

Analyzing the Sentiment of Social-Media for Predicting Depression using Supervised Learning and Radial Basis Function

Yogesh Sahu¹, Pinaki Ghosh²

¹Research Scholar, ²Head and Professor

^{1,2}School of Advanced Computing, Sanjeev Agrawal Global Educational University, Bhopal

Abstract– Sentiment analysis is a recent phenomenon that involves comprehending individuals' feelings across many situations in their everyday existence. The use of social media data, including textual information as well as emoticons, emojis, and other visual representations, will be employed throughout the whole of the process, including the analysis and classification operations. Previous study conducted several experiments utilising Binary and Triple Classification methods; however, it has been shown that multi-class classification offers more accurate and precise categorization. The data would be partitioned into many subcategories according to the polarity in a multi-class classification. Supervised Machine Learning Methods will be used throughout the categorization operation. Sentiment levels may be monitored or analysed using social media. This study investigates the use of artificial intelligence techniques to analyse sentiment in social media data with the purpose of understanding or detecting it. The poll used visual campaigns to analyse social media data, which included words, emoticons, and emojis, for the purpose of emotion detection using diverse machine learning methods. The SL-RBF Algorithm demonstrates higher accuracy in sentiment analysis.

Keywords– Sentiment Analyst, Radial Basis Function, Accuracy, Multi Class Classification, Precision.

I. INTRODUCTION

In the present day, a large proportion of people use social media and the Internet as platforms to disseminate their experiences and articulate their views and feelings. This often leads to extensive data transmission via the Internet. Nonetheless, a significant portion of this data has value for analysis. For instance, many companies and political campaigns rely on communication platforms to collect public sentiment and ascertain if it is impartial, positive, or negative.

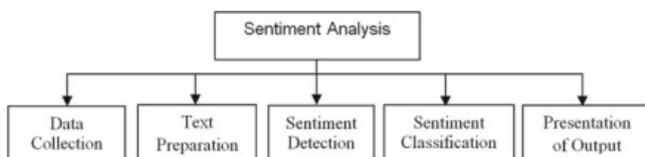


Fig.1- Sentiment analysis process steps

The extensive exchange of information on the Internet has resulted in the rise of the SA. Nasukawa [1] first proposed the notion of SA. The Sentiment Analysis (SA) is used in the process of Natural Language Processing (NLP) [2]. It examines the opinions, feelings, and reactions of people and writers on the internet, namely on social media and commercial websites, about a diverse array of products and services. Opinion mining, also known as sentiment analysis, refers to the broad field of research that scientists use to classify thoughts and attitudes into categories of neutrality, negativity, or positivity. Sentiment analysis (SA) is a text analysis technique often used in online reviews, polls, and social media postings. It monitors and analyses consumer feedback and responses on commercial websites to determine the acceptance or rejection of a product. This information assists the firm in boosting sales by uncovering client preferences. Due to the widespread presence of various perspectives on social networking platforms, policymakers, psychologists, researchers, manufacturers, and system developers generated novel concepts and critically evaluated them to arrive at optimal decisions. Sentiment analysis is a very effective technique for extracting and categorising sentiment information from a text using machine learning, statistics, and natural language processing (NLP).

II. LITERATURE REVIEW

Currently, sentiment analysis is regarded as one of the most advanced fields within the realm of natural language processing. Sentiment analysis determines the precise polarity of a text. Our objective is to use several machine learning classification analysis methods to determine if the Bengali text correlates to a state of happiness or sadness. We are collecting data for this assignment from various social media platforms, Bengali blogs, and other sources. By overcoming several obstacles, we successfully attained a favourable result. Sheikh Abujar, Abu Kaisar Mohammad Masum, Umme Sanzida Afroz, and Md. Rafidul Hasan Khan (2020)

Users are engaging in virtual socialism because to the rapid spread of social media, generating a substantial volume of textual and visual data. Users' online activity is manifested by their engagement with other posts, as seen by their likes and their sharing of content, such as tweets, status updates, and shared posts. Examining digital footprints to anticipate an individual's personality has evolved into a computationally intensive undertaking. Utilising user-generated textual content

inside a profile-based methodology might be advantageous in accurately portraying the social media identity. Hasan Mahmud, Hasan Al Marouf, and Md. Kamrul Hasan are individuals that are associated with the year 2020.

This study presents a profile-based approach for determining the authorship of Bangla songs written by two well-known writers and poets, Kazi Nazrul Islam and Rabindranath Tagore, via the use of supervised learning methods. The paper's problem statement is on the use of stylometric aspects to identify authorship to Bangla lyrics. The BanglaMusicStylo dataset was used, including 856 songs by Rabindranath Tagore and 620 songs by Kazi Nazrul Islam. The customary authorship attribution works in literature do not rely on Bangla song lyrics as a foundation; instead, they are based on the books written by the authors. Rafayet Hossian and Ahmed Al Marouf conducted the study in 2019.

Due to the actions of users sharing material, expressing viewpoints, disseminating news, capturing images, and recording events, social media has expanded into an extensive repository of words and images. It might be argued that publishing and tweeting status updates is a common feature of popular social networking platforms such as Facebook, Google+, Twitter, and others. User-generated textual information, such as tweets and status updates, plays a crucial role in interpersonal communication on social media platforms. Ahmed Al Marouf, Hasan Mahmud, and Md. Kamrul Hasan conducted a study in 2019.

Social media has expanded into a vast repository of words and images due to people sharing material, expressing views, disseminating news, capturing photographs, and recording events. One may argue that the act of sharing and posting status updates is a common feature of popular social networking platforms such as Facebook, Google+, Twitter, and others. The language used by individuals in their tweets and status updates plays a crucial role in interpersonal communication on social media platforms. Ahmed Al Marouf, Hasan Mahmud, and Md. Kamrul Hasan conducted a study in 2019.

The area of sentiment analysis has seen significant growth in the last decade as a result of the vast amount of opinionated data accessible on blogs and social networking platforms. The purpose of Natural Language Processing is to ascertain the presence of subjective information, such as positive or negative sentiments, in a given text or material. Social media and other online platforms provide a vast opportunity for swiftly uncovering human potential, allowing ordinary individuals to express their sentiments via comments that unequivocally showcase their embrace of others' potential. The authors of the publication are Tapasy Rabeya, Ahmed Al Marouf, Manoranjan Dash, Sanjida Ferdous, and Narayan Ranjan Chakraborty, and the publication was released in 2019.

Sentiment analysis is now a highly discussed topic that seeks to assist in extracting significant information from vast datasets. Its primary objective is to analyse and interpret the emotions derived from the patterns of the text. The system automatically classifies the manner in which feelings are conveyed, distinguishing between neutrality, positivity, and negativity towards any given subject. Data analysis may be conducted utilising many sources such as social media,

newspapers, medical publications, and movie reviews. We have collected review data for films and used five distinct machine learning classifiers to analyse the data. Md. Sharif Hossen and Atiqur Rahman (2019)

With few exceptions, females have a higher likelihood than males to suffer the prevalence, incidence, and morbidity risk of depressive disorders. These diseases often begin in mid-puberty and persist into adulthood. To examine possible risk variables that may contribute to gender disparities in depressive disorders. An exhaustive examination of the literature that investigates the genuine and artificial elements that independently contribute to gender differences in depressive diseases. Gender disparities exist in depressive diseases, however artificial variables may partially contribute to a higher prevalence among females. Greg Wilkinson and Marco Piccinelli conducted the study in 2018.

This article provides a concise summary of eRisk 2018. This lab was established by CLEF for the second consecutive year. The main objective of eRisk was to analyse assessment techniques, efficacy indicators, and other methods for early risk identification. Early detection systems have several uses, particularly in the field of health and safety. The second version of eRisk included two tasks: one focused on the early diagnosis of anorexia risk, and the other on the early detection of depression risk. (Fabio Crestani, Javier Parapar, and David E. Losada; 2018)

III. PROBLEM IDENTIFICATION

The following are the issue identifications based on the previous work:

An unrelated depression detection occurs as a result of insufficient accuracy and recall.

The precision of depression identification is compromised by poor accuracy and F1-Score.

The time it takes to identify depression is prolonged owing to the elevated mistake rate.

IV. RESEARCH OBJECTIVES

The proposed effort aims to achieve the following objectives:

- The objective is to enhance the accuracy and completeness of identifying depression cases that are connected.
- To enhance the precision and F1-Score in order to increase the accuracy of depression identification.
- The objective is to minimise the error rate in detecting depression.

V. METHODOLOGY

The algorithm of the suggested technique. The SVM-RBF (Support Vector Machine - Radial Basis Function) may be described as follows:

Data set preparation involves the selection of any form of data, which may be obtained either from various sources or downloaded from the internet. The greater the amount of data, the higher the accuracy of the forecast.

Prediction Models	TN	FP	FN	TP
Naïve Bayes	39.64	10.35	9.47	40.55
Linear SVM	40.47	9.51	8.64	41.37
SL-RBF (Proposed)	40.94	9.05	8.11	41.91

Table 2- Estimation of Precision, Recall, F1-Score and Accuracy among different models and SL-RBF (Proposed Prediction Model)

Prediction Models	Precision	Recall	F1-Score	Accuracy
Naïve Bayes	0.79	0.81	0.8	80.1
Linear SVM	0.81	0.83	0.82	81.2%
SL-RBF (Proposed)	0.83	0.84	0.84	83.43%

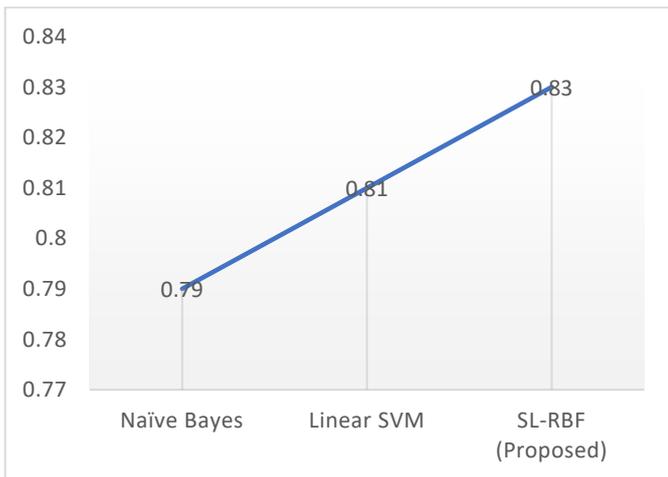


Fig.5- Graphical Analysis of Precision among different models and SL-RBF (Proposed Prediction Model)

The graph (Fig.5) shows that the proposed model gives better precision for depression prediction as compare than other models. The precision of SL-RBF is improved by 0.02 as compare than Linear SVM model.

The graph (Fig.6) shows that the proposed model gives better recall for depression prediction as compare than other models. The recall of SL-RBF is improved by 0.01 as compare than Linear SVM.

The graph (Fig.7) shows that the proposed model gives better F1-Score for depression prediction as compare than other models. The F1-Score of SL-RBF is improved by 0.02 as compare than Linear SVM.

The graph (Fig.8) shows that the proposed model gives better Accuracy for depression prediction as compare than other models. The Accuracy of SL-RBF is improved by 2.23% as compare than Linear SVM model.

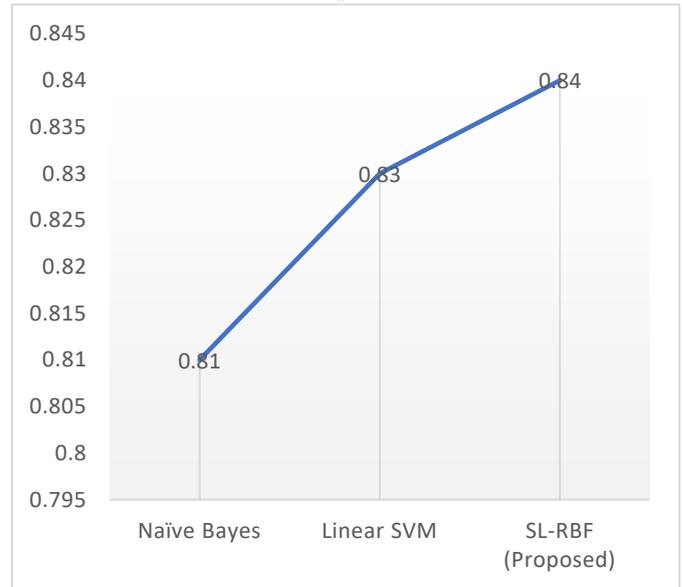


Fig.6- Graphical Analysis of Recall among different models and SL-RBF (Proposed Prediction Model)

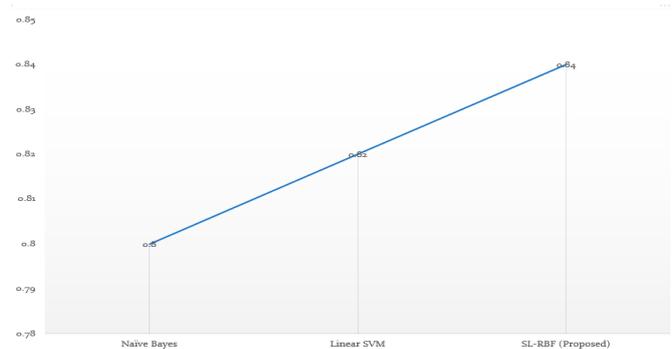


Fig.7- Graphical Analysis of F1-Score among different models and SL-RBF (Proposed Prediction Model)

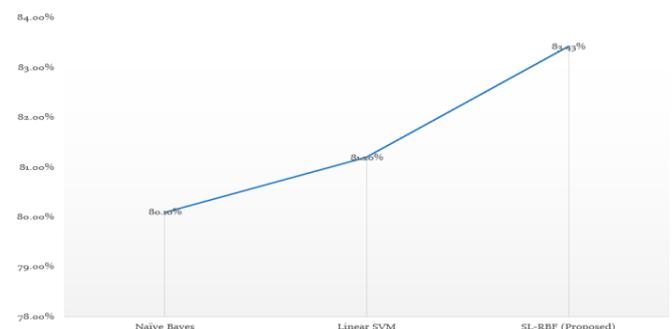


Fig.8- Graphical Analysis of Accuracy among different models and SL-RBF (Proposed Prediction Model)

VIII. CONCLUSIONS

The findings of this study may be summarized as follows:

1. The suggested model exhibits superior prediction accuracy compared to Linear SVM. The precision increases by 2.23%.
2. The suggested model exhibits superior prediction accuracy compared to Linear SVM. The accuracy increases by 0.02.
3. The suggested model exhibits superior prediction recall compared to Linear SVM. The recall increases by 0.01.

4. The suggested model outperforms Linear SVM in terms of prediction F1-Score. The F1-Score increases by 0.02.

Therefore, the prediction of depression in software is more accurately achieved by the suggested approach SL-RBF (Supervised Learning with Radial Basis Function).

IX. FUTURE RECOMMENDATION

Our suggested technique enhances the precision of depression prediction and is very beneficial for future enhancements. For future improvements, it is necessary to evaluate the accuracy using various datasets and use other AI methods to verify the accuracy estimate. The proposed model is limited by its processing time due to the large volume of data required to estimate the performance of the training data. In the future, the same algorithms will be used to real-time data from platforms such as Instagram, Facebook, LinkedIn, etc., in order to assess the efficiency of the system.

REFERENCES

- [1] Jitimon Angskun, Suda Tipprasert and Thara Angskun, "Big data analytics on social networks for real-time depression detection", *Journal of Big Data*, 2022.
- [2] Md. Rafidul Hasan Khan; Umme Sunzida Afroz; Abu Kaisar Mohammad Masum; Sheikh Abujar; Syed Akhter Hossain, "Sentiment Analysis from Bengali Depression Dataset using Machine Learning", 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2020.
- [3] Tanna D, Dudhane M, Sardar A. Deshpande K, Deshmukh N., "Sentiment analysis on social media for emotion classification", In: International Conference on intelligent computing and control systems, 2020.
- [4] Md Kamrul Hasan, Hasan Mahmud, Ahmed Al Marouf, "Comparative Analysis of Feature Selection Algorithms for Computational Personality Prediction From Social Media", *IEEE Transactions on Computational Social Systems*, 2020.
- [5] Yang X, McEwen R, Ong LR, Zihayat M., "A big data analytics framework for detecting user-level depression from social networks", *Int J Inf Manag.* 2020.
- [6] Lyua YW, Chow JC-C, Hwang J-J, "Exploring public attitudes of child abuse in mainland China: a sentiment analysis of China's social media", *Weibo. Child Youth Serv Rev.* 2020.
- [7] Chen B, Cheng L, Chen R, Huang Q, Phoebe Chen Y-P. Deep neural networks for multiclass sentiment classification. In: IEEE 20th International Conference on high performance computing and communications, IEEE 16th International Conference on Smart City, IEEE 4th International Conference on Data Science and Systems 2018; pp. 854–59.
- [8] Sethi M, Pande S, Trar P, Soni P. Sentiment identification in COVID-19 specific tweets. In: International Conference on electronics and sustainable communication systems (ICESC 2020), pp. 509–16, <https://doi.org/10.1109/ICESC48915.2020.9155674>.
- [9] Kundale JU, Kulkarni NJ. Language independent multi-class sentiment analysis. In: 5th International Conference on computing communication control and automation (ICCUBEA), 2019; pp. 1–7, <https://doi.org/10.1109/ICCUBEA47591.2019.9128383>.
- [10] Ruz GA, Henriquez PA, Mascareno A. Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. *Future Gener Comput Syst.* 2020; 106:92–104.
- [11] Yang X, McEwen R, Ong LR, Zihayat M. A big data analytics framework for detecting user-level depression from social networks. *Int J Inf Manag.* 2020; 54:102141.
- [12] Tao X, Dharmalingam R, Zhang J, Zhou X, Li L, Gururajan R. Twitter analysis for depression on social networks based on sentiment and stress. In: 6th International Conference on behavioral, economic and socio-cultural computing, 2019; pp. 1-4, <https://doi.org/10.1109/BESC48373.2019.8963550>.
- [13] Tanna D, Dudhane M, Sardar A. Deshpande K, Deshmukh N. Sentiment analysis on social media for emotion classification. In: International Conference on intelligent computing and control systems (ICICCS 2020), pp. 911–15, <https://doi.org/10.1109/ICICCS48265.2020.9121057>.
- [14] Arora P, Arora P. Mining Twitter data for depression detection. In: IEEE International Conference on signal processing and communication (ICSC), 2019; pp. 186–89, <https://doi.org/10.1109/ICSC45622.2019.8938353>.
- [15] Chen Y, Zhou B, Zhang W, Gong W, Sun G. Sentiment analysis based on deep learning and its application in screening for perinatal depression. In: IEEE Third International Conference on data science in cyberspace. 2018; pp. 451–6. <https://doi.org/10.1109/DSC.2018.00073>.
- [16] Uddin AH, Bapery D, Arif ASM. Depression analysis from social media data in Bangla language using long short term memory (LSTM) recurrent neural network technique. In: International Conference on computer, communication, chemical, materials and electronic engineering (IC4ME2), 11–12 July, 2019; pp. 1-4, <https://doi.org/10.1109/IC4ME247184.2019.9036528>.
- [17] Cheng L-C, Tsai S-L. Deep learning for automated sentiment analysis of social media. In: IEEE/ACM International Conference on advances in social networks analysis and mining. 2019; pp. 1001–4. <https://doi.org/10.1145/3341161.3344821>.
- [18] Al Asad N, Pranto MAM, Afreen S, Islam MM. Depression detection by analyzing social media posts of user. In: IEEE International Conference on signal processing, information, communication & systems (SPICSCON). Dhaka, Bangladesh, 2019; pp. 13–17, <https://doi.org/10.1109/SPICSCON48833.2019.9065101>.
- [19] Lyua YW, Chow JC-C, Hwang J-J. Exploring public attitudes of child abuse in mainland China: a sentiment analysis of China's social media Weibo. *Child Youth Serv Rev.* 2020; 116:102520.
- [20] Abid F, Li C, Alam M. Multi-source social media data sentiment analysis using bidirectional recurrent convolutional neural networks. *Comput Commun.* 2020; 157:102–15.
- [21] Hammou BA, Lahcen AA, Mouline S. Towards a real-time processing framework based on improved distributed recurrent neural network variants with fastText for social big data analytics. *Inf Process Manag.* 2020; 57:102122.
- [22] Tadessi MM, Lin H, Xu B, Yang L. Detection of depression-related posts in reddit social media forum. *IEEE Access.* 2019; 7:44883–93. <https://doi.org/10.1109/ACCESS.2019.2909180>.
- [23] Trotzek M, Koitka S, Friedrich CM. Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. *IEEE Trans Knowl Data Eng.* 2018; 32:588–601.
- [24] Tariq S, Akhtar N, Afzal H, Khalid S, Mufti MR, Hussain S, Habib A, Ahmad G. A novel co-training-based approach for the classification of mental illnesses using social media posts. *IEEE Access.* 2019; 7:166165–72. <https://doi.org/10.1109/ACCESS.2019.2953087>.
- [25] Rao G, Zhang Y, Zhang L, Cong Q, Feng Z. MGL-CNN: a hierarchical posts representations model for identifying depressed individuals in online forums. *IEEE Access.* 2020; 8:32395–403. <https://doi.org/10.1109/ACCESS.2020.297373>.