Stress Detection Based on Social Interaction Using Web Mining

[1]Swapnaja Shahaji Darole, [2]Safwan Burhan Kurawlay, [3]S. N. Mhatre
[1]B.E student, [2]B.E student, [3]Assistant Professor
[1]Department of Information Technology,
[1]Bharati Vidyapeeth College of Engineering, Navi Mumbai, India

Abstract : Nowadays , everyone has mental stress in their life. People are used to share their feelings, daily activities and interacting with friends on the social media platforms. In this, we find that users stress state is closely related to their interaction in the social media. From this, We are taking live users data from twitter so that we can employ large scale dataset from social media platform to study their correlation, stress state and social interaction. We define a set of stress related textual, visual and social attributes from various contents based on that we are applying some classification rules in web mining. From this, we can get number of stressed users and non-stressed users.

Keywords **- Mental Stress, Social Media Platform, Stress State, User Data, Twitter, Dataset, Correlation, Social Attributes, Classification Rules, Web Mining.**

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 the information gathered over the World Wide Web. Web mining is used to analyze customer behaviors, evaluate the effectiveness of a particular Web site, and help quantify the success of a marketing campaign.

Classification is the data mining function that assign item in a collection to target categories or classes. The aim of the 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 to risks. Classification models are tested by comparing the values which are predicted to known target values in a set of test data.

A Naive Bayes Classifier is an algorithm which is used to classify objects by using Bayes Theorem. Naive Bayes Classifier will assume strong, or naive, independence between attributes present in the data points. Major uses of naive Bayes classifiers are spam filters, test analysis and medical diagnosis. Naive Bayes Classifier are widely used for machine learning because they are simple in implementation. Naive Bayes is also called as Simple Bayes or Independence Bayes.

II. LITERARTURE REVIEW

Huijie Len revealed that proposed model can improve the detection performance by 6-9 percent. On analyzing the social data further, several intriguing phenomena were discovered, that is 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]

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]

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 text, visual attributes from images and social attributes from comments. Atleast 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]

III. RESEARCH METHODOLOGY

The Admin uses the API of the social media sites to connect and retrieve user data. user data is then stored in the database, from there the data is pre-processed. The pre-processed data gets on applied with classification techniques in our system, Naïve Bayes and Support Vector Machine are used. Ranking is based on the number of stressed posts tweeted by each user. Classification techniques gives us accuracy which is specify the stressed and non-stressed user.
The system can be divided into following units:

API Connection

Storing Data in Database

Preprocessing

Classification

Fig3.1 describes the architectural diagram of the proposed system.

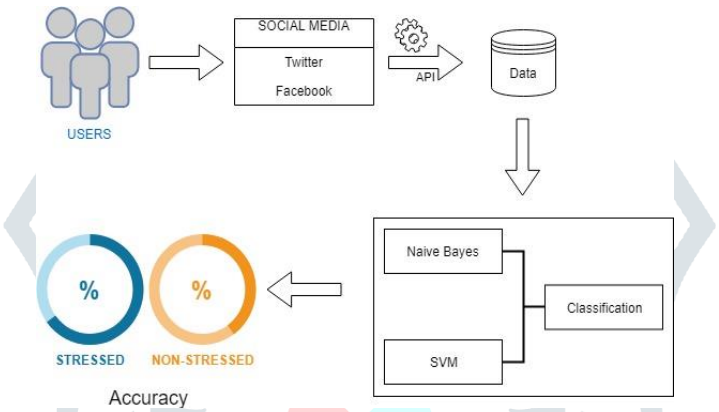

Fig3.1 Architectural Diagram

API Connection

Social Media provides API to developers who want to use their data for research or other purposes. We have applied for developer account to get an access to API which contains keys to be entered in the code.

Storing Data in Database

A code is written which gets the data from the API connection with the social media sites. The data gets stored in our database by making a connection of the database with the code.

Preprocessing

Data preprocessing specifies any type of processing performed on raw data to prepare it for another processing procedure. Commonly used as a initiatory data mining practice, data preprocessing transforms the data into a format that will be more easily and effectively processed for the purpose of the user.

Classification

Classification is used to classify each and every item in a set of data into one of predefined set of classes or groups. The data analysis task classification is where a model or classifier is constructed to predict categorical labels or the class label attributes. Classification is a data mining function that assigns items in a collection to target number of categories or classes. The goal of classification is to predict the target class accurately for each case in the data. In the Proposed system we have used Naïve Bayes and Support Vector Machine to classify the data in stressed and on non-stressed user.

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).This classifier assumes that the effect of the value of a predictor (x) on a given class (c) is independent of the other predictors values. 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 JAVA/PYTHON

Support vector machine is a supervised learning algorithm which can be used for classification and regression. Support vector machine gives correct and incorrect instances. It finds root mean squared error, mean absolute error of the training data. 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. RESULT AND DISCUSSION

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.

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

TABLE I. ACCURACY COMPARISON FOR CNN AND NB

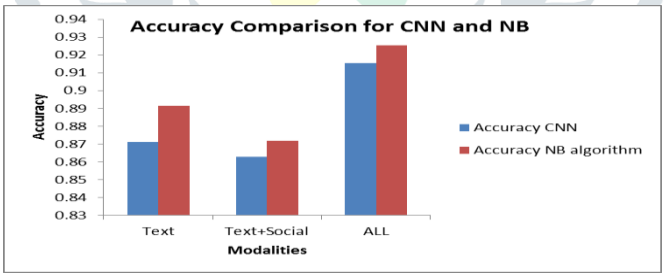

Fig. 1. Comparison of accuracy for Convolutional Neural Networks and Naïve Bayes algorithm

TABLE II. F-MEASURE COMPARISON FOR CNN AND NB

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

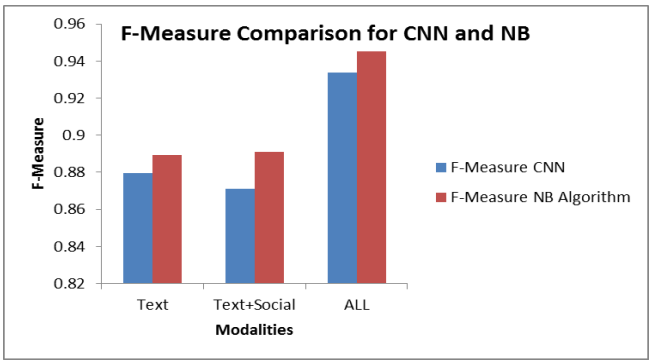

Fig. 2. Comparison of F-Measure for Convolutional Neural Networks and Naïve Bayes algorithm

V. CONCLUSION

In this paper, we presented a model for detecting the mental stress level of the users by leveraging the tweets of each user and their social behaviour. 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

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 Knowldege 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. 18th 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), 504507.

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