International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 7 Issue 6, November-December 2023 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Analysis of Social Media Sentiment for Depression Prediction using Supervised Learning and Radial Basis Function

Yogesh Sahu[1], Dr. Pinaki Ghosh[2]

[1]Research Scholar, [2]Head and Professor,
[1, 2]SSOAC, SAGE University, Bhopal, Madhya Pradesh, India

ABSTRACT

Sentiment analysis is a new trend in understanding people's emotions in a variety of scenarios in their daily lives. Social media data, which includes text data as well as emoticons, emojis, and other images, would be used throughout the process, including the analysis and categorization procedures. Numerous trials were carried out in previous research using Binary and Triple Classification, however multi-class classification provides more exact and precise classification. The data would be separated into many sub-classes based on the polarity in multi-class classification. During the categorization procedure, Supervised Machine Learning Methods would be used. Sentiment levels may be tracked or studied via social media. This work examines sentiment analysis on communal media data for apprehension or detection using various artificial intelligence approaches. In the poll, it was visually campaigned that social media data, which included words, emoticons, and emojis, was used for sentiment recognition using various machine learning approaches. For sentiment analysis, the Supervised Learning with Radial Basis Function (SL-RBF) Algorithm has a greater precision value.

KEYWORDS: *Sentiment Analysist, Radial Basis Function, Accuracy, Multi Class Classification, Precision*

How to cite this paper: Yogesh Sahu | Dr. Pinaki Ghosh "Analysis of Social Media Sentiment for Depression Prediction using Supervised Learning and Radial Basis Function" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-7 | Issue-6, December 2023, pp.259-267, URL: www.ijtsrd.com/papers/ijtsrd60158.pdf

IJTSRD60158

1. INTRODUCTION

Nowadays, the majority of individuals use social media and the Internet to share their experiences and to express their thoughts and emotions. This often results in huge data communication across the Internet. However, the majority of this data may be usefully examined; for instance, the majority of businesses and political campaigns depend on communication sites to gather public opinion and determine if it is neutral, favorable, or unfavorable.

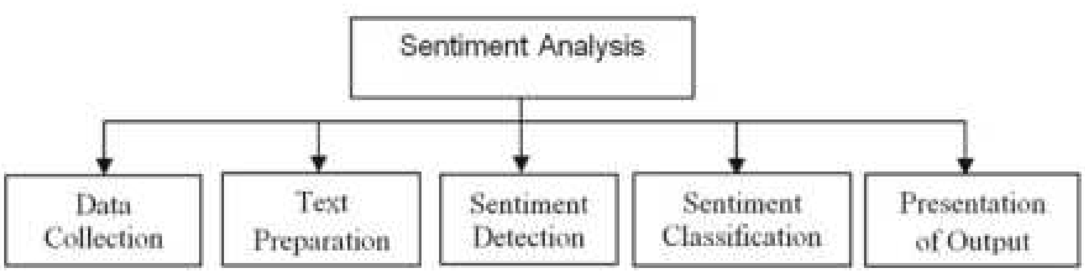

Figure 1: Sentiment analysis process steps

The massive information interchange on the Internet has led to the emergence of the SA. It was Nasukawa [1] who initially put up the concept of SA. First, the SA is used in the natural language processing (NLP) [2] process, which analyzes the thoughts, emotions, and responses of individuals and authors on the Internet through social media and commercial websites on a wide range of goods and services. Opinion mining, another name for sentiment analysis, is another term for the large area of study that many scholars use to categorize thoughts and attitudes as neutral, negative, or positive. SA is a textual research that is often utilized in online reviews and surveys as well as social media posts. It manages consumer comments and reactions on commercial websites to find out if a product is accepted or rejected; this helps the business increase sales since it reveals the customer's preference. As a result of the proliferation of diverse viewpoints on social networking sites, policymakers, psychologists, researchers, manufacturers, and system developers came up with fresh ideas and analyzed them to make the best judgments possible. Sentiment analysis is a very efficient method of extracting and defining sentiment information from a text unit utilizing machine learning, statistics, and natural language processing (NLP).

2. Related Work

Sentiment analysis is one of the most sophisticated areas of natural language processing these days. Sentiment analysis identifies a paragraph's specific polarity. Our goal is to use several kinds of machine learning classification analysis techniques to identify if the Bengali text corresponds to a joyful or sad feeling. We are gathering information for this task from many social media platforms, Bengali blogs, etc. Through several challenges, we were able to achieve a satisfactory outcome. (Sheikh Abujar, Abu Kaisar Mohammad Masum, Umme Sanzida Afroz, and Md. Rafidul Hasan Khan; 2020)

Users are participating in virtual socialism as a result of social media's explosive expansion, producing a vast amount of text and visual information. Users' online activity is shown in their like of other postings, which is reflected in their sharing of material, including tweets and status updates, and shared posts. Analyzing digital traces to predict a user's personality has become a computationally demanding task. Using user-generated textual material in a profile-based approach might be helpful in reflecting the social media personality. (Hasan Mahmud; Hasan Al Marouf; Md. Kamrul Hasan; 2020)

This work proposes a profile-based method to identify the lyricist of Bangla songs composed by two renowned novelists and poets, Kazi Nazrul Islam and Rabindranath Tagore, using supervised learning techniques. The use of stylometric elements to attribute authorship to Bangla lyrics might be regarded as the paper's issue statement. We have used the BanglaMusicStylo dataset, which includes 856 songs by Rabindranath Tagore and 620 songs by Kazi Nazrul Islam. Bangla song lyrics are not the basis for the conventional authorship attribution works found in literature; rather, they are based on the books authored by the writers. (Rafayet Hossian and Ahmed Al Marouf, 2019)

As a result of individuals posting content, exchanging opinions and news, taking pictures, and documenting events, social media has grown into a massive word and image archive. One may argue that sharing and tweeting status updates is a frequent function of well-known social networking sites like Facebook, Google+, Twitter, and so forth. User-generated textual content, like tweets and status updates, may be the most important language for interpersonal communication on social media. (Ahmed Al Marouf; Hasan Mahmud; Md. Kamrul Hasan; 2019)

As a result of individuals posting content, exchanging opinions and news, taking pictures, and documenting events, social media has grown into a massive word and image archive. One may argue that sharing and tweeting status updates is a frequent function of well-known social networking sites like Facebook, Google+, Twitter, and so forth. User-generated textual content, like tweets and status updates, may be the most important language for interpersonal communication on social media. (Ahmed Al Marouf; Hasan Mahmud; Md. Kamrul Hasan; 2019)

Sentiment analysis has been a thriving experimental research field in the past ten years due to the abundance of opinionated data available on blogs and social networking sites. The assignment of Natural Language Processing is cited in order to determine whether or whether text or material includes any subjective information, such as positive or negative information. Social media and other internet platforms are providing a huge platform for quickly discovering human potential, and regular people may express their feelings by making comments that clearly demonstrate how they are welcoming of others' potential. (Tapasy Rabeya, Ahmed Al Marouf, Manoranjan Dash, Sanjida Ferdous, and Narayan Ranjan Chakraborty, 2019)

Sentiment analysis is now the most talked about subject that aims to help extract meaningful information from enormous datasets. It focuses on analyzing and deciphering the emotions from the text patterns. It automatically categorizes how sentiments are expressed, such as whether they are neutral, positive, or negative about anything existing. Data analysis may be done using a variety of sources, including social media, newspapers, medical journals, and movie reviews. Here, we've gathered review data for movies and used five different types of machine learning classifiers to examine the information. (Md. Sharif Hossen; Atiqur Rahman; 2019)

With very few exceptions, girls are more likely than men to experience the prevalence, incidence, and morbidity risk of depressive disorders, which start in mid-puberty and last into adulthood. to go through potential risk factors that might cause gender variations in depressive illnesses. a rigorous analysis of the literature that addresses the real and artifactual factors that contribute to gender disparities in depressive illnesses separately. There are true gender differences in depressive illnesses, even if artifactual factors may somewhat contribute to a female predominance. (Greg Wilkinson and Marco Piccinelli, 2018)

In this publication, eRisk 2018 is summarized. This was the second year that CLEF had set up this lab. Examining assessment technique, efficacy metrics, and other early risk detection procedures was the primary goal of eRisk. Early detection technologies have several applications, especially in the health and safety domain. Two tasks were included in the second version of eRisk: one on early risk detection of anorexia and the other on early risk detection of depression. (Fabio Crestani, Javier Parapar, and David E. Losada; 2018)

3. Problem Identification

Following are the problem identification on the basis of existing work:

- An unrelated depression detection generates due to low precision and recall.
- The exactness of depression detection is low due to low accuracy and F1- Score.
- The depression detection time is high due to the high error rate.

4. Research Objectives

Following are the objectives of the proposed work:

- To improve precision and recall for related depression detection.
- To improve accuracy and F1- Score for improving exactness of depression detection.
- To reduce the error rate for depression detection.

5. Methodology

The Algorithm of proposed methodology SVM-RBF (Support Vector Machine - Radial Basis Function) is as follows

Preparation of Data set- one can take any type of data or can download from net also. More the data more will be accuracy of the prediction.

Data pre-processing- In this step we make the words simpler so that the prediction becomes easy. Some common data pre-processing methods are- tokenization (dividing into each word), lemmatization, stemming and removing stop words (unwanted words) and characters.

Feature extraction-For all classification algorithms, features are necessary to either plot or make a precise detail so that the predictions are based on that feature. here we will use ICA algorithm

Classifier algorithms- Here we use SVM (Support Vector Machine) with RBF function

Prediction- Once all the above steps are done the model is ready to do the predictions. We will do the predictions on the testing dataset.

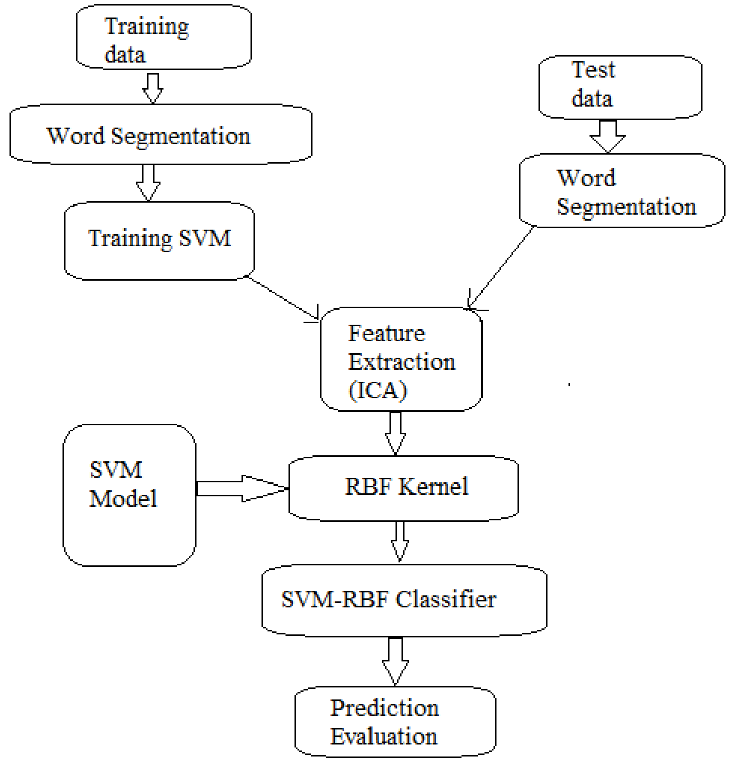

Figure 2: Proposed model for depression detection from social media

6. Experimental Setup

This section outlines the particular procedures of the experiment after outlining various presumptions and constraints. The following are the presumptions made in this work:

1. The training and testing data come from a single dataset, and our study focuses on sentiment analysis of depression identification from social media. When using the Twitter dataset for experimentation, for instance, the training set is chosen and the test set is created from the remaining portion of the dataset.
2. The trained model favors the non-faulty classes during the tests because of a limited number of defective classes. Thus, before training the model, class imbalance is applied to the whole dataset.
3. A 3-fold cross-validation is used in order to more accurately measure the algorithm's performance.

The validity of the sentiment analysis of depression identification from social media framework described in this study may be validated under the aforementioned assumptions. The particular protocol for the experiment is as follows:

Step 1: The class dependency of the depression tweets is extracted using the code analysis tool, and a CSV file is subsequently created.

Step 2: From the Kaggle dataset (https://www.kaggle.com/datasets/ferno2/training1600000processednoemoticoncsv), the labeled nodes and feature metrics are extracted.

Step 3: To address the imbalance in data classes, the SL-RBF (Supervised Learning with Radial Basis Function) technique is used.

7. Results and Analysis

The following observations are performed on anaconda navigator with python 3.11.1 with jupyter lab toolbox. The proposed procedure SL-RBF (Logistic Regression with Radial Basis Function) perform on (Kaggle Repository) training.1600000.processed.noemoticon.csv dataset and calculate precision, recall, F1-Score and accuracy parameters are calculated as follows:

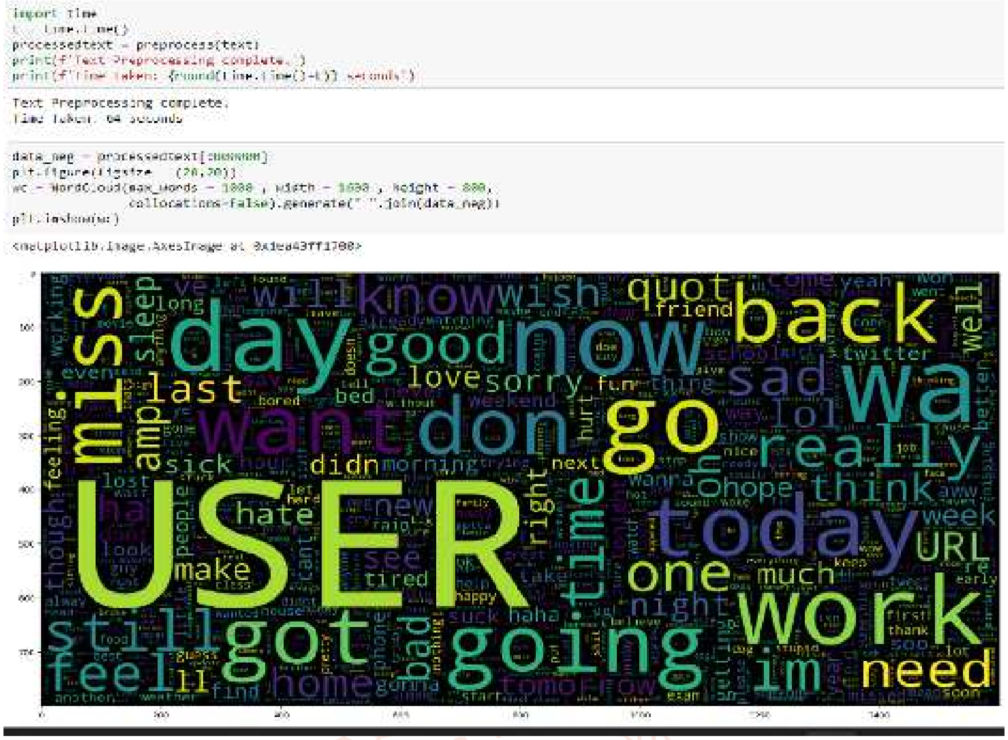

Figure 3: Text Preprocessing Process

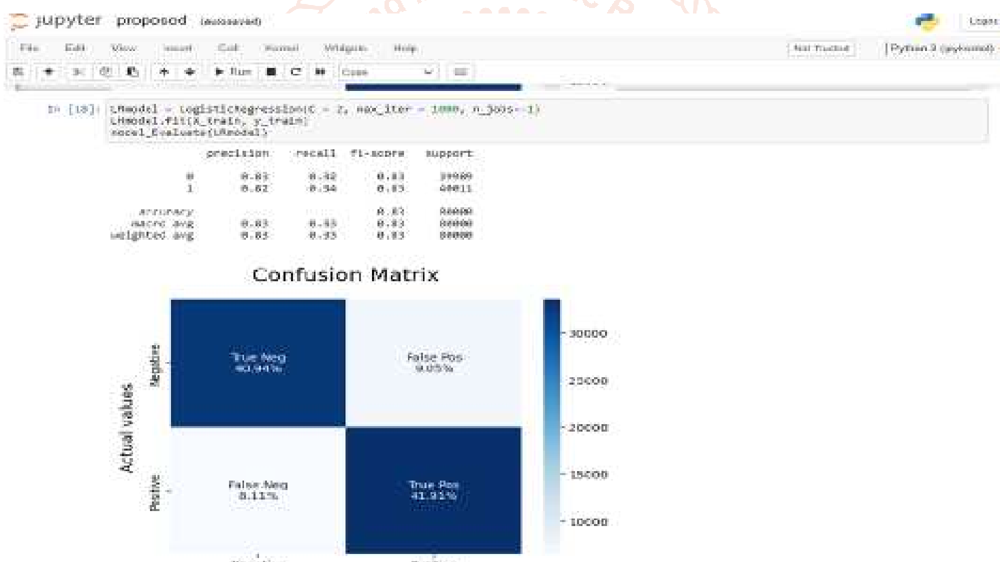

Figure 4: Evaluation of Confusion Matrix SL-RBF (Proposed Prediction Model)

Table 1: Estimation of Confusion Matrix among different models and SL-RBF (Proposed Prediction Model)

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%**

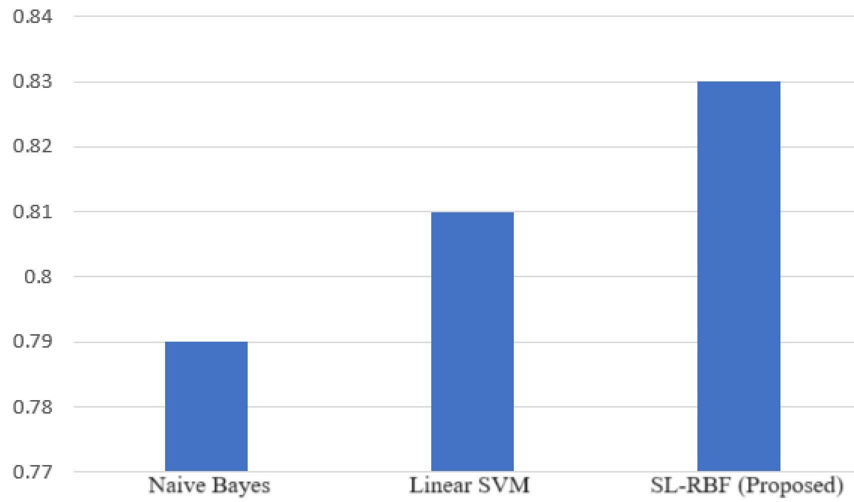

Figure 5: Graphical Analysis of Precision among different models and SL-RBF (Proposed Prediction Model)

The above graph show 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.

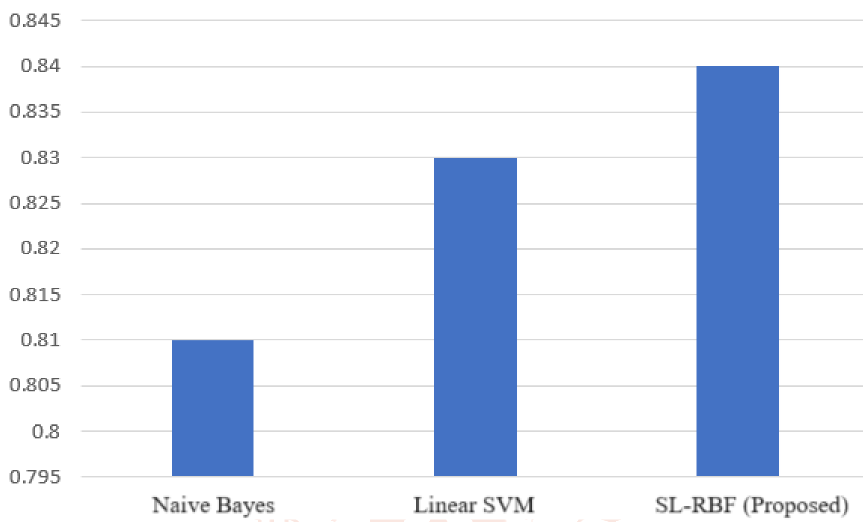

Figure 6: Graphical Analysis of Recall among different models and SL-RBF (Proposed Prediction Model)

The above graph show 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.

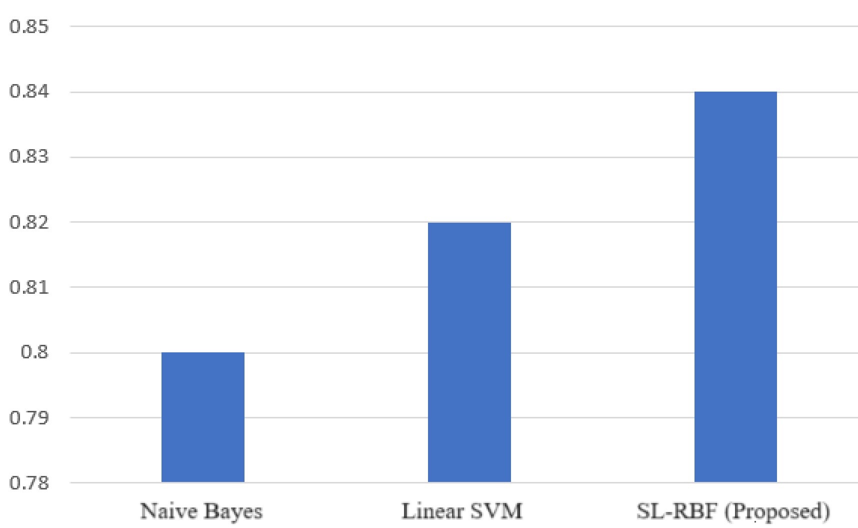

Figure 7: Graphical Analysis of F1-Score among different models and SL-RBF (Proposed Prediction Model)

The above graph show 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.

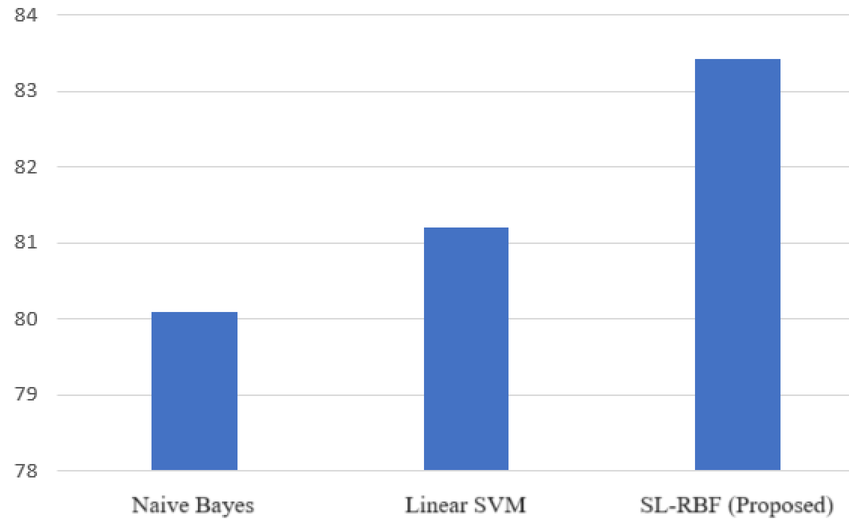

Figure 5.6: Graphical Analysis of Accuracy among different models and SL-RBF (Proposed Prediction Model)

The above graph show 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.

8. Conclusions

The conclusions of this work are as follows:

1. The proposed model gives better prediction accuracy as compare than Linear SVM. The accuracy improves by 2.23%.
2. The proposed model gives better prediction precision as compare than Linear SVM. The precision improves by 0.02.
3. The proposed model gives better prediction recall as compare than Linear SVM. The recall improves by 0.01.
4. The proposed model gives better prediction F1-Score as compare than Linear SVM. The F1-Score improves by 0.02.

Hence, depression prediction in software is better predict through proposed method SL-RBF (Supervised Learning with Radial Basis Function).

9. Future Recommendation

Our proposed methodology helps to improve the accuracy of depression prediction and greatly helpful for further improvement. In future enhancements, the accuracy has to be tested with different dataset and to apply other AI algorithms to check the accuracy estimation. The limitation of the proposed model is processing time, because of huge amount of data taken for estimating the performance of train data. In future, the same algorithms to be implemented with real-time data (like instgram, facebook, linkedIn etc) for estimating the effectiveness 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 classifers. 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 classifcation. 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 analy sis 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 Confer ence 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 senti ment 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 recur rent 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 classifcation 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.