts	28
3.3	Inaccurate translation from Bengali to English using NLP	30
4.1	Punctuation Removal	38
4.2	Bengali Pronouns	40
4.3	Pronoun Removal	40
4.4	Emoji Removal	42
4.5	Tokenization	43
4.6	Sentiment Words	45
4.7	Variable Weight	46
4.8	Linguistic Rules	46
5.1	Experts' Comments on the Model	57
5.2	Result of Dataset Validation by Experts	60
5.3	Summary of Expert Classification	66
5.4	Accuracy of Sentiment Analysis models in Bengali	67
5.5	Accuracy of Depression Analysis Models in Bengali	68
5.6	Result of Test Run with Random Dataset	71
5.7	Word identification performance of the enhanced linguistic dictionary	72

LIST OF FIGURES

Figure	Page	
2.1	Approaches of Sentiment Analysis	12
2.2	Steps in Dey's model (Dey & Sarker, 2019)	15
2.3	The Conceptual Framework of Tao (Tao et al., 2016)	17
3.1	Facebook groups used for data collection	23
3.2	Research Methodology Steps	24
3.3	Sample Dataset	26
3.4	Translated Sample Dataset for Expert Validation	31
4.1	DepSymdetect Model	34
4.2	DepSymdetect Model Flow Diagram	35
4.3	Input panel of the DepSymdetect model	36
4.4	Punctuation Removal	37
4.5	Pronoun removal	39
4.6	Use of Wrong Emojis	41
4.7	Emoji Removal	41
4.8	Tokenization	43
4.9	DepSymdetect model algorithm	49
4.10	Verified words for linguistic dictionary enhancement	51
5.1	Question about Depressive Symptoms	55
5.2	Symptoms identified by the Experts	56
5.3	Statement Validation Table	58
5.4	Classification of Statements by the DepSymdetect Model	66
5.5	Identified Depressive Symptoms by DepSymdetect	67
5.6	Accuracy of Sentiment Analysis Models in Bengali	68
5.7	Precision Comparison Graph	69

5.8	Recall Comparison Graph	69
5.9	F-Score Comparison Chat	69



LIST OF APPENDICES

Appendix		Page
A	Expert Profile	83
B	Questionaries & Answers	84
C	Participant Consent Form	101
D	Sample Dataset For Expert Validation	105
E	Ethical Documents	118
E1	Apply for Ethical Approval	118
E2	Response from Ethical Committee	118
E3	First Correction List	119
E4	Submission of First Correction	119
E5	Second Correction List	120
E6	Submission of Second Correction	120
E7	Online Meeting with Local Expert	120
F	Verified Words List For Linguistic Dictionary Enhancement	121
G	Data Validation By The Experts	129
H	Ethical Approval Letter	174

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
CNN	Convolution Neural Networking
D_MW	Degree Of Modifiers
Did	Data ID
DL	Deep Learning
E_NW	Emotional Negative Words
E_PW	Emotional Positive Words
FB	Facebook
fp	False Positive
fn	False Negative
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
ML	Machine Learning
NLP	Natural Language Processing
N_PS	Number of Positive Statement
N_NS	Number of Negative Statement
Pid	Patient ID
RNN	Recurrent Neural Network
SA	Sentiment Analysis
SVM	Support-Vector Machines
SP	Statement Polarity
Sid	Statement ID
TD	Text Decision
TF-IDF	Term Frequency-Inverse Document Frequency
ToP	Time of Post

TP	Text Polarity
tn	True Negative
tp	True Positive
UPM	University Putra Malaysia



CHAPTER 1

INTRODUCTION

1.1 Introduction

With the advancement of technology, we now live in an era of open borders and abundant knowledge. The technical revolution has merged our social and personal lives. Social media has enabled us to connect with people anywhere in the world. Facebook and Twitter are two of the most well-known and prominent names in this world of social media (Heravi et al., 2018). The number of social media users is increasing day by day, especially on Facebook. In 2019, there were 2.77 billion users of Facebook worldwide (Drus & Khalid, 2019). People use Facebook not only to express their societal opinions, but also to voice out their personal views about a certain topic or issue of their interest. As a result, Facebook has expanded into such a vast medium of communication and sharing of emotional data through text, image and video that it has become a huge database of human emotions that portray people's personalities and reflect the states of their mental health (Andriessen et al., 2019; Harper, 2014).

Depression is currently the fastest growing and most prevalent mental illness affecting all ages. Its long-term effects are increasing as well (Shrestha & Spezzano, 2019; Almeida et al., 2017; Tao et al., 2016). In the US, 6.7% of adults experience this illness that adversely impacts their thoughts and behaviour. In Japan alone, approximately 24% of young adults (aged 20-24 years) and 18% of teenagers (aged 15-20 years) became victims of unipolar depression (Diego De Leo, 2015). Depression is the cause of frequent mood swings and the reason for impermanent emotional responses, which make life difficult for the sufferers and the people around them. In everyday living, depression makes routine activities harder and gradually manifests itself as a significant health issue. Suicide is the worst-case consequence of depression, with approximately 800,000 people committing suicide each year because of depression (Andriessen et al., 2019). According to a 2015 report by the World Health Organization, three hundred million people have been affected by depression. Of great concern is that its rate of incidence has been rising among adolescents and young adults in recent years, from 8.7% to 11.3% and from 8.8% to 9.6% respectively (Shrestha & Spezzano, 2019). Girls and women are at greater risk, having a 50% higher probability of suffering from depression as a result of hypertension, dementia and silent stroke than men (Kumar et al., 2019).

People who are depressed tend to become introverted and therefore participate more actively on social media sites like Facebook (Alexopoulos, 2003) than in real life. Because people post so much on Facebook about their opinions and emotions, it is easy to learn about their personalities and understand their moral values just by analysing their posts (Goswami & Poray, 2017).

Much research has been undertaken using Sentiment Analysis in the English language to detect depressive symptoms (M. Rahman et al., 2020; Shrestha & Spezzano, 2019; A. H. Uddin, Bapery, & Mohammad Arif, 2019). Present-day research has shown that

Sentiment Analysis has crossed the boundaries of language and is used around the world. The other languages that have adopted Sentiment Analysis to identify depressive symptoms, besides English, are Korean (Song et al., 2019), Thai (Sanguansat, 2016), Arabic (Itani et al., 2017), Malay (Hossen et al., 2018), Portuguese and Chinese (Peng et al., 2017).

Sentiment Analysis has also been done in the national language of Bangladesh to detect depression. In literature, we have found both “Bangla” and “Bengali” being used to refer to the national language of Bangladesh. In this study, we use the term Bengali according to the language convention of the Bangla Academy. The Bangla Academy is a government-funded independent institution whose mission is to promote the Bengali language, literature and culture, to design and execute the national language policy, and to carry out original research in Bengali.

Bengali is increasingly being used in Facebook. In Bangladesh, people use Facebook daily to communicate and to transact business. Facebook is also used as a channel for sharing random feelings or updates on daily activities (A. H. Uddin, Bapery, & Arif, 2019). People share every moment in Facebook. Their posts contain both positive and negative emotive text and reveal the hidden emotions of a person, especially negative ones (Almeida et al., 2017).

Extensive research has been conducted in Bengali on the topic of opinion evaluation and the identification of depression (M. Rahman et al., 2020; A. H. Uddin, Bapery, & Arif, 2019; A. H. Uddin, Bapery, & Mohammad Arif, 2019; Arafin Mahtab et al., 2018). Initially, Sentiment Analysis was employed in the e-commerce industry to gather and record public sentiments about certain products. To detect depression using Sentiment Analysis, two main approaches have been taken, namely Machine Learning and Lexicon-based Analysis. The Lexicon-based Analysis method is used in this investigation. In Bangladesh, relatively little research has been done utilizing linguistic dictionaries, which is why this strategy was chosen (Dey & Sarker, 2019). Another point to note is that in Lexicon-based Analysis, the Machine Learning model is not believed to have a significant impact on accuracy (Goswamil & Poray, 2017).

1.2 Research Motivation

During the period between 2013 and 2019, the rate of suicides in Bangladesh kept rising. In fact, 128.08 persons in every 100,000 commit suicide every year (Ferdous & Alam, 2021). This figure is rather alarming. Police investigation revealed that in a majority of those cases, the person was depressed and a number of them had also been attending psychiatrist consultations. A pattern found to be common among these suicide cases was that all of them had put up many sad and negative posts on Facebook. Some of them even posted farewell notes on Facebook before committing suicide. The growing suicide rates motivated this research study to be conducted.

1.3 Problem Statement

Depression analysis models in the Bengali language have been developed in past research works (Asad et al., 2019; A. H. Uddin, Bapery, & Arif, 2019; A. H. Uddin, Bapery, & Mohammad Arif, 2019). However, all past studies mentioned the problem of detecting depressive symptoms in order to perform depression analysis.

- **Problem #1:** There is no mention of any detection technique to identify depressive symptoms in past research studies. Existing studies claim to detect depression based only on feature sets, without mentioning any symptoms identified for depression.

In the Bengali language, research studies have been conducted to detect depression (Asad et al., 2019; A. H. Uddin, Bapery, & Arif, 2019; A. H. Uddin, Bapery, & Mohammad Arif, 2019). Text and word polarities were the main features of those studies but depressive symptoms were not mentioned.

Another study described the detection of depression using Lexicon-based Analysis (Tao et al., 2016). Although word usage and frequency were used as the key features, the identification of depressive symptom was not mentioned.

Techniques for detecting depression have been discussed in past studies. One study used a set of ten features including emoticons, sentiment polarity, word frequency and time (Wang et al., 2013). However, the study is very old and based on Machine Learning, which requires a large amount of data for training purposes.

All the above-mentioned research studies claim to detect depression, but did not mention any depressive symptoms. Our interview with experts in psychology revealed that if we mapped those features with depressive symptoms, then the identified symptom would be sadness. In addition, a number of potential symptoms can be identified from social media posts and need to be considered as well for depression analysis. No such study has ever been conducted in Bengali on depression analysis using a linguistic dictionary.

- **Subproblem:** A linguistic dictionary in Bengali has been introduced for opinion review and is able to identify positive and negative words (Dey & Sarker, 2019). In depression analysis, several other words need to be considered as negative, carrying a depressive meaning. During our interview with the experts, we learned that depression analysis differs slightly from opinion review when text analysis is used. Opinion reviews involve one-off statements about something or someone, but depression is a long-term condition. In some cases, even positive words could carry a depressive meaning. Hence, these words need to be considered as well. We therefore need to expand the existing dictionary with more words based on linguistic criteria to be able to identify more negative sentiment words.

- **Problem #2:** There was no involvement of domain experts in the verification and validation of existing depression analysis models in the Bengali language (Asad et al., 2019; A. H. Uddin, Bapery, & Arif, 2019; A. H. Uddin, Bapery, & Mohammad Arif, 2019). As these models work with depression, it should be necessary for experts in psychology or psychiatry to verify and validate the classification made by these models. Without this step, the result could be questionable.
- A general depression analysis model in the English language has been described in a past study (Tao et al., 2016). A model for depression analysis in the Chinese language using the Sentiment Analysis technique has also been thought of (Wang et al., 2013). Among the features used were emoticons, first person pronoun and blog posting time. The model uses data mining, which requires a lot of data training. Considering almost a decade has passed, the study is now outdated and needs to be reviewed. Applying the linguistic dictionary approach can reduce the overhead of data training and the analysis can be performed with fewer features.

Another study mentions a depression analysis model in Bengali based on Deep Learning (A. H. Uddin, Bapery, & Ari, 2019). It uses word polarity and frequency as features for depression analysis. In this study, there is no mention of insomnia as a depressive symptom. The study analysed depression based only on the identification of sadness.

From a psychiatrist's point of view, identifying only sadness would be insufficient to detect depression, as confirmed during our interview with the experts. According to the experts, there are other major symptoms that need to be considered in order to determine that a person is depressed. This is a major limitation of Uddin's research. Unless it is resolved, the results could mislead the analysis and cause misclassification.

There has been no study of note in the Bengali language on depression analysis using a linguistic dictionary. A linguistic dictionary has only been used for opinion review (Dey & Sarker, 2019). Nevertheless, all research done thus far in Bengali have taken the Machine Learning approach to detect depression. Machine Learning requires a large amount of training data and feature sets, which is an overhead, making the implementation of the model very difficult. To reduce this overhead and ease implementation, the linguistic dictionary can be combined with the depressive symptom analysis.

1.4 Research Objectives

For this research, we narrowed down our focus to the following objectives:

1. RO1: To identify other depressive symptom from social media using Expert Validation and Comparison method.
2. RO2: To identify more negative sentiment words by enhancing the existing linguistic dictionary.

3. RO3: To evaluate the accuracy of the identified sentiment words and verify the symptoms detected by the model with the help of experts.

1.5 Research Questions

1. RQ1: What other depressive symptoms can be identified from social media in the Bengali language?
2. RQ2: How can the improved DepSympdetect model identify more negative sentiment words to detect sadness?
3. RQ3: How can the accuracy of the model be evaluated from a psychiatric point of view?

1.6 Research Scope

In this study, we experimented by combining an existing depression analysis model (Tao et al., 2016) with an existing linguistic dictionary (Dey & Sarker, 2019) to identify depressive symptoms in Bengali. We also applied an existing format for linguistic rules (Wang et al., 2013) to build a set of linguistic rules for the Bengali language. We also used an existing feature list (Wang et al., 2013) and modified it by adding two new features, which are ToP (time of post) and statement polarity. After combination and modification, we propose our new model for detecting depressive symptoms, which we call DepSympdetect. In our DepSympdetect model, the feature set we use comprises text polarity (TP), time of post (ToP) and statement polarity (SP).

1.7 Summary

This thesis is organized into seven chapters, beginning with the current introductory chapter, followed by:

Chapter 2 covers the literature review and our analysis of the existing models for detecting depression. It also introduces our proposed model, named DepSympdetect, that is designed based on text analysis using a linguistic approach.

Chapter 3 outlines the research methodology adopted to construct the model using a new parameter.

Chapter 4 puts forth the implementation of the model.

Chapter 5 presents the analysis and interpretation of results produced by DepSympdetect.

Chapter 6 concludes the thesis, states the research findings, suggests future work that can be done and lists down the contributions of this research.

REFERENCES

- Malaysia: Facebook users 2018-2023 | Statista. (n.d.). Retrieved December 1, 2020, from <https://www.statista.com/statistics/490484/number-of-malaysia-facebook-users/>
- Thailand: number of Facebook users 2017-2025 | Statista. (n.d.). Retrieved December 1, 2020, from <https://www.statista.com/statistics/490467/number-of-thailand-facebook-users/>
- Agarwal, B., Mittal, N., Bansal, P., & Garg, S. (2015). Sentiment analysis using common-sense and context information. *Computational Intelligence and Neuroscience*, 2015. <https://doi.org/10.1155/2015/715730>
- Alexopoulos, G. S. (2003). *Vascular Disease, Depression, and Dementia*. *Ci*, 1178–1180.
- Almeida, H., Briand, A., & Meurs, M. (2017). *Detecting Early Risk of Depression from Social Media User-generated Content*.
- Andriessen, K., Krysinska, K., Kölves, K., & Reavley, N. (2019). Suicide Postvention Service Models and Guidelines 2014–2019: A Systematic Review. *Frontiers in Psychology*, 10(November). <https://doi.org/10.3389/fpsyg.2019.02677>
- Arafin Mahtab, S., Islam, N., & Mahfuzur Rahaman, M. (2018). Sentiment Analysis on Bangladesh Cricket with Support Vector Machine. *2018 International Conference on Bangla Speech and Language Processing, ICBSLP 2018*, 1–4. <https://doi.org/10.1109/ICBSLP.2018.8554585>
- Asad, N. Al, Mahmud Pranto, M. A., Afreen, S., & Islam, M. M. (2019). Depression Detection by Analyzing Social Media Posts of User. *2019 IEEE International Conference on Signal Processing, Information, Communication and Systems, SPICSCON 2019*, 13–17. <https://doi.org/10.1109/SPICSCON48833.2019.9065101>
- Aziz, M. T. (2016). Sentiment Analysis On Facebook Group Using Lexicon Based Approach. *2016 3rd International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*, 1–4. <https://doi.org/10.1109/CEEICT.2016.7873080>
- Baker, D. A., & Algorta, G. P. (2016). The Relationship between Online Social Networking and Depression: A Systematic Review of Quantitative Studies. *Cyberpsychology, Behavior, and Social Networking*, 19(11), 638–648. <https://doi.org/10.1089/cyber.2016.0206>
- Barrett, L. F., Mesquita, B., Ochsner, K. N., & Gross, J. J. (2007). The experience of emotion. *Annual Review of Psychology*, 58, 373–403. <https://doi.org/10.1146/annurev.psych.58.110405.085709>

- Cacheda, F., Fernandez, D., Novoa, F. J., & Carneiro, V. (2019). *Early Detection of Depression: Social Network Analysis and Random Forest Techniques*. *Corresponding Author: 21*. <https://doi.org/10.2196/12554>
- Casarella, J. (2020). *Types of Depression: Major, Chronic, Manic, and More Types*. WebMD. <https://www.webmd.com/depression/guide/depression-types>
- Dey, R. C., & Sarker, O. (2019). Sentiment analysis on bengali text using lexicon based approach. *2019 22nd International Conference on Computer and Information Technology, ICCIT 2019*, 18–20. <https://doi.org/10.1109/ICCIT48885.2019.9038250>
- Dhaoui, C., Webster, C. M., & Tan, L. P. (2017). Social media sentiment analysis: lexicon versus machine learning. *Journal of Consumer Marketing*, 34(6), 480–488. <https://doi.org/10.1108/JCM-03-2017-2141>
- Diego De Leo, K. K. (2015). Child, Adolescent and Young Adult Suicides: A Comparison Based on the Queensland Suicide Registry. *Journal of Child and Adolescent Behaviour*, 03(03). <https://doi.org/10.4172/2375-4494.1000209>
- Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, 161, 707–714. <https://doi.org/10.1016/j.procs.2019.11.174>
- El Rahman, S. A., Alotaibi, F. A., & Alshehri, W. A. (2019). Sentiment Analysis of Twitter Data. *2019 International Conference on Computer and Information Sciences, ICCIS 2019*. <https://doi.org/10.1109/ICCISci.2019.8716464>
- Facebook users in Bangladesh - March 2020 / NapoleonCat*. (n.d.). Retrieved December 1, 2020, from <https://napoleoncat.com/stats/facebook-users-in-bangladesh/2020/03>
- Fatyanosa, T. N., & Bachtiar, F. A. (2018). Classification method comparison on Indonesian social media sentiment analysis. *Proceedings - 2017 International Conference on Sustainable Information Engineering and Technology, SIET 2017, 2018-Janua*, 310–315. <https://doi.org/10.1109/SIET.2017.8304154>
- Ferdous, M. Z., & Alam, A. S. M. M. (2021). PRESENT SITUATION OF SUICIDE IN BANGLADESH: A REVIEW. *MedRxiv*, 2021.02.23.21252279. <https://doi.org/10.1101/2021.02.23.21252279>
- Giri, K. J. (2015). *International Journal of Advanced Research in Big Data - Overview and Challenges*. July 2014.
- Goswamil, S., & Poray, J. (2017). Human computer interaction for sentiment analysis and opinion mining: A review. *2016 International Conference on Computer, Electrical and Communication Engineering, ICCECE 2016*. <https://doi.org/10.1109/ICCECE.2016.8009586>
- Gu, S., Wang, F., Patel, N. P., Bourgeois, J. A., & Huang, J. H. (2019). A model for basic emotions using observations of behavior in *Drosophila*. *Frontiers in Psychology*, 10(APR), 1–13. <https://doi.org/10.3389/fpsyg.2019.00781>

- Hamm, M. P., Newton, A. S., Chisholm, A., Shulhan, J., Milne, A., Sundar, P., Ennis, H., Scott, S. D., & Hartling, L. (2015). Prevalence and effect of cyberbullying on children and young people: A scoping review of social media studies. *JAMA Pediatrics*, 169(8), 770–777. <https://doi.org/10.1001/jamapediatrics.2015.0944>
- Haimson, O. L., Cotten, S. R., Clark, J. L., & Shin, N. (2021). Designing Trans Health Technologies: Expert Perspectives on Patient-Centeredness and Cultural Humility. *Journal of Medical Internet Research*, 23(2), e20277. doi: 10.2196/20277
- Hao, J., & Dai, H. (2016). Social media content and sentiment analysis on consumer security breaches. *Journal of Financial Crime*, 23(4), 855–869. <https://doi.org/10.1108/JFC-01-2016-0001>
- Harper, C. (2014). Diagnostic and statistical manual of mental disorders. In *A Companion to Criminal Justice, Mental Health & Risk*. <https://doi.org/10.1097/00002093-198802020-00022>
- Hasan, M.M. and Arefin, M.S., 2021. Sentiment Analysis of Bengali Language using Transfer Learning Approach. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 87-92). IEEE.
- Hassan, A. U., Hussain, J., Hussain, M., Sadiq, M., & Lee, S. (2017). Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression. *International Conference on Information and Communication Technology Convergence: ICT Convergence Technologies Leading the Fourth Industrial Revolution, ICTC 2017, 2017-Decem*, 138–140. <https://doi.org/10.1109/ICTC.2017.8190959>
- Heravi, B. R., Mcleod, J., Eds, P. W., & Hutchison, D. (2018). Transforming Digital Worlds. In *Transforming Digital Worlds* (Vol. 10766). Springer International Publishing. <https://doi.org/10.1007/978-3-319-78105-1>
- Hossen, I., Goh, M., Connie, T., & Aris, A. (2018). *Sentiment Analysis of Malay Social Media Text*. 3(January 2019), 241–251. <https://doi.org/10.1007/978-981-10-8276-4>
- Ikoru, V., Sharmina, M., Malik, K., & Batista-navarro, R. (2018). Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers. *2018 5th International Conference on Social Networks Analysis, Management and Security, SNAMS 2018*, 95–98.
- Isah, H., Trundle, P., & Neagu, D. (2014). Social media analysis for product safety using text mining and sentiment analysis. *2014 14th UK Workshop on Computational Intelligence, UKCI 2014 - Proceedings*, 1–7. <https://doi.org/10.1109/UKCI.2014.6930158>
- Islam, M. A., Islam, A. B. M. A. Al, & Anik, M. S. H. (2017). Polygot: An approach towards reliable translation by name identification and memory optimization using semantic analysis. *Proceedings of 2017 International Conference on Networking, Systems and Security, NSysS 2017, 2018-Janua*, 1–8. <https://doi.org/10.1109/NSYSS2.2017.8267795>

- Itani, M., Roast, C., & Al-Khayatt, S. (2017). Developing Resources for Sentiment Analysis of Informal Arabic Text in Social Media. *Procedia Computer Science*, 117, 129–136. <https://doi.org/10.1016/j.procs.2017.10.101>
- Joyce, B., & Deng, J. (2018). Sentiment analysis of tweets for the 2016 US presidential election. *2017 IEEE MIT Undergraduate Research Technology Conference, URTC 2017, 2018-Janua*, 1–4. <https://doi.org/10.1109/URTC.2017.8284176>
- Karamollaoğlu, Doğru, D., & Yıldız, U. &. (2018). *Sentiment Analysis on Turkish Social Media Shares through Lexicon Based Approach*. 5–9.
- Khan, M. T., Durrani, M., Ali, A., Inayat, I., Khalid, S., & Khan, K. H. (2016). Sentiment analysis and the complex natural language. *Complex Adaptive Systems Modeling*, 4(1). <https://doi.org/10.1186/s40294-016-0016-9>
- Kumar, A., Sharma, A., & Arora, A. (2019). *Anxious Depression Prediction in Real-time Social Data*. 2019, 1–7.
- Lazar, J., Feng, J. H., & Hochheiser, H. (2017). Research Methods in Human-Computer Interaction. In *Research Methods in Human-Computer Interaction*. <https://doi.org/10.1016/b978-0-444-70536-5.50047-6>
- Leventhal, A. M. (2008). Behavior Modification and Avoidance Behavior. *Behavior Modification*, 32(6), 759–779.
- List of Emotions: 5 Ways to Express Your Feelings*. (n.d.). Retrieved January 8, 2023, from <https://greatist.com/health/list-of-emotions>
- Lup, K., Trub, L., & Rosenthal, L. (2015). Instagram #Instasad?: Exploring Associations Among Instagram Use, Depressive Symptoms, Negative Social Comparison, and Strangers Followed. *Cyberpsychology, Behavior, and Social Networking*, 18(5), 247–252. <https://doi.org/10.1089/cyber.2014.0560>
- Maciej Serda, Becker, F. G., Cleary, M., Team, R. M., Holtermann, H., The, D., Agenda, N., Science, P., Sk, S. K., Hinnebusch, R., Hinnebusch A, R., Rabinovich, I., Olmert, Y., Uld, D. Q. G. L. Q., Ri, W. K. H. U., Lq, V., Frxqwu, W. K. H., Zklfk, E., Edvhg, L. V, ... فاطمی ح. (2013). Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza. *Uniwersytet Śląski*, 7(1), 343–354. <https://doi.org/10.2/JQUERY.MIN.JS>
- Mamun, M. A., Hossain, M. S., & Griffiths, M. D. (2019). Mental Health Problems and Associated Predictors Among Bangladeshi Students. *International Journal of Mental Health and Addiction*. <https://doi.org/10.1007/s11469-019-00144-8>
- Mansour, S. (2018). Social media analysis of user's responses to terrorism using sentiment analysis and text mining. *Procedia Computer Science*, 140, 95–103. <https://doi.org/10.1016/j.procs.2018.10.297>
- Mäntylä, M. V., Graziotin, D., & Kuuttila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16–32. <https://doi.org/10.1016/j.cosrev.2017.10.002>

- Martin-Domingo, L., Martín, J. C., & Mandsberg, G. (2019). Social media as a resource for sentiment analysis of Airport Service Quality (ASQ). *Journal of Air Transport Management*, 78(January), 106–115. <https://doi.org/10.1016/j.jairtraman.2019.01.004>
- Mishra, N., & K. Jha, C. (2012). Classification of Opinion Mining Techniques. *International Journal of Computer Applications*, 56(13), 1–6. <https://doi.org/10.5120/8948-3122>
- Moallem, A., & Alemi, M. (2021). Usability evaluation of mobile health applications: Expert review versus user testing. *International Journal of Medical Informatics*, 150, 104429. doi: 10.1016/j.ijmedinf.2021.104429
- Nadeem, M. (2016). *Identifying Depression on Twitter*. 1–9. <http://arxiv.org/abs/1607.07384>
- Peng, H., Cambria, E., & Hussain, A. (2017). A Review of Sentiment Analysis Research in Chinese Language. *Cognitive Computation*, 9(4), 423–435. <https://doi.org/10.1007/s12559-017-9470-8>
- Rahman, L., Sarowar, G., & Kamal, S. (2018). Teenagers Sentiment Analysis from Social Network Data. *Social Networks Science: Design, Implementation, Security, and Challenges*, 3–23. https://doi.org/10.1007/978-3-319-90059-9_1
- Rahman, M., Haque, S., & Rahman, Z. (2020). Identifying and Categorizing Opinions Expressed in Bangla Sentences using Deep Learning Technique. *International Journal of Computer Applications*, 176(17), 13–17. <https://doi.org/10.5120/ijca2020920119>
- Ramesh, K., Anto Bennet, M., Mari Arathy, T., Kavia, V., Preetha, S., & Nivedha, G. (2019). Analysis of the attitude of social network analysis using r3 tool. *Proceedings of the 3rd International Conference on Computing Methodologies and Communication, ICCMC 2019, Iccmc*, 998–1003. <https://doi.org/10.1109/ICCMC.2019.8819774>
- Riyadh, A. Z., Alvi, N., & Talukder, K. H. (2018). Exploring human emotion via Twitter. In *20th International Conference of Computer and Information Technology, ICCIT 2017* (Vols. 2018-Janua, pp. 1–5). <https://doi.org/10.1109/ICCITECHN.2017.8281813>
- Sanguansat, P. (2016). Paragraph2Vec-based sentiment analysis on social media for business in Thailand. *2016 8th International Conference on Knowledge and Smart Technology, KST 2016*, 175–178. <https://doi.org/10.1109/KST.2016.7440526>
- Seabrook, E. M., Kern, M. L., & Rickard, N. S. (2016). Social Networking Sites, Depression, and Anxiety: A Systematic Review. *JMIR Mental Health*, 3(4), e50. <https://doi.org/10.2196/mental.5842>
- Shayaa, S., Wai, P. S., Chung, Y. W., Jaafar, N. I., & Zakaria, S. B. (2017). *Social Media Sentiment Analysis on Employment in Malaysia. October*.

- Shrestha, A., & Spezzano, F. (2019). Detecting depressed users in online forums. *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2019*, 945–951. <https://doi.org/10.1145/3341161.3343511>
- Singapore: number of Facebook users 2023 | Statista. (2020). Statista Research Department. <https://www.statista.com/statistics/490492/number-of-singapore-facebook-users/>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Song, M., Park, H., & Shin, K. shik. (2019). Attention-based long short-term memory network using sentiment lexicon embedding for aspect-level sentiment analysis in Korean. *Information Processing and Management*, 56(3), 637–653. <https://doi.org/10.1016/j.ipm.2018.12.005>
- Suman, N., Gupta, P. K., & Sharma, P. (2018). Analysis of stock price flow based on social media sentiments. *Proceedings - 2017 International Conference on Next Generation Computing and Information Systems, ICNGCIS 2017*, i, 64–68. <https://doi.org/10.1109/ICNGCIS.2017.34>
- Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of depression-related posts in reddit social media forum. *IEEE Access*, 7, 44883–44893. <https://doi.org/10.1109/ACCESS.2019.2909180>
- Tao, X., Zhou, X., Zhang, J., & Yong, J. (2016). Sentiment analysis for depression detection on social networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10086 LNAI, 807–810. https://doi.org/10.1007/978-3-319-49586-6_59
- Timothy, J. (2019). *9 Depression Symptoms to Look Out For*. Healthline. <https://www.healthline.com/health/depression/recognizing-symptoms>
- Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in everyday life. *PLoS ONE*, 10(12). <https://doi.org/10.1371/JOURNAL.PONE.0145450>
- Uddin, A. H., Bapery, D., & Arif, A. S. M. (2019). Depression Analysis from Social Media Data in Bangla Language using Long Short Term Memory (LSTM) Recurrent Neural Network Technique. *5th International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering, IC4ME2 2019*, 1–4. <https://doi.org/10.1109/IC4ME247184.2019.9036528>
- Uddin, A. H., Bapery, D., & Mohammad Arif, A. S. (2019). Depression Analysis of Bangla Social Media Data using Gated Recurrent Neural Network. *1st International Conference on Advances in Science, Engineering and Robotics Technology 2019, ICASERT 2019, 2019(Icaser)*, 1–6. <https://doi.org/10.1109/ICASERT.2019.8934455>

- Uddin, S. B., Hossain, F., & Biswas, K. (2015). *Bangla to English Text Conversion Using OpenNLP Tools. January 2013.*
- Wang, X., Zhang, C., Ji, Y., Sun, L., Wu, L., & Bao, Z. (2013). A depression detection model based on sentiment analysis in micro-blog social network. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7867 LNAI, 201–213. https://doi.org/10.1007/978-3-642-40319-4_18
- Wong-Anuchit, C., Chantamit-o-pas, C., Schneider, J. K., & Mills, A. C. (2019). Motivational Interviewing–Based Compliance/Adherence Therapy Interventions to Improve Psychiatric Symptoms of People With Severe Mental Illness: Meta-Analysis. *Journal of the American Psychiatric Nurses Association*, 25(2), 122–133. <https://doi.org/10.1177/1078390318761790>
- Wu, S. (2019). *The Relationship Between Social Media Use and Depression in College Students : a Scoping Review.*
- Yuliyanti, S., Djatna, T., & Sukoco, H. (2017). Sentiment mining of community development program evaluation based on social media. *Telkomnika (Telecommunication Computing Electronics and Control)*, 15(4), 1858–1864. <https://doi.org/10.12928/TELKOMNIKA.v15i4.4633>