RESEARCH ARTICLE OPEN ACCESS

TWEET BASED STRESS ANALYSIS USING WEB MINING AND CLASSIFIED CLUSTERS

S.Nivetha[1], Dr. T.Meyyappan[2]
1(Department of Computer Science, Alagappa University, Karaikudi
Email: skn.nivetha@gmail.com)
2 (Department of Computer Science, Alagappa University, Karaikudi
Email:meyyappant@alagappauniverity.ac.in)

Abstract:

Web mining is a category of data mining technique and it is mostly used to discover interesting data from a web store. Web mining plays an important role in understanding the customer's need, analysing user's behaviour and their needs through the raw data. The discovered data can be applied into various fields to offer newer services.

This proposed research work aims to discover the stress possibilities of web users. The proposed work focus on prediction of stress level of twitter users based on their tweets. The tweet history of every user can be captured from the twitter server and the raw data will be clustered based on the tweets and session. The clustered data will be extracted to detect the stress patterns. The proposed approach implements the K-Means cluster approach with Query mining based classification technique to generate quick and accurate reports of stress possibilities. The main objective of the proposed approach is to provide faster prediction compared to existing methods. The proposed method is implemented with CNN and NB techniques

.Keywords —**Stress, Web Mining, Clusters, Prediction, Query Mining.**

I. INTRODUCTION

Twitter is a most popular social network in today's digital communication media. Approximately 600 - 700 tweet messages are floated on twitter by its users, every second. Twitter lets its users to post short messages to their followers and also offers hash tagging. The hash tagging feature leads to popularize the current issue and a topic which is related to the tweets [2].

The social media environments are essential which reflects the behaviour of users. Many kinds of useful information can be derived from the social media users and their interactions. Most of the users are sharing messages based on thought and their mental stress situation. Hence, it is possible to predict their stress level based on their tweets and what they are sharing [5].

Anybody can be affected by stress and the stress can be due to various factors and issues. The major factors influencing stress are peer pressure, parental pressure, exam, working environments, social factors and personal depression and others. Handling user's stress is an important activity which helps to promote the business through online transactions. For example, if an online servicing portal automatically predicts the stress level and mood of the user to determine related contents, it can attract and retain users to promote their business.

It is mandatory to predict the stress and interest factors before providing customer services. The same is applicable for online social media, online shopping portals and other services. This paper proposes and implements the new ways and algorithms to predict user stress level with web data mining approaches.

II. RELATED WORK

Table 1: Related Works

Paper Title	Newly Introduced	Algorithms Used	Data Collection
Detecting Adolescent Psychological Pressures from Micro Blog [1, 6]	Tweet Based Pressure Prediction	Naïve Bayes, SVM, Neural Networks	Twitter
Helping Teenagers Relieve Psychological Pressures [4]	T Helper	Gaussian Process	Twitter

A Mathematical approach in Gauging influence by identifying shifts in the emotions of social media users [7]	-	-	Twitter
User Level Psychological Stress Detection from social media using Deep Neural Networks [3]	-	Neural Networks	Weibo (China) and Twitter
Stress Detection using Classification Algorithms [9]	-	Neural Networks and Naïve Bayes	Twitter

III. PROPOSED APPROACH

The existing research works in the literature employ Naive Bayes and CNN techniques. The proposed research work is implemented with Clusters and Classification algorithms.

The proposed method comprises various stages. The web mining is the initial stage, data cluster creation is the second stage and applying classification is the final stage. Web usage mining steps are as follows: Data Pre-processing, Pattern analysis and discovering patterns.

At first, the twitter data set is loaded into the architecture. The architecture loads the twitter raw data set in this stage. The raw data can't be used as it is and it should be cleaned before using it. The loaded dataset is cleaned with unwanted data locator mechanism. The mechanism is designed to detect unwanted data like default twitter navigations and separation marks and others. The cleaned data are temporarily stored in a DAT file.

The data detection and data conversion steps are taken in the same stage. The cleaned data will be loaded and the available data detection mechanism detects which data are available in the dataset. The data conversion is the final step which converts the detected data as readable.

The second stage is the cluster stage. In this state, the converted readable data are partitioned with small clusters based on requirements. To create these clusters this proposed approach uses K-Means clustering algorithm. The sequence of steps in K-Means clustering algorithm are:

Step 1: Begin

Step 2: Choose Number of Cluster centers

Step 3: Set Initial Cluster - Random

Step 4: Put Object closest cluster center

Step 5: Find New Cluster

Step 6: Move objects to clusters

Step 7: End

The final stage called resulting stage is classification. The primary aim of classification is to predict the required patterns accurately based on data. In this stage, the stress possibilities and non stress possibilities are calculated based on clustered data. The classification is achieved in this proposed system with the help of query mining approach. The exact query is generated to detect the tweets from clusters based on gender, twitter ID and message vice versa.

IV. PROPOSED APPROACH – ARCHITECTURE

The architecture diagram illustrates the three stages and its internal operations performed in the proposed approach. The proposed approach is easier to implement. It is more efficient in generating accurate predictions than the existing Naïve bayes approach. The algorithm K-Means and query mining based classification offers a faster performance than existing approaches.

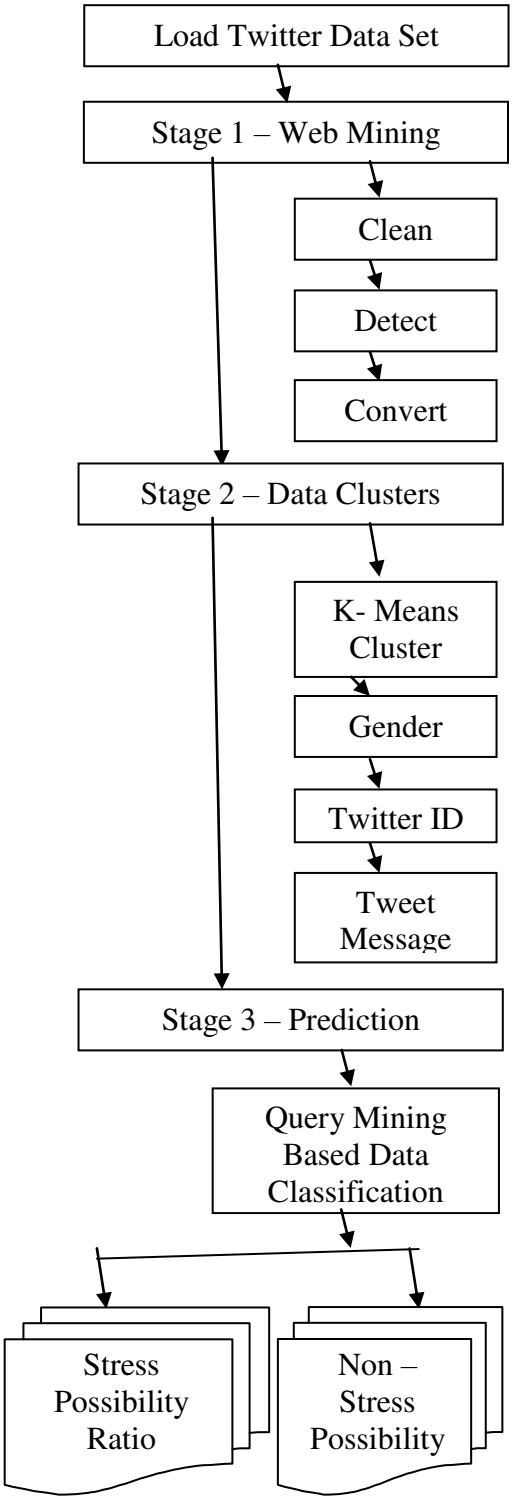


Fig. 1: Proposed Architecture

V. EXPERIMENTAL RESULTS

The proposed method is implemented in MATLAB. The proposed method is experimented with various twitter datasets. The performance results are illustrated below:

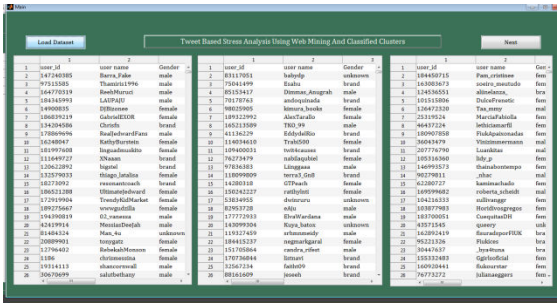


Fig. 2: A Loaded Twitter Dataset after Data Cleaning

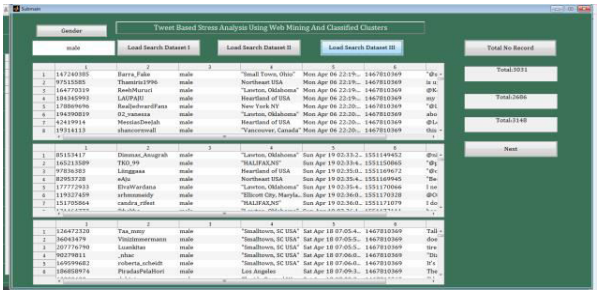


Fig. 3: Gender Analytics after Cluster clean

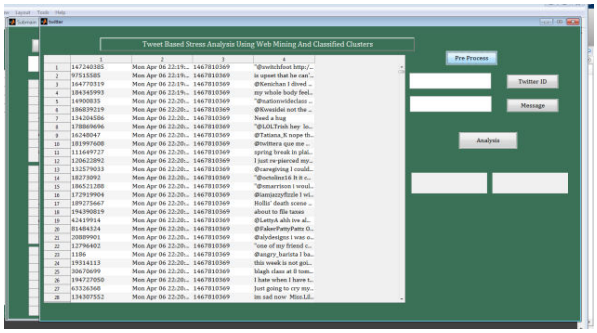
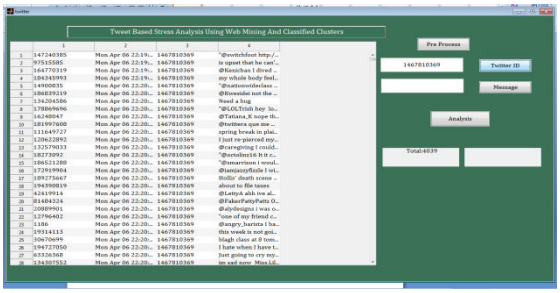


Fig.4. Classified Data



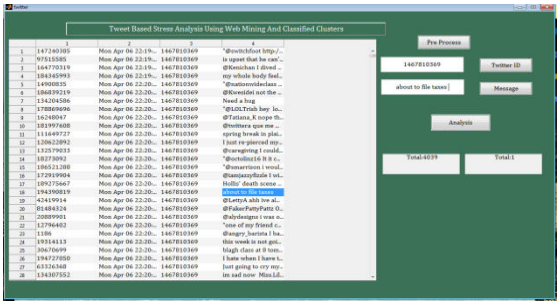


Fig. 5: Twitter ID based Cluster and Mined Data

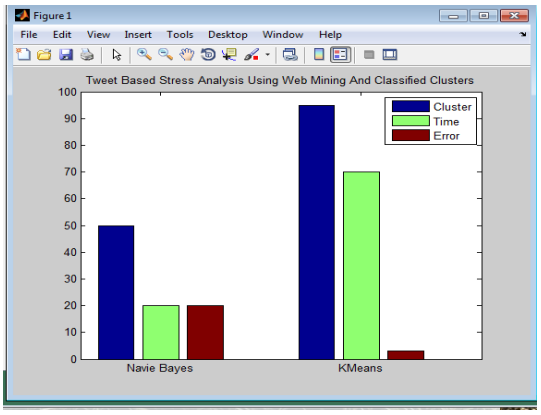


Fig. 6: Comparison of Stress Prediction Results: Naïve Bayesian Vs Proposed Approach

VI. CONCLUSION

The method proposed in this research work is unique and efficient model or architecture to predict stress level of twitter users. This proposed approach is implemented with web mining strategies,

K-Means clustering and query mining based classification techniques. It is implemented with MATLAB tool. It generates accurate and readable results. Also it offers faster performance than the existing one and the performance chart is also generated using the proposed architecture. In future, the same can be predicted by their shared photos and videos. Also the same architecture can be implemented for other databases like facebook, instagram and etc.

REFERENCES

[1] Li, Y., Huang, J., Wang, H. and Feng, L., 2015, June. Predicting Teenager's Future Stress Level from Micro-Blog. In 2015 IEEE 28th International Symposium on Computer-Based Medical Systems (pp. 208-213). IEEE.

[2] Asur, S. and Huberman, B.A., 2010, August. Predicting the future with social media. In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01 (pp. 492-499). IEEE Computer Society.

[3] Lin, H., Jia, J., Guo, Q., Xue, Y., Li, Q., Huang, J., Cai, L. and Feng, L., 2014, November. User-level psychological stress detection from social media using deep neural network. In Proceedings of the 22nd ACM international conference on Multimedia (pp. 507-516). ACM.

[4] Li, Q., Xue, Y., Jia, J. and Feng, L., 2014. Helping Teenagers Relieve Psychological Pressures: A Micro-blog Based System. In EDBT (pp. 660-663).

[5] Spear, L.P., 2000. The adolescent brain and age-related behavioral manifestations. Neuroscience & Biobehavioral Reviews, 24(4), pp.417-463.

[6] Xue, Y., Li, Q., Jin, L., Feng, L., Clifton, D.A. and Clifford, G.D., 2014, April. Detecting adolescent psychological pressures from micro-blog. In International Conference on Health Information Science (pp. 83-94). Springer, Cham.

[7] Servi, L. and Elson, S.B., 2014. A mathematical approach to gauging influence by identifying shifts in the emotions of social media users. IEEE Transactions on Computational Social Systems, 1(4), pp.180-190.

[8] Mogadala, A. and Varma, V., 2012, October. Twitter user behavior understanding with mood transition prediction. In Proceedings of the 2012 workshop on Data-driven user behavioral modelling and mining from social media (pp. 31-34). ACM.

[9] JS Kanchana,JS.,ThaqneemFathima, H., Surya, R.,and Sandhiya, R., 2018, "Stress Detection Using Classification Algorithm", International Journal of Engineering Research & Technology, 7(04), pp.84-86.