

# Stress Detection Using Classification Algorithm

J. S. Kanchana

Associate Professor

Department of Information Technology  
K.L.N College of Engineering  
Madurai, India

R. Surya

Department of Information Technology  
K.L.N College of Engineering  
Madurai, India

H. Thaqneem Fathima

Department of Information Technology  
K.L.N College of Engineering  
Madurai, India

R. Sandhiya

Department of Information Technology  
K.L.N College of Engineering  
Madurai, India

**Abstract**— Psychological problems are becoming a major threat to people's life. It is important to detect and manage stress before it turns into a severe health issue. Nowadays people share their feeling in social media regularly. It becomes easy to detect the stress of the users based on their social behavior. Also, traditional stress detection methods are time consuming and costly. So the linguistic attributes in tweets can be leveraged to detect individual user stress. In this project, the stress states of users are classified using Naive Bayes classification algorithm and are categorized into stressed and non-stressed user.

## I. INTRODUCTION

Web Mining is the application of data mining techniques which is used to discover interesting usage patterns from Web data so as to understand and better serve the needs of Web-based applications. Web mining is the collection of information gathered by using traditional data mining methods and techniques with information gathered over the World Wide Web. Web mining is used to analyze customer behavior, evaluate the effectiveness of a particular Web site, and help quantify the success of a marketing campaign.

Classification is a data mining function that assigns item in a collection to target categories or classes. The aim of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks. Classification models are tested by comparing the predicted values to known target values in a set of test data.

A Naive Bayes classifier is an algorithm that uses Bayes' theorem to classify objects. Naive Bayes classifiers will assume strong, or naive, independence between attributes present in the data points. Major uses of naive Bayes classifiers are spam filters, text analysis and medical diagnosis. Naive Bayes classifiers are widely used for machine learning because they are simple in implementation. Naive Bayes is also called as simple Bayes or independence Bayes.

## II. RELATED WORK

Huijie Lin analysed that users stress state is closely related to that of their friends in social media, and a huge dataset from real-world social platforms is used to analyze the correlation of users' stress states and social interactions. Stress-related attributes like textual, visual, and social attributes are defined, and then a factor graph model combined with Convolutional Neural Network is proposed to make use of the tweet content and social interaction information for stress detection. The

paper revealed that proposed model can improve the detection performance by 6-9 percent. On analyzing the social interaction data further, several intriguing phenomena were discovered, i.e., the number of social structures of sparse connections of stressed users is around 14 percent higher than that of non-stressed users, indicating that the social structure of stressed users tends to be less connected and less complicated than that of non-stressed users. [1]

Jia Jia proposed a stress detection method automatically using cross-media microblog data. A framework of three-level is constructed to design the problem. A set of low-level features is obtained from the tweets. Then, middle-level representations based on psychological and art theories is defined and extracted: linguistic attributes from texts, visual attributes from images, and social attributes from comments. At last, a Deep Sparse Neural Network is created to learn the stress categories. The proposed method is effective and efficient on detecting psychological stress from microblog data. [4]

Ling Feng investigated the correlations between stress state of user and their tweeting content, social engagement and behavior patterns. Then stress-related attributes are defined as follows: 1) low-level content attributes from a single tweet, including text, images and social interactions; 2) user-scope statistical attributes through their weekly micro-blog postings, leveraging information of tweeting time, tweeting types and linguistic styles. The content attributes are combined with statistical attributes, by a convolutional neural network (CNN) with cross auto encoders to generate user-scope content attributes. Finally, a deep neural network (DNN) model is proposed to incorporate the two types of user-scope attributes to detect users' psychological stress. Experimental results show that the proposed model is effective and efficient on detecting psychological stress from micro-blog data. [2]

Jichang Zhao has built a system called MoodLens in which 95 emoticons are mapped into four categories of sentiments, i.e. angry, disgusting, joyful, and sad. These sentiments act as the class labels of tweets. Around 3.5 million labeled tweets are collected as the training data and a Naive Bayes classifier is trained, with an empirical precision of 64.3%. Using MoodLens for real-time tweets obtained from Weibo, several interesting temporal and spatial patterns are observed. Also, sentiment variations are well-captured by MoodLens to effectively detect abnormal events in China. [3]

### III. PROPOSED SYSTEM

In the proposed system, Naive Bayes algorithm and support vector machine algorithm are used to classify stressed and non-stressed user. Ranking is based on the number of stressed posts tweeted by each user. Then the user with highest stress is found.

#### Data Preprocessing

In data preprocessing, the noisy and incomplete data are removed. The twitter dataset has the user ID, stress category and user's tweet. For example: Incomplete tweet data like "u guys know y" and "very much so" are removed from the dataset.

#### Naive Bayes Classification

Naive Bayes algorithm is the algorithm that learns the probability of an object with certain features belonging to a particular group or class. Bayes theorem provides a method of calculating the posterior probability,  $P(c|x)$ , from  $P(c)$ ,  $P(x)$ , and  $P(x|c)$ . Naive Bayes classifier assumes that the effect of the value of a predictor ( $x$ ) on a given class ( $c$ ) is independent of the values of other predictors. This assumption is called class conditional independence.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

- Where  $P(c|x)$  is the Posterior probability
- $P(x|c)$  is the likelihood
- $P(c)$  is the class prior probability
- $P(x)$  is the Predictor prior probability

#### Support Vector Machine using WEKA

Support vector machine is a supervised learning algorithm which can be used for classification and regression. Support vector machine in WEKA gives correct and incorrect instances. It finds root mean squared error, mean absolute error of the training data. In WEKA the performance is measured by using confusion matrix in terms of TP, FP, and F-Measure in terms of recall and precision.

#### Performance Computation

The performance of both Naïve Bayes algorithm and Support Vector Machine algorithms are compared. Parameters like Accuracy, Recall, Precision and f-measure are used for performance analysis.

- Performance =  $TP / (TP+FP)$
- Recall =  $TP / (TP+FN)$
- F1 Score =  $2 * (Recall * Precision) / (Recall + Precision)$

### IV. EXPERIMENTAL RESULTS

From the twitter dataset, the algorithm classifies the users as either stressed or non-stressed and finds the count of each type of user. It also finds the number of tweets posted per week by the users. The following tables show the accuracy (TABLE I) and F-measure (TABLE II) of Convolutional Neural Networks used in the existing system and Naïve Bayes algorithm used in the proposed system. Accuracy is calculated

individually for text, text and social behavior and overall tweets.

TABLE I. ACCURACY COMPARISON FOR CNN AND NB

Modalities	Accuracy	
	CNN	NB
Text	0.8713	0.8913
Text+Social	0.8628	0.8718
All	0.9155	0.9254

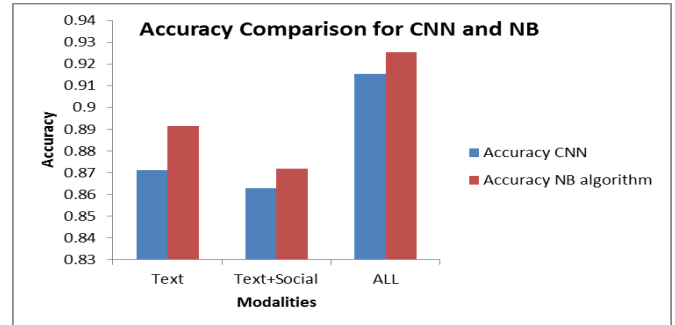


Fig. 1. Comparison of accuracy for Convolutional Neural Networks and Naïve Bayes algorithm

TABLE II. F-MEASURE COMPARISON FOR CNN AND NB

Modalities	F-Measure	
	CNN	NB
Text	0.8794	0.8893
Text+Social	0.8711	0.8911
All	0.9340	0.9451

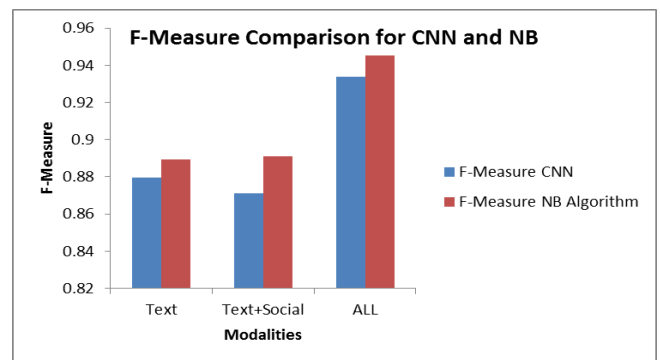


Fig. 2. Comparison of F-Measure for Convolutional Neural Networks and Naïve Bayes algorithm

### V. CONCLUSION

In this paper, we presented a model for detecting the psychological stress level of the users by leveraging the tweets of each user and their social behavior. The classification revealed the number of stressed and non-stressed users, the number of tweets posted per week by each user and the number of stressed and non-stressed posts per week.

Our future work will include classification based on the tweet text and images. The visual attributes will also be included in the process of detecting user stress.

## REFERENCES

- [1] 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 Knowledge and Data Engineering, 2017.
- [2] H. Lin, et al., "User-level psychological stress detection from social media using deep neural network," in Proc. ACM Int. Conf. Multimedia, 2014, pp. 507-516.
- [3] J. Zhao, L. Dong, J. Wu, and K. Xu, "Moodlens: An emoticonbased sentiment analysis system for chinese Tweets," in Proc. 18<sup>th</sup> ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2012, pp. 1528-1531.
- [4] 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.
- [5] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from Twitter," in Proc. IEEE 3rd Int. Conf. Privacy, Security, Risk Trust, IEEE 3rd Int. Conf. Soc. Comput., 2011, pp. 149-156. [17] M. S. Granovetter, "The strength of weak ties," Amer. J. Sociology, vol. 78, pp. 1360-1380, 1973.
- [6] Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., ... & Choudhury, T. (2012, September). StressSense: Detecting stress in unconstrained acoustic environments using smartphones. In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (pp. 351-360). ACM.
- [7] Paredes, P., Sun, D., & Canny, J. (2013, May). Sensor-less sensing for affective computing and stress management technology. In Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on (pp. 459-463). IEEE.
- [8] Sadilek, A., Kautz, H. A., & Silenzio, V. (2012, June). Modeling Spread of Disease from Social Interactions. In ICWSM.
- [9] Paul, M. J., & Dredze, M. (2011, July). You are what you Tweet: Analyzing Twitter for public health. In ICWSM.
- [10] De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013, July). Predicting depression via social media. In AAAI Conference on Weblogs and Social Media.
- [11] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313(5786), 504-507.
- [12] Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A. (2008, July). Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th international conference on Machine learning (pp. 1096-1103). ACM.