Stress Detection Using Bernoulli Naive Bayes Machine Learning Technique.

B.Akhil Bharadwaz[1]
Department of Electronics and Communication Engineering,
Presidency University,
Bangalore, India.
akhilbharadwazbrahmadevi@gmail.com

B.Sasidhar[2]
Department of Electronics and Communication Engineering,
Presidency University,
Bangalore, India.
bonganisasidhar08@gmail.com

Ms. Sapna R[3]
Assistant professor, Department of Computer Science and Engineering,
Presidency University,
Bangalore, India.
sapna.aradhya@gmail.com

***Abstract*—The lives of people are increasingly being threatened by psychological issues. Before stress develops into a serious health problem, it is critical to recognize and manage it. Social Networking has become a necessary component of our daily lives. On social media sites, people frequently communicate their feelings and emotions, including stress. Social media makes it simple to identify and reduce stressors for both individuals and organizations. Based on user social behavior, stress may be easily identified in users. Also, traditional methods for stress detection are expensive and time-consuming. This study suggests a machine learning method for detecting stress in social media posts using the Bernoulli Naive Bayes algorithm. On a dataset of 2,839 social media posts that includes posts from WhatsApp, Facebook, and Twitter, the proposed approach is assessed, and it yields promising results. We have developed a graphical user interface for the system using Streamlit, which is an open-source app framework that helps us quickly deploy the development into an interactive GUI. This GUI provides the option for the user to enter text. The model determines whether the user is stressed or not based on the input from the user.**

Keywords—Bernoulli Naive Bayes Algorithm, CountVectorizer, Stemming, Stress detector, Streamlit.

I. INTRODUCTION

Stress is a common emotion experienced by individuals, and social media has become an essential platform for individuals to express their thoughts and feelings. Stress detection on social networking sites can offer important insights into a person's mental health and aid in preventing and managing stress-related disorders. Machine learning algorithms can be employed to detect stress in social media posts [3]. In this research paper, we propose the use of the Bernoulli Naive Bayes algorithm for stress detection on social media posts [3]. [11] The Naive Bayes algorithm is a simple yet powerful classification algorithm that can be used to classify texts based on their content. The proposed approach involves collecting a dataset of social media posts containing stress-related keywords and applying pre-processing techniques such as stop word removal, stemming, and punctuation removal [11]. The Naive Bayes algorithm is then trained on the pre-processed dataset, and the resulting model is used to classify new social media posts as stressed or not stressed. The proposed approach has the potential to aid in the early detection of stress in social media posts, which can aid in the prevention and management of stress-related disorders. This research paper presents the details of the proposed approach, including the dataset, pre-processing techniques, methodology, evaluation metrics, and results [10]. The findings of this research paper demonstrate that the Naive Bayes algorithm can be an effective approach for stress detection on social media posts.

II. RELATED WORK

In order to detect stress in social media messages, several researchers have studied the use of machine learning algorithms. In one study, researchers used the Support Vector Machine (SVM) approach to categorize tweets as stressful or non-stressful. The study has a 79.1% accuracy rate using a dataset of 2,000 tweets [1]. Using deep learning techniques based on convolutional neural networks (CNNs), experts in a different study classified tweets as stressed or not worried. 80.8% accuracy was observed in the study, which examined a dataset of 3,000 tweets [2]. Some of the other studies can be summarized as follows:

Reshma Radheshamjee et al. [3] proposed a social media post-based stress identification method. The datasets for this research were acquired from social media. Nowadays, people who are upset or depressed will post words or other images on Facebook, Twitter, or any other social media platform. Twitter is used by a huge number of individuals to post comments and other postings using quotes and other terms. They have presented a method that will gather data based on the Twitter dataset using support vector machines and Nave Bayes techniques. The algorithm that predicts the detection suggests that the person could be stressed out or sad. Stress, depression, and other factors are taken into consideration in the Twitter dataset. Stress and depression have been categorized using emotional analysis. We may acquire the greatest outcomes and also see the precision and recall levels when they apply more strategies or algorithms. Stress and depression have been predicted using the confusion matrix.

In order to identify human stress, Alana Paul Cruz et al. [4] suggested a machine-learning approach based on human biosignals. ECG was chosen as the biosignal, and its characteristics were retrieved. It is advantageous to use ECG as the biosignal since it is simple to determine respiratory signals and EDR properties without the use of additional sensors. They used an Optimised SVM algorithm, which combines the SVM and Decision Tree algorithms, to develop this model. They used the Physionet's database to train and validate their model. This model is 96.3% accurate.

A technique for stress detection utilizing deep neural networks was proposed by Russell Li and Zhandong Liu [5]. The need for manually created features is a drawback of conventional machine learning techniques. If features are recognized incorrectly, accuracy declines. They created a 1-dimensional (1D) convolutional neural network and a multilayer perceptron neural network to overcome this shortcoming. The deep convolutional neural network achieved accuracy rates of 99.80% and 99.55%, respectively, for binary and 3-class categorization. The deep multilayer perceptron neural network successfully classified objects into binary and three classes with accuracy rates of 99.65% and 98.38%, respectively. The only limitation of this project is that the datasets utilized were gathered from 15 human volunteers, which might not be a good enough representation of the entire human population.

A method that uses physiological devices to detect stress was proposed by Madhavi Ganapathiraju et al [6]. We can obtain accurate values by using physiological instruments. When we look at somebody from the exterior, we cannot tell whether they are healthy or not. Because of this issue, we can quickly determine whether a person is ill or under stress. This physiological apparatus makes use of a motion sensor, body temperature, pulse rate, and galvanic skin response. When we utilize a physiological device, it records values depending on those values, and when those values change, it is simple to determine if someone is experiencing issues.

A system for driver stress detection was proposed by Wan-young Chung et al [7]. We hear about car accidents in the news so frequently these days. As a result of the collisions, the driver will pass away. We don't understand what the issue is with drivers dying in collisions. The likelihood of an accident increases when the driver is fatigued or experiencing some other issue. The suggested system will gather data and should store that data in a database since it uses physiological devices including motion sensors, galvanic skin reaction sensors, and body temperature sensors. Based on these values, we may predict that the driver is having certain problems and can warn him or her when necessary. Based on the information, we can warn the driver by sounding a bell.

Jia Jia [8] suggested an automated approach for stress detection that makes use of cross-media microblog data. A three-level structure is built to design the issue. The tweets yield a number of low-level properties. After then, middle-level representations are developed and extracted in accordance with psychological and aesthetic ideas, including the linguistic qualities of texts, the visual qualities of images, and social characteristics from the comments. In order to learn the many types of stress, a Deep Sparse Neural Network is created. The suggested strategy works well and effectively to identify psychological stress in microblog data.

Huijie Lin [9] found that a user's stress level is closely correlated with that of their friends on social media. To examine this link, Huijie Lin employed a sizable dataset from real-world social platforms. The text of tweets and data on social interactions are then suggested to be used in a factor graph model with a convolutional neural network for stress identification. Textual, visual, and social aspects that connect to stress are defined. The study found that the suggested model could improve detection performance by 6-9%. Stressed users have about 14% more sparsely linked social structures than non-stressed users, indicating that their social networks are often less connected and less complex. This is just one of several intriguing phenomena that were discovered after further analysis of the social interaction data.

III. METHODOLOGY

The steps that are involved in the process of making this project can be described as shown in Figure 1.

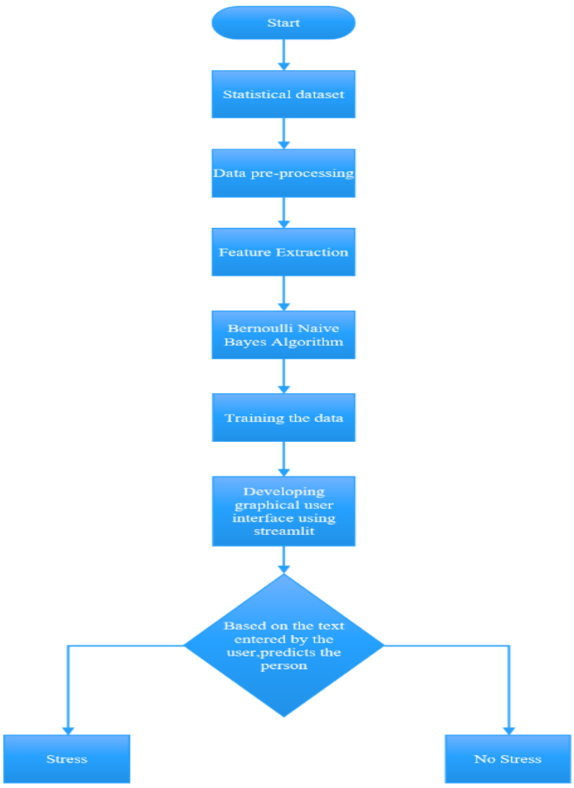

Figure 1 The flowchart of the project

A. Data Collection

The dataset for this study was gathered from social media posts made by people who had recently experienced stress or depression [11]. Totally, a dataset of 2,839 messages during the period of January 2022 to April 2023 was gathered. The data set contains posts that contain stress-related keywords, including "anxiety," "depression," and "stress."

B. Data pre-processing

The dataset was pre-processed by us by having stop words, punctuation, and URLs removed [11]. Additionally, we eliminated noisy and incomplete data. In order to get the words down to their base form, we also performed stemming. A ratio of 80:20 was used to divide the pre-processed dataset into training and testing sets.

C. Bernoulli Naive Bayes Algorithm

The Bernoulli Naive Bayes algorithm requires an understanding of the Naive Bayes principle. The supervised machine learning technique Naive Bayes predicts the likelihood of distinct classes based on a range of attributes [3][11]. It shows how likely it is that an event will occur. The term “conditional probability” also refers to Naive Bayes. Bayes Theorem serves as the foundation for Naive Bayes.

$$P\left(\frac{A}{B}\right) = \frac{P\left(\frac{B}{A}\right) * P(A)}{P(B)}$$

Where: -
A: event 1
B: event 2
P (A|B): the posterior probability that A will be true if B is true.
P (B|A): The likelihood that B is true assuming that A is true.
P(A): Prior probability that A is true.
P(B): The marginalized probability that B is true.

The denominator, or marginal probability, is ignored in the Naive Bayes classifier since we are only concerned with the largest posterior probability.

The Naive Bayes classifier is based on two fundamental principles:

- Conditional Independence: Every feature stands alone from the rest. This suggests that the functionality of one feature has no impact on that of another. This is the only explanation for the ‘Naive’ in ‘Naive Bayes’.
- Feature Importance: Each feature is equally crucial. To produce sound forecasts and obtain the most precise outcomes, it is necessary to be familiar with all the features.

Multinomial, Bernoulli, and Gaussian Bayes are the three basic subtypes of Naive Bayes [3][11][12]. The following paper will cover Bernoulli Naive Bayes.

Before moving on, let’s first examine the Bernoulli Distribution.

Let there be a random variable ‘X’ and let the probability of success be denoted by ‘p’ and the likelihood of failure be represented by ‘q’.
Success: p
Failure: q
q = 1 - (probability of Success)
q = 1 – p

$$p(x) = P[X = x] = \begin{cases} q = 1 - p & x = 0 \\ p & x = 1 \end{cases}$$

$$X = \begin{cases} 1 \; Bernoulli\ trail = S \\ 0 \; Bernoulli\ trail = F \end{cases}$$

As we can see above, x can only accept a binary value of 0 or 1. Bernoulli Naive Bayes is a member of the Naive Bayes family. It is built on the Bernoulli Distribution and only accepts binary input or values of 0 or 1. If the dataset’s attributes are binary, we can assume that Bernoulli Naive Bayes is the best algorithm to use.

D. Implementation

In this approach, text data from social media posts are extracted and tokenized using a natural language processing toolbox. We categorize the tokens as 0: no stress and 1: stress. We used CountVectorizer, a fantastic utility offered by the Python SCI-kit-learn module. It is used to turn a text into a vector depending on how often (or how many times) each word appears across the whole thing. When we have several of these texts and want to turn each word into a vector, this is useful. We implemented the Bernoulli Naive Bayes algorithm using the sci-kit-learn library in Python. We used the testing set to evaluate the algorithm's performance after training it on the training set. We have developed a graphical user interface for the system using Streamlit, which is an open-source app framework that helps us quickly deploy the development into an interactive GUI. This GUI provides the option for the user to enter text. The model determines whether the user is stressed or not based on the input from the user.

E. Performance Computation

The performance of all three Naive Bayes algorithms is compared. Parameters like Accuracy, Recall, Precision, and F-measure are used for performance analysis [11].

- $Performance = \frac{TP}{TP+FP}$
- $Recall = \frac{TP}{TP+FN}$
- $F1\ Score = 2 * \frac{Recall*Precision}{Recall+Precision}$

IV. RESULT

The results of the suggested method, which was trained and tested on the collected dataset using the Bernoulli Naive Bayes Algorithm, will be denoted as 0: no stress and 1: stress. The following table shows the accuracy, recall, precision, and f1 score of all three Naive Bayes algorithms, which will prove that the proposed system using Bernoulli Naive Bayes is best in all the considered parameters and can help us detect stress effectively.

TABLE I. PARAMETERS COMPARISON

Parameters	**Naive Bayes Algorithms**		
	Gaussian	***Multinomial***	***Bernoulli***
Accuracy	0.60	0.78	0.80
Recall	0.72	0.89	0.88
Precision	0.62	0.75	0.78
F1-Score	0.67	0.81	0.83

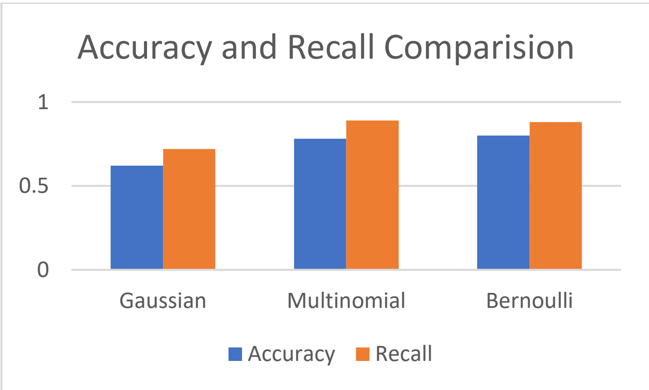

Figure 2 Comparison of accuracy and recall of all three algorithms.

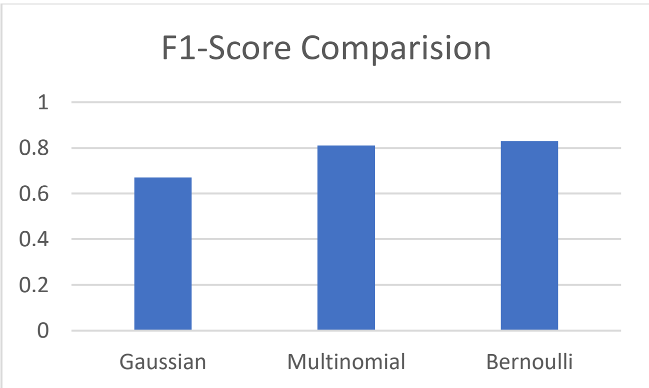

Figure 3 Comparison of F1-Score of all three algorithms.

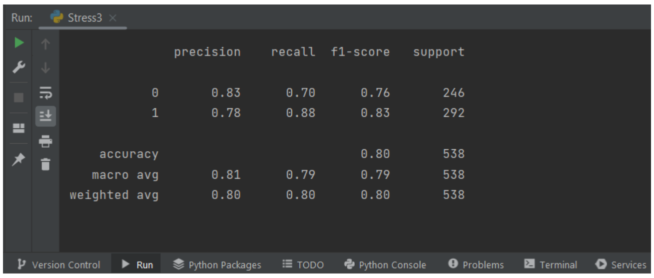

Figure 4 Results

The proposed approach revealed that the accuracy of the Bernoulli Naive Bayes algorithm in detecting stress is 80%. The algorithm's precision and recall were, respectively, 78%, and 88%. The F1 score, which is a measure of the model's overall performance, was 83%, as shown in Figure 4.

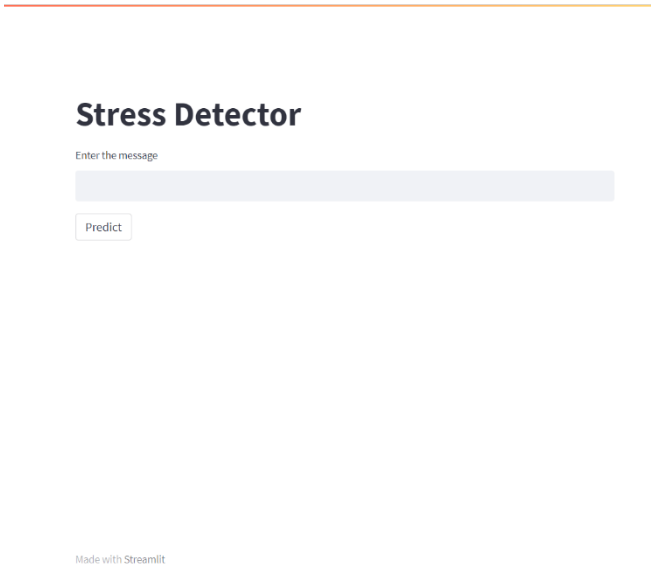

Figure 5 GUI of the proposed system

Figure 5 shows the graphical user interface of the project, where the user can enter input in the form of text, and the proposed system predicts the nature based on the number of stressed keywords given by the user.

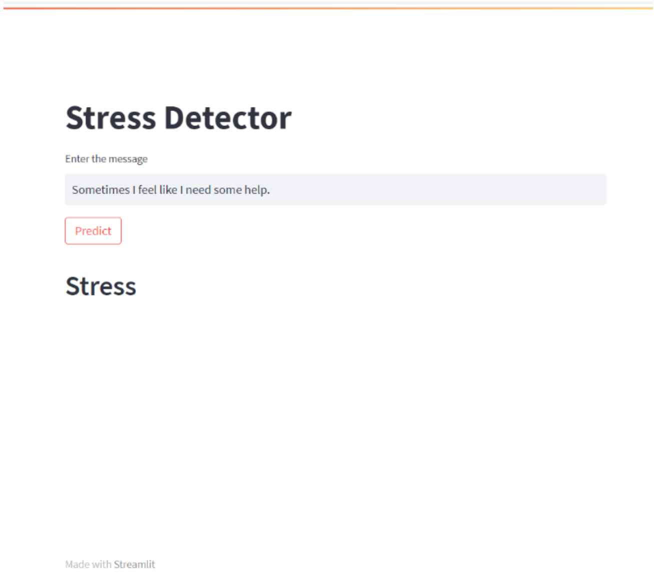

Figure 6 Output of the person showing stress

If it found any stress words in the text entered by the user. It will predict it as stress, as shown in figure 6.

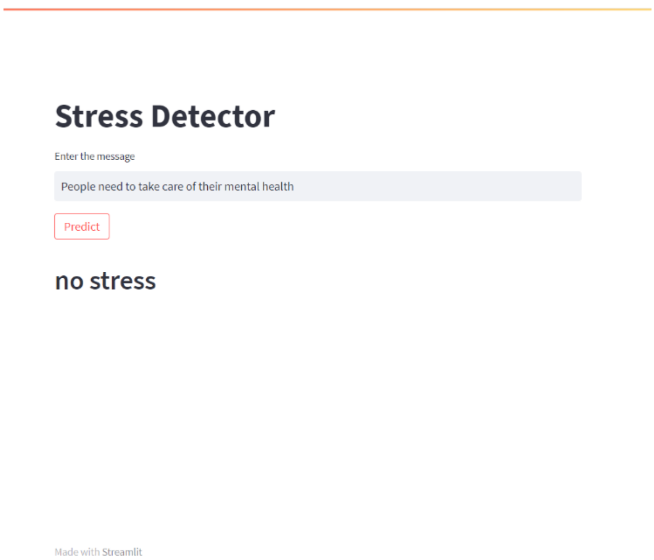

Figure 7 Output of the person showing no stress

If it doesn't find any words related to stress, then it will predict it as no stress, as shown in figure 5.

V. DISCUSSION

The outcomes of our tests demonstrate that the Bernoulli Naive Bayes algorithm can detect stress with excellent accuracy. The technique is a viable choice for real-time stress detection applications due to its effectiveness and simplicity. However, the dataset we used for our research was modest in size. To assess the algorithm's performance on larger datasets gathered in real-world contexts, more investigation is required.

VI. CONCLUSION

In conclusion, we suggested a machine learning method for detecting stress in social media messages using the Bernoulli Naive Bayes algorithm. The suggested method produced encouraging outcomes, showing that the algorithm can be applied for stress detection. The strategy offers both individuals and organizations a useful tool for monitoring and reducing stressors on social media sites. Future research can further improve the accuracy of the model by incorporating more features and using other machine-learning algorithms.

REFERENCES

[1] Saif, H., Fernandez, M., & He, Y. (2013). Semantic sentiment analysis of Twitter. In Proceedings of the 2013 International Conference on Social Computing (SocialCom) (pp. 508-515). IEEE.

[2] Li, F. L., Hsu, Y. F., & Hsu, W. L. (2016). Tweet stress detection and its correlation with social network metrics. In Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 1195-1202). IEEE.

[3] Reshma Radheshamjee, and Supriya Kinariwala "Detection and Analysis of Stress using Machine Learning Techniques" International Journal of Engineering and Advanced Technology, pp. 2249-8958, 2019.

[4] Cruz, A.P., Pradeep, A., Sivasankar, K.R. and Krishnaveni, K.S., 2020, July. A decision tree optimised SVM model for stress detection using biosignals. In *2020 International Conference on Communication and Signal Processing (ICCSP)* (pp. 0841-0845). IEEE.

[5] Li, R. and Liu, Z., 2020. Stress detection using deep neural networks. *BMC Medical Informatics and Decision Making*, *20*, pp.1-10.

[6] Lavanya Rachakonda, Prabha Sundaravadivel, Saraju P. Mohanty, Elias Kougianos, Madhavi Ganapathiraju, "A smart sensor in the IoMT for stress level detection".

[7] Boon Giin Lee, Wan-Young Chung "Wearable Glove-Type Driver Stress Detection Using a Motion Sensor," IEEE Transaction on Intelligent Transportation systems," 2018.

[8] H. Lin, J. Jia,Q. Guo, Y. Xue, J. Huang, L. Cai, and L. Feng,"Psychological stress detection from cross-media microblog data using deep sparse neural network," in Proc. IEEE Int. Conf. Multimedia Expo, 2014, pp. 1–6.

[9] Huijie Lin, Jia Jia, Jiezhong Qiu, Yongfeng Zhang, Guangyao Shen, Lexing Xie,Jie Tang, Ling Feng, and Tat-Seng Chua "Detecting Stress Based on Social Interactions in Social Networks" in IEEE Transactions on Knowldege and Data Engineering, 2017.

[10] Archana, V.R. and Devaraju, B.M., 2020. Stress Detection Using Machine Learning Algorithms. *International Journal of Research in Engineering, Science and Management*, *3*(8), pp.251-256.

[11] Kanchana, J.S., Fathima, H.T., Surya, R. and Sandhiya, R., 2018. Stress detection using classification algorithm. *Int. J. Eng. Res. Technol*, *7*(04).

[12] Bobade, P. and Vani, M., 2020, July. Stress detection with machine learning and deep learning using multimodal physiological data. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 51-57). IEEE.