

DECLARATION OF RESEARCHER

I, Rakesh Dutta (Reg. No. : Ph.D/Comp. Sc.(1523)/694/R-2022), declare that this thesis titled, "Text Classification and Sentiment Analysis using Deep Learning techniques with Special Reference to COVID-19" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- The thesis is based on work done by myself with my supervisors Dr. Mukta Majumder.
- This thesis has not been previously submitted elsewhere for a degree or any other qualification at any University or any other institution.
- This has been always clearly attributed where I have consulted the published works of others.
- The sources have been always given where I have quoted from the works of others. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Rakesh Dutta

.....
Rakesh Dutta
Department of Computer Science & Technology
North Bengal University
April, 2024

CERTIFICATE OF SUPERVISOR

It is certified that the work contained in the thesis titled "Text Classification and Sentiment Analysis using Deep Learning techniques with Special Reference to COVID-19," by "Rakesh Dutta (Reg. No. : Ph.D/Comp. Sc.(1523)/694/R-2022)", has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

.....*Mukta Majumder*..... **SUPERVISOR**
Signature of Supervisor Department of Computer
Dr. Mukta Majumder Science & Technology
Assistant Professor University of North Bengal
Department of Computer Science & Technology
North Bengal University



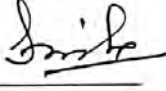
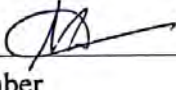
April, 2024

DRC CERTIFICATE

It is certified that the work contained in the thesis titled "Text Classification and Sentiment Analysis using Deep Learning techniques with Special Reference to COVID-19," by "Rakesh Dutta (Reg. No. : Ph.D/Comp. Sc.(1523)/694/R-2022)", has been carried out following University guidelines and has been approved by the Departmental research Committee for processing by the Board of Research Studies.

Approved:

Date:

<p><u>Dr. Ardhendu Mandal</u> </p> <p>Name and Signature of DRC Member</p>	<p><u>06/05/2024</u></p>
<p><u>Dr. Rakesh Kumar Mandal</u> </p> <p>Name and Signature of DRC Member</p>	<p><u>06/05/2024</u></p>
<p><u>Dr. Sharad Sinha</u> </p> <p>Name and Signature of DRC Member</p>	<p><u>6/5/2024</u></p>
<p><u>Dr. Swapnil Das</u> </p> <p>Name and Signature of DRC Member</p>	<p><u>06/05/2024</u></p>
<p><u>Dr. Munira Majumder</u> <u>M. Majumder</u></p> <p>Name and Signature of DRC Member</p>	<p><u>06-05-2024</u></p>
<p><u>Dr. JKM Sadique Wazman</u> <u>J. Wazman</u></p> <p>Name and Signature of DRC Member</p>	<p><u>06/5/2024</u></p>
<p><u>Dr. Md. Arif</u> <u>M. Arif</u></p> <p>Name and Signature of DRC Member</p>	<p><u>06.05.2024</u></p>
<p>_____</p> <p>Name and Signature of DRC Member</p>	<p>_____</p>

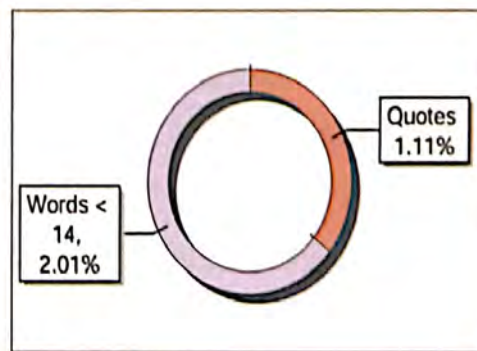
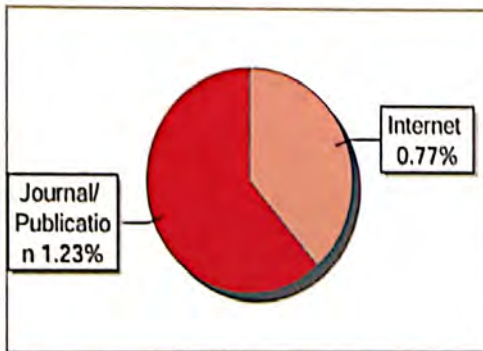
Department of Computer Science & Technology
North Bengal University

Submission Information

Author Name Rakesh Dutta
Title TEXT CLASSIFICATION AND SENTIMENT ANALYSIS USING DEEP LEARNING TECHNIQUES WITH SPECIAL REFERENCE TO COVID-19
Paper/Submission ID 1660837
Submitted by nbuplg@nbu.ac.in
Submission Date 2024-04-18 16:58:05
Total Pages 124
Document type Thesis

Result Information

Similarity 2 %



Exclude Information

Quotes Excluded
References/Bibliography Excluded
Sources: Less than 14 Words % Not Excluded
Excluded Source 2 %
Excluded Phrases Not Excluded

Database Selection

Language English
Student Papers Yes
Journals & publishers Yes
Internet or Web Yes
Institution Repository Yes

A Unique QR Code use to View/Download/Share Pdf File



Mukta Majumdar
06-05-24
SUPERVISOR
Department of Computer
Science & Technology
University of North Bengal

Rakesh Dutta
06/05/2024

ACKNOWLEDGEMENTS

First of all, I would like to thank my Ph.D. advisor Dr. Mukta Majumder for his guidance and unconditional support. His relentless encouragement and faith in my ability was the single most important reason for the completion of this work.

I would like to thank the members of my Doctoral Committee for giving me timely directions. I also thank all other faculty members of the department of Computer Science & Technology, North Bengal University for their kind supports and for providing me with all the facilities needed to carry out my research work. I am thankful to my thesis reviewers and other anonymous reviewers of my articles. I am thankful to Google Co-Lab for using GPU. I would like to thank all of the co-authors Dr. Biswapati Jana, Dr. Dibya Ranjan Das Adhikary, and Nilanjana Das for their support and suggestions.

The support which I have received from my friends and colleagues is invaluable. They extended their cooperation each time when I really needed it. I pay my gratitude to all the members of the Department of Computer Science and Application, Hijli College, Kharagpur. I am thankful to Dr. Tirthankar Dasgupta, Dr. Manjira Sinha, and Dr. Soumyadip Bandyopadhyay for their supports. There are a few more friends whose company I have enjoyed thoroughly during my PhD. It may not be possible to name all of them, however, I sincerely thank all of them for their support.

My parents have always been my greatest inspiration. The guidance of my parents, brothers, uncles and the seniors throughout my career have always been the most active support in the journey towards my dreams and aspirations. My deepest gratitude goes to my wife, Susmita De Dutta, for her tireless support that helped me through this journey. I would like to express my gratitude to my little daughter, Hoimasri Dutta, who always cheers me up. Finally, I thank my God for letting me through all the difficulties. "I will keep on trusting You for my future". Thank you, Lord.

This thesis is dedicated to my parents . . .

List of Tables

2.1	Name of COVID-19 Textual datasets with their applications	13
2.2	Description of COVID-19 Textual datasets.	14
3.1	Experimental outcomes on LitCovid dataset.	55
3.2	Comparison of the performance of our proposed A-BiLSTM model with the existing models on the LitCovid dataset.	56
4.1	Example of labels in the dataset.	65
4.2	Some case study on the rumor dataset.	68
4.3	Comparative result of proposed system with other state of art systems.	69
4.4	Comparative result of proposed system with other existing systems.	69
5.1	CI table	82
5.2	Consistency Ratio (CR) maximum thresholds for different pairwise comparison System.	82
5.3	Dependency relationship features are selected according to the dependency rules for sentiment analysis.	84
5.4	Dependency relationship features are selected according to the dependency rules for sentiment analysis.	87
5.5	Dependency relationship features are selected according to the dependency rules for sentiment analysis.	88
5.6	Comparison of our proposed model with LSTM, Bi-LSTM, and GRU.	94
6.1	The number of elements/criteria is represented by a random index (RI).	108
6.2	Assign the relative importance value of each factor and calculate normalized pair-wise comparison matrix.	109
6.3	Results of our proposed model versus other existing models.	118

List of Figures

3.1	Basic architecture of Deep Neural Network.	39
3.2	Basic architecture of RNN.	41
3.3	Demonstrate the four units of LSTM.	44
3.4	Illustration of a LSTM model.	45
3.5	Illustration of a BiLSTM model.	45
3.6	The architecture of A-BiLSTM.	46
3.7	Continuous Bag-of-Words (CBOW) model.	53
3.8	Skip-Gram model.	54
3.9	Comparative study of A-BiLSTM, Longformer, and BioBERT in terms of accuracy on validation dataset.	56
3.10	Comparison of A-BiLSTM, Longformer, and BioBERT in terms of loss on validation dataset.	57
4.1	BERT network model.	60
4.2	The architecture of BERT-ABiLSTM.	63
4.3	Number of veracity present in each class of the rumor dataset.	64
4.4	Comparison of the proposed model and BERT + LSTM (VAE) with ADASYN (Over-sampling) in terms of accuracy on validation dataset.	70
4.5	Comparison of the proposed model and BERT + LSTM (VAE) with ADASYN (Over-sampling) in terms of loss on validation dataset.	70
5.1	General block diagram of the proposed method.	74
5.2	Number of clusters obtained by Elbow scheme for Lockdown dataset.	77
5.3	Number of clusters obtained by the Elbow scheme for Unlock dataset.	77
5.4	Silhouette analysis for K-means clustering on Lockdown dataset.	78
5.5	Silhouette analysis for K-means clustering on Unlock dataset.	78
5.6	Flow diagram of semantic orientation approach for sentiment analysis.	83

5.7	Aspect wise people's sentiment analysis and compare our method with the existing model on the nationwide Lockdown Twitter dataset.	89
5.8	Aspect wise people's sentiment analysis and compare our method with the existing model on the Unlock Twitter dataset.	89
5.9	Illustration for the architecture of GRU.	91
5.10	Illustration for the architecture of Bi-GRU.	92
5.11	Comparison of Bi-GRU, GRU, Bi-LSTM, and LSTM in the form of accuracy on Lockdown dataset.	95
5.12	Comparison of Bi-GRU, GRU, Bi-LSTM, and LSTM in the form of accuracy on Unlock dataset.	95
5.13	Comparison of Bi-GRU, GRU, Bi-LSTM, and LSTM in the form of loss on Lockdown dataset.	96
5.14	Comparison of Bi-GRU, GRU, Bi-LSTM, and LSTM in the form of loss on Unlock dataset.	96
6.1	The block diagram of aspect terms identification.	100
6.2	Number of clusters generated by the Elbow technique for the Agree dataset.	103
6.3	Number of clusters generated by the Elbow technique for the Disagree dataset.	103
6.4	Number of clusters generated by the Elbow technique for the Neutral dataset.	103
6.5	Analysis of silhouettes score for K-Mediod clustering on Agree dataset.	104
6.6	Analysis of silhouettes score for K-Mediod clustering on Disagree dataset.	104
6.7	Analysis of silhouettes score for K-Mediod clustering on Neutral dataset.	104
6.8	TextRank graph built from a collection of tweets (text).	106
6.9	Number of tweets that agree on each aspect terms.	110
6.10	Number of tweets that disagree on each aspect terms.	110
6.11	Number of tweets that neutral on each aspect terms.	110
6.12	The block diagram of the proposed sentiment classification approach.	111
6.13	The architecture of the TRANS-Att-BiGRU.	112

6.14	Word vector representation input based on BERT model.	113
6.15	Comparison of TRANS-Att-BiGRU and BERT in the form of accuracy.	117
6.16	Comparison of TRANS-Att-BiGRU and BERT in the form of loss. .	117

LIST OF ABBREVIATIONS

SARS Severe Acute Respiratory Syndrome

COV-2 Coronavirus 2

NLP Natural Language Processing

CNN Convolutional neural network

RNN Recurrent neural network

LSTM Long short-term memory

BiLSTM Bi-directional long short-term memory

AM Attention mechanism

RL Reinforcement Learning

BIDAF Bi-Directional Attention Flow

POS Part-of-speech

SNLI Stanford Natural Language Inference

A-BiLSTM Attention based Bi-directional long short-term memory

CBOW Continuous bag-of-words

BOW Bag of words

TF-IDF Term Frequency–Inverse Document Frequency

BioBERT Bidirectional Encoder Representations from Transformers for Biomedical
Text Mining

AdamW Adam optimizer with a linear warmup

BERT Bidirectional Encoder Representations from Transformers

FN False negative

FP False positive

TP True positive

TN True negative

GDP Gross domestic product

GPU Graphics processing unit

ICMR Indian Council of Medical Research

LDA Latent Dirichlet allocation

RTX Ray Tracing Texel eXtreme

SSD Solid state drive

DL Deep Learning

BiGRU Bi-directional Gated Recurrent Unit

MCDM Multi-Criteria Decision Making

ABSA Aspect-Based Sentiment Analysis

NLTK Natural Language Toolkit

SVM Support Vector Machine

NB Naive Bayes

RF Random forest

LOC Lockdown

UNLOC Unlock

WCSS Within-Cluster Sum of Square

PMI Point-wise mutual information

BWM Best-worst method

CR Consistency Ratio

DS Dempster-Shafer Aggregation

W2V Word2vec

BERT-ABiLSTM Attention-based Bidirectional LSTM with BERT Embedding

SMOTE synthetic minority over-sampling method

DBSCAN Density-Based Spatial Clustering of Applications with Noise

DSMOTE Density-based synthetic minority oversampling

ADASYN Adaptive Synthetic Sampling Approach

VAE Variational AutoEncoders

Att BiLSTM Attention based Bi-directional long short-term memory

BERT-ABiGRU Transformer based Attention Bi-Gated Recurrent Unit

HEIs Higher Educational Institutions

OL Online Learning

GloVe Global vector

RAKE Rapid Automatic Keyword Extraction

AHP Analytical Hierarchy Process

CI Consistency Index

DT Decision Tree

KNN K-Nearest Neighbor

LR Logistic Regression

LIST OF SYMBOLS

\sim	Tilde
\otimes	Pair Wise Multiplication
σ	Sigma
\rightarrow	Forward
\leftarrow	Backward
$+$	Addition
$-$	Substraction
\times	Multiplication
Σ	Summation
$[[$	Third Parentheses
$()$	First Parentheses
\cup	Union
\oplus	Pair Wise Addition
$\ $	Mod
$/$	Division
$\%$	Percentage
\sim	Equivalent
$=$	Equals To
ϵ	Belongs To
$exp(.)$	Exponential Function
ϕ	Null Set
$-$	Complement
\leq	Less Than Equals To
$\{\}$	Second Parentheses
\cap	Intersection

\geq	Greater Than Equals To
\forall	For All
Π	Pi
$\hat{}$	Hat
$*$	Astrick
$\#$	Hash
β	Beta
$\sqrt{}$	Square Root
$<$	Less Than
ρ	Ratio of Degree to Frequency
λ	Lambda