International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

DETECTING STRESS BASED ON SOCIAL INTERACTIONS IN SOCIAL NETWORKS USING STACKING ENSEMBLING

[1]N.SIVA PARVATHI, [2]M.PRADEEP
[1]M.Tech Schoalr, Dept of CSE, St. Mary's Women's Engineering College, Budampadu, Guntur, A.P, India
[2]Associate Professor, Dept of CSE, St. Mary's Women's Engineering College, Budampadu, Guntur, A.P, India

ABSTRACT: Social media currently plays an important role in current social structure of humanity. People are used to sharing their daily activities and interacting with friends on social media platforms which provides a unique opportunity with mining, measuring, modeling different user behavioral patterns. The project is based on the idea of detecting the users stress in a proactive manner. The proposed system uses ensemble learning model which gives better accuracy compared to other standard machine learning models with the help of meta classifier and the applied algorithms are KNN, Random Forest, Convolutional Neural Network, Support Vector Machine.

Index Terms— Stress detection, factor graph model, micro-blog, social media, healthcare, social interaction

I. INTRODUCTION

The computer-assisted process of digging through and analyzing enormous quantities of data and then extracting the data's significance is known as data mining or knowledge exploration. Data mining instruments for predicting preferences and future trends providing businesses with the tools they need to make strategic decisions, guided by expertise. Data mining software can provide answers to market concerns that have previously taken so long to resolve. They explore hidden trend databases, uncovering predicting data that experts might overlook because it falls outside of their assumptions.

Data mining (also known as data exploration or knowledge mining) is the act of evaluating and summarizing data from many perspectives into valuable information – information [1] that can be used to improve income, reduce costs, or do both. Data mining software is one of several analytical approaches to data analysis. It allows users to examine, categories, and summaries data-related relationships from a variety of perspectives. Data mining is the process of identifying patterns, [2] or similarities between thousands of fields in massive relational databases.

Though the term "data mining" is new, the technology is old. For years, powerful machines have been used to sift through massive amounts of supermarket scanner data and analyze market research studies. Continuous advances in computer processing power, disc storage, and statistical methodologies, on the other hand, are drastically enhancing study accuracy while driving down costs. Although data mining is still in its infancy, firms in a wide range of industries, including retail, banking, Data To make advantage of historical data, mining tools and techniques are already being applied in health care, manufacturing, transportation, and aerospace.

Data mining employs pattern recognition technologies and statistical and mathematical methodologies to help analysts uncover significant information, linkages, trends, patterns [2], exceptions, and anomalies that might otherwise go undiscovered when sifting through warehoused knowledge. Data mining is a technique for identifying patterns and correlations in data so that businesses may make more informed decisions. Data mining can aid in the identification of sales patterns, the development of more effective marketing efforts, and the accurate prediction of customer loyalty. Data mining's most basic applications.

II. LITERATURE SURVEY

The detection of psychological stress is connected to the topics of sentiment analysis and emotion detection.

Emotion recognition at the tweet level in social networks is the subject of research.

Computer-assisted emotion identification, analysis, and application, particularly on social networks, has received a lot of interest in recent years Relationships between stress and psychology, Personality traits may be a significant factor to consider.

For instance, demonstrates that Daily is a reliable source of information. Stress can be reliably identified using behavioral metrics. Mobile phone users provide service. Many tweet-level social media emotion analysis studies use text-based linguistic features and traditional categorization methodologies. Proposed the use of a technology called Mood Lens to undertake emotion analysis Weibo, a Chinese micro blogging site, divides emotions into four categories: furious, nasty, glad, and sad.

Investigated the issue of emotional transmission on social networks and discovered that anger had a stronger link between users than joy, meaning that negative emotions could spread faster and further. We can utilize this inference to combine the social influence of users who detect stress because stress is typically considered as a bad emotion. However, these activities primarily rely on textual communication via social media networks. Content In practice, social network data is made up of sequential and interrelated items from different sources and modalities, making it cross-media data.

Emotion detection at the user level in social networks is the subject of research.

While the identification of tweet-level emotions reflects the rapid feeling expressed People's emotions or psychological stress states are often more long-lasting varying throughout time in a single tweet In recent years, a lot of research has focused on user-level emotion identification in social networks. Our prior work proposed utilising a deep convolutional network to learn user-level presentation on sequential tweet series over time to detect social media psychological stress situations using a deep convolutional network.

Employed the homophilia concept to incorporate social ties into Twitter's user-level sentiment analysis. Although some user-level emotion detection research has been done, the importance of social links in one's states of psychological stress, as well as the value of social relationships in one's states of psychological stress

Utilizing social interactions for social media analysis is the subject of research.

Social networking is one of the most important features of social media networks. Many scholars are now focusing on how to leverage social interaction information to improve the efficacy of social media analysis.

Investigated the relationship between social connections and users' beliefs and behaviors, finding that interactions centered on Twitter can lead to successful cognitions. Used Flicker comments to forecast how photographs published on Flicker will make people feel. These studies, on the other hand, focused solely on the content of social interactions, such as textual comment material, while disregarding the underlying structural aspects, such as how users are related.

III. PROBLEM DEFINITION

While the identification of tweet-level emotions reflects the rapid feeling expressed in a single tweet, people's emotions or psychological stress states are often more long-lasting, shifting over time. In recent years, a lot of research has focused on user-level emotion identification in social networks. Our previous work proposed using a deep convolutional network to learn user-level presentation on sequential tweet series over a period of time to detect social media psychological stress conditions. People's reactions to various types of demands or problems are referred to as stress.

While this reaction will help us stay focused, energetic, and mentally engaged while working appropriately, if it is exaggerated, sadness, anxiety,

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

hypertension, and a slew of other life-threatening conditions will undoubtedly result. People use cyberspace as a large soapbox to air their grievances and share what they go through in their daily lives. As a result, it can be utilized as a highly useful way to evaluate a person's stress levels based on the posts and status changes he/she shares. This is a suggestion for a website that accepts the subject's Twitter username as an input. Sentiment Analysis is used to scan and analyze the subject's profile and provide results. Such findings reveal the subject's total stress levels as well as a description of his or her mental and emotional state.

Implementation Study

While tweet-level emotions reflect the fleeting emotion communicated in a single tweet, people's emotions or psychological stress states are often more long-lasting, fluctuating over time. A lot of research has concentrated on user-level emotion identification in social networks in recent years [27]. Our prior work [27] proposed utilizing a deep convolutional network to learn user-level presentation on sequential tweet series over time to detect social media psychological stress situations using a deep convolutional network. Stress refers to people's reactions to many forms of demands or issues. While this reaction might help us stay focused, energetic, and mentally engaged while working, if it is excessive, it can lead to depression, anxiety, hypertension, and a plethora of other life-threatening illnesses. People use the internet as a giant soapbox to voice their frustrations and share their daily experiences. As a result, it can be used to assess a person's stress levels based on the posts and status updates that he or she shares. This is an idea for a website that takes the subject's Twitter handle as an input. Sentiment Analysis is a technique for scanning and analyzing a person's profile and providing findings. Findings like these reveal entire stress levels of the person, as well as a description of his or her mental and emotional state

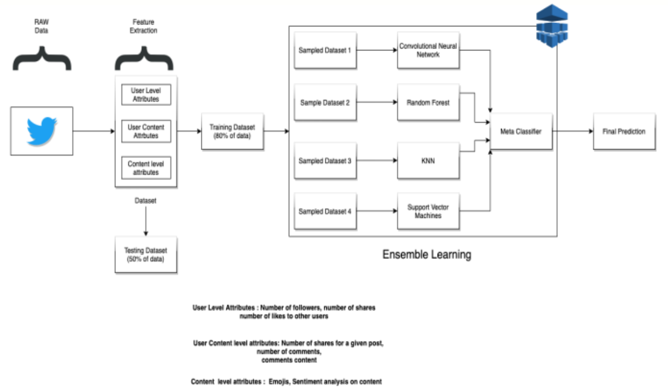

Fig. 1: key components of the proposed framework

Represents how all of the previously acquired knowledge can be examined, as well as how all of the sentences can be extracted utilizing emotions. Following the extraction of words to determine the sentences following recognition of the expression, the score for each sensation for each category polarity and categorized result is calculated.

IV PROPOSED APPROACH

Intense negative thoughts and the lack of positive emotions, which can be understood using machine learning methods, are the cardinal sign of psychological stress. In The cardinal indicator of psychological stress is intense negative thoughts and a lack of happy feelings, which can be understood using machine learning algorithms. Optical Character Recognition (OCR) for image processing, Natural Language Processing (NLP), and Convolutional Neural Network (CNN) for text content processing are all included in the proposed system.

Image processing that detects and extracts text tweets from photographs includes preprocessing, attribute extraction, classification, and linking to NGOs. The key components of the proposed framework are depicted in Figure 1. The information was gathered through social media channels like Twitter and Facebook. A data selection of graphics and text was used to determine the user's stress level. The photo dataset is extracted and analyzed using OCR, which extracts the text. After that, the text tweet material dataset and image extracted text dataset were used as preprocessing input and feature extraction using NLP. CNN plays a crucial role in categorizing favorable and negative

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

tweet content. Finally, users' disparaging tweets and information are compiled and given to the NGO for advice. The information about the stressed user (i.e. user ID) is extracted and supplied to the CLASSIFER. This is one of the ways that a suicidal individual is discovered in a social network, and it is quite likely to reduce the number of suicide attempts.

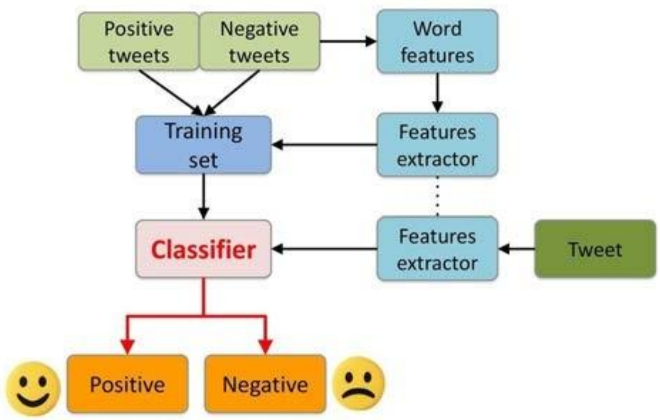

Fig 2: Architecture of psychological stress prediction

The non-stressed user is represented by a positive (1), whereas the stressed user is represented by a negative (0). In order to alert CLASSIFER to the fact that it interacts with After the CNN classification of social media dropout layer is eventually applied to the network regularisation and to avoid over-fitting concerns value, which offers the value of likelihood as an output, the projected result may be classified as positive (1). Ensemble approaches, at the network's conclusion, assist improve machine learning results by integrating numerous models. When compared to a single model, ensemble approaches provide for superior predictions. Many prominent machine learning contests, including Netflix Competition, KDD 2009, and Kaggle, awarded first place to ensemble approaches. The main hypothesis is that by correctly combining weak models, we can get more accurate and resilient models.

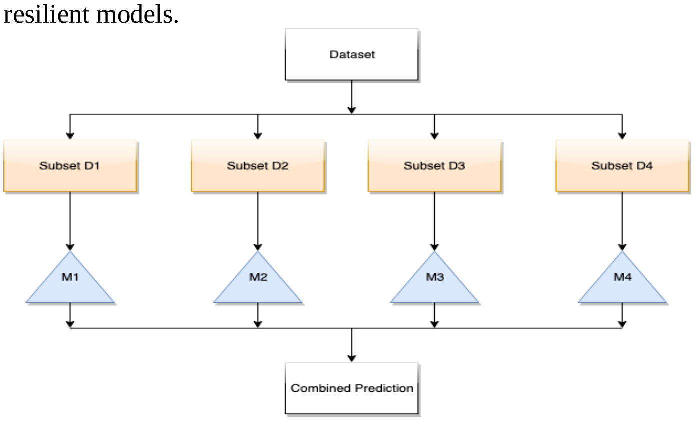

Fig 3: Proposed method implementation

4.1 Dataset Description

Twitter dataset:

Twitter allows its users to mine tweets and make datasets. So different organizations and educational institutions including Harvard and Cornell publish different twitter datasets which are found in the below dataset which can be used for testing and training.

- https://www.isi.edu/~lerman/downloads/twitter/twitter2010.html

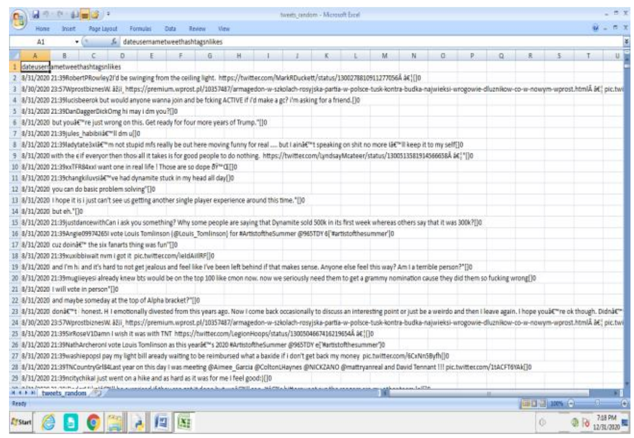

Fig 4: Scrapping tweets form the online tweets

V. ALGORITHMS USED

5.1 Random Forest

Random forest is a supervised learning algorithm. Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset.

Working of Random Forest Algorithm:

Step1 : First, start with the selection of random samples from a given dataset.

Step2 : Following that, this algorithm will create a decision tree for each sample. The forecast result from each decision tree will then be obtained.

Step 3: Each projected outcome will be voted on in this round.

Step 4: Finally, choose the prediction result with the most votes as the final forecast result.

5.2 KNN Model

- The KNN algorithm implies that items that are similar are close together. To put it another

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

way, related items are close together.

- For example "Birds of a feather flock together."

It can be used to solve both classification and regression problems

5.3 Support Vector Machine

- The Support Vector Machine (SVM) algorithm is a simple yet effective Supervised Machine Learning approach that can be used to create both regression and classification models.The main goal of SVM is to construct the hyperplane and divide the So that datamining and categorization may be done successfully, divide the dataset into separate groups. SVM can be applied to real-world problems such as detecting tension in tweets.

Convolutional neural network model:

- CNNs are regularized versions of multilayer perceptrons.
- • Fully linked networks usually refer to multilayer perceptrons, in which each neuron in one layer is connected to all neurons in the next layer.
- The input layer, hidden layer, and output layer are the three types of layers in CNN.The input from input layer is feed into hidden layer
- There can be many hidden layers depending upon our model and data size
- Each layer's output is calculated by multiplying the preceding layer's output by the learnable weights of the layers in a matrix.

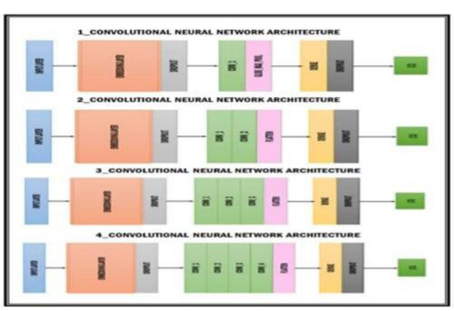

Fig 5: Layered Convolution Neural Networks for stress classification

1-CNN is the first layer composed of 1-convculation.The layer and the temporal convolution operation are carried out with the kernel size and 0-padding in this layer. The functionality of relu activation is then applied to the layer and the pooling process is performed. To decrease the dimensionality of the data, global maximum pooling is carried out here. For the fully connected convolution layer that produces the single value, the pooling output is supplied as an input. The sigmoid activation function is performed with this

VI.RESULTS AND EVOLUTION METRICS

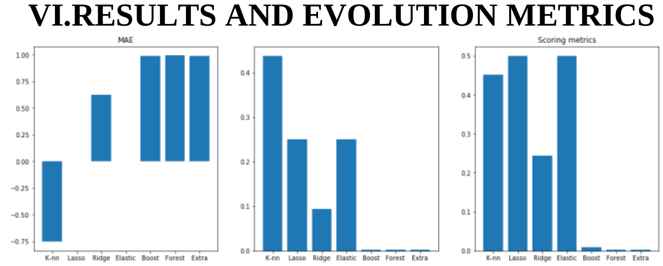

Fig 6 Scoring metrics like Mean absolute error and scoring metrics results graph

	model	mse	mae	score
0	K-nn	0.4	0.45	-75.05
1	Lasso	0.2	0.50	-0.02
2	Ridge	0.1	0.24	62.26
3	Elastic	0.2	0.50	-0.02
4	Boost	0.0	0.01	99.17
5	Forest	0.0	0.00	99.19
6	Extra	0.0	0.00	99.15

Fig 7 scoring metrics results using machine learning approach

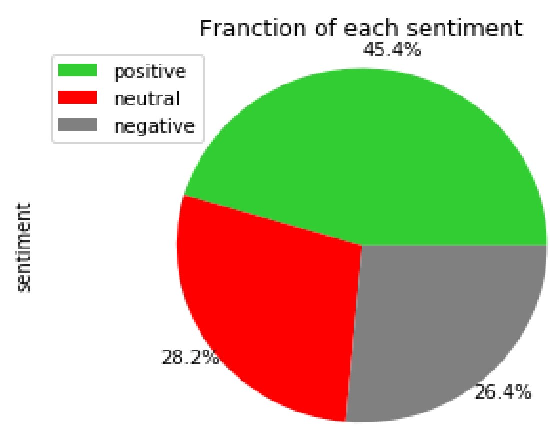

Fig .8:-fraction of each tweet for stress classification

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

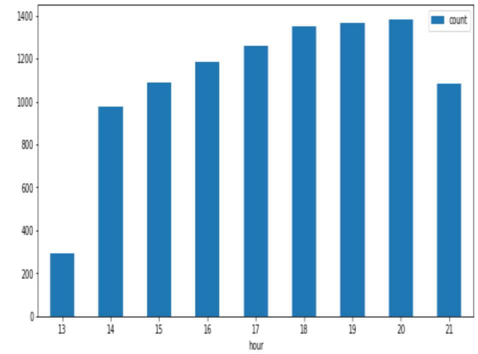

Fig 9-frequency plot of tweets

Out[16]:

	hour	count
0	13	295
1	14	977
2	15	1087
3	16	1186
4	17	1258
5	18	1353
6	19	1367
7	20	1383
8	21	1085

Fig. 10:-Hours count of each tweet users

Out[12]:

	date	username	tweet	hashtags	nlikes	sentiment
0	2020-08-31 21:39:04	RobertPRowley2	I'd be swinging from the ceiling light. https...	[]	0	positive
1	2020-08-30 23:57:17	Wprostbiznes	W. ➡ https://premium.wprost.pl/10357487/armag...	[]	37	neutral
2	2020-08-31 21:39:04	lucisbeer	ok but would anyone wanna join and be fcking A...	[]	0	positive
3	2020-08-31 21:39:04	DanDaggerDick	Omg hi may i dm you?	[]	0	neutral
4	2020-08-31 21:39:04	realMarkMoney	I love your work, but you're just wrong on thi...	[]	0	positive

Fig. 11: Stress tweet classification

```
Model: "sequential_5"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding_5 (Embedding)      (None, 280, 300)          12325200
_________________________________________________________________
global_average_pooling1d_1 ( (None, 300)               0
_________________________________________________________________
dense_6 (Dense)              (None, 24)                7224
_________________________________________________________________
dense_7 (Dense)              (None, 1)                 25
=================================================================
Total params: 12,332,449
Trainable params: 12,332,449
Non-trainable params: 0
```

Fig. 12: Total parameter for fine tuning

```
Epoch 20/30
542/542 [==============================] - 70s 129ms/step - loss: 0.0048 - accuracy: 0.9986 - val_loss: 0.1165 - val_
accuracy: 0.9720
Epoch 21/30
542/542 [==============================] - 70s 130ms/step - loss: 0.0037 - accuracy: 0.9988 - val_loss: 0.1604 - val_
accuracy: 0.9720
Epoch 22/30
542/542 [==============================] - 72s 132ms/step - loss: 0.0039 - accuracy: 0.9988 - val_loss: 0.1277 - val_
accuracy: 0.9709
Epoch 23/30
542/542 [==============================] - 70s 128ms/step - loss: 0.0042 - accuracy: 0.9987 - val_loss: 0.1325 - val_
accuracy: 0.9720
Epoch 24/30
542/542 [==============================] - 69s 127ms/step - loss: 0.0037 - accuracy: 0.9990 - val_loss: 0.1324 - val_
accuracy: 0.9699
Epoch 25/30
542/542 [==============================] - 69s 128ms/step - loss: 0.0024 - accuracy: 0.9994 - val_loss: 0.1412 - val_
accuracy: 0.9725
Epoch 26/30
542/542 [==============================] - 70s 130ms/step - loss: 0.0032 - accuracy: 0.9991 - val_loss: 0.1968 - val_
accuracy: 0.9683
Epoch 27/30
542/542 [==============================] - 72s 133ms/step - loss: 0.0027 - accuracy: 0.9991 - val_loss: 0.1646 - val_
accuracy: 0.9714
Epoch 28/30
542/542 [==============================] - 72s 133ms/step - loss: 0.0029 - accuracy: 0.9992 - val_loss: 0.1810 - val_
accuracy: 0.9424
Epoch 29/30
542/542 [==============================] - 76s 140ms/step - loss: 0.0031 - accuracy: 0.9989 - val_loss: 0.1997 - val_
accuracy: 0.9330
Epoch 30/30
542/542 [==============================] - 75s 138ms/step - loss: 0.0034 - accuracy: 0.9990 - val_loss: 0.1471 - val_
accuracy: 0.9699
```

Fig 13: Training and validation loss Result

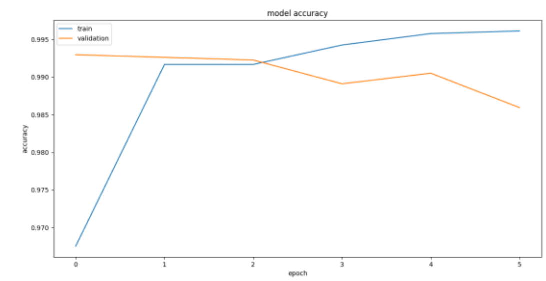

Fig. 14: training and validation accuracy graph for CNN ensemble model with DT

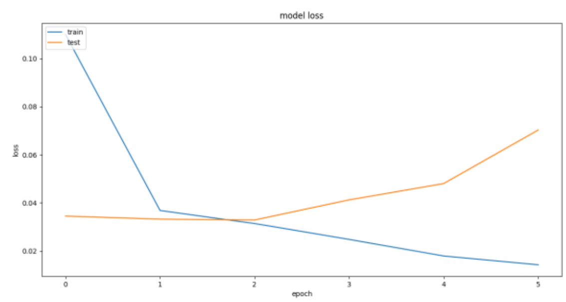

Fig 15: Training and validation loss graph for CNN ensemble model with DT

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1981
1	1.00	1.00	1.00	1872
accuracy			1.00	3853
macro avg	1.00	1.00	1.00	3853
weighted avg	1.00	1.00	1.00	3853

Fig. 16: Evolution Metrics Results

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

Classifers Used	Accuracy
Support Vector Machine	93.6%
Random Forest	46%
KNN	81%
Decision Tree	98.5%

Fig 17 : Comparative Result Analysis

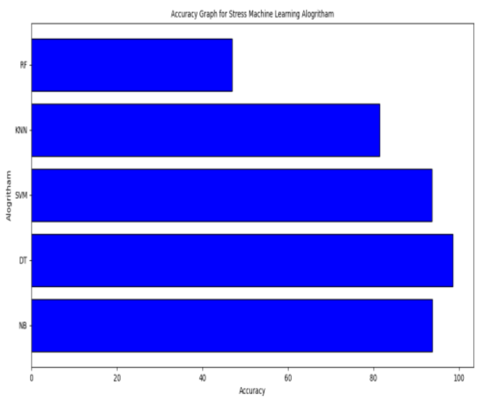

Fig 18: Comparative Result Analysis graph

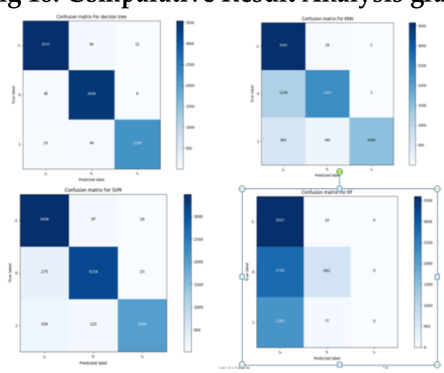

Fig 19: Confusion matrix report for all machine learning algorithms

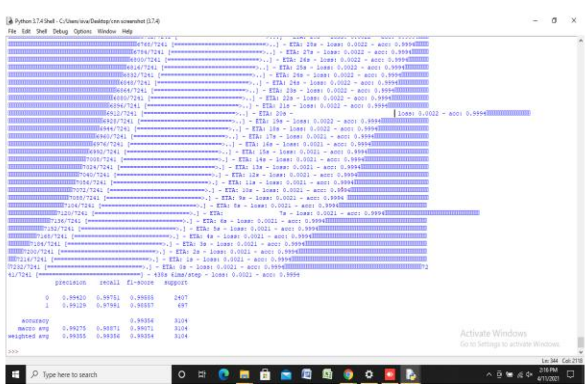

Fig. 20: CNN ensemble model with DT Meta-Classifier with 99.35 % accuracy

VII.CONCLUSION

This research focuses on detecting stress in users by deducing their tweeting patterns and user behaviors from twitter data using an reassembling model based on the features of the user's above-mentioned qualities. Running this model allows us to detect and assist stressed individuals. This research is based on a data collection that was scraped directly from Twitter API utilizing emphasized keywords. The experimental results shows the comparison of Machine Learning Algorithms and it tells that Decision Tree algorithm accuracy (98.5%) is best among Random forest (46%), support vector machine (93.6%), KNN (81%) and Decision Tree (98.5%). Hence it was selected for the meta-classifier in the CNN ensemble model which gave us an accuracy of 99.35%

VIII.REFERENCES

[1] Ben Verhoeven, Walter Daelemans, and Barbara Plank. Twisty: A multilingual twitter stylometry corpus for gender and personality profiling. In Proceedings of the Tenth International Conference on Language Resources and Evaluation LREC, pages 1632–1637, 2016

[2] Andrey Bogomolov, Bruno Lepri, Michela Ferron, Fabio Pianesi, and Alex Pentland. Daily stress recognition from mobile phone data, weather conditions and individual traits. In ACM International Conference on Multimedia, pages 477–486, 2014.

[3] Chris Buckley and Ellen M Voorhees. Retrieval evaluation with incomplete information. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, pages 25–32, 2004.

[4] Xiaojun Chang, Yi Yang, Alexander G Hauptmann, Eric P Xing, and Yao-Liang Yu. Semantic concept discovery for large-scale zero-shot event detection. In Proceedings of International Joint

Conference on Artificial Intelligence, pages 2234–2240, 2015.

[5] Wanxiang Che, Zhenghua Li, and Ting Liu. Ltp: A chinese language technology platform. In Proceedings of International Conference on Computational Linguistics, pages 13–16, 2010.

[6] Chih chung Chang and Chih-Jen Lin. Libsvm: a library for support vector machines. Acm transactions on intelligent systems and technology, 2(3):389–396, 2001.

[7] Dan C Ciresan, Ueli Meier, Jonathan Masci, Luca Maria Gambardella, and Jurgen¨ Schmidhuber. Flexible, high performance convolutional neural networks for image classification. In Proceedings of International Joint Conference on Artificial Intelligence, pages 1237– 1242, 2011.

[8] Sheldon Cohen and Thomas A. W. Stress, social support, and the buffering hypothesis. Psychological Bulletin, 98(2):310–357, 1985.

[9] Glen Coppersmith, Craig Harman, and Mark Dredze. Measuring post traumatic stress disorder in twitter. In Proceedings of the International Conference on Weblogs and Social Media, pages 579–582 , 2014.

[10] Rui Fan, Jichang Zhao, Yan Chen, and Ke Xu. Anger is more influential than joy: Sentiment correlation in weibo. PLoS ONE, 2014.

[11] Zhanpeng Fang, Xinyu Zhou, Jie Tang, Wei Shao, A.C.M. Fong, Longjun Sun, Ying Ding, Ling Zhou, , and Jarder Luo. Modeling paying behavior in game social networks. In In Proceedings of the Twenty-Third Conference on Information and Knowledge Management (CIKM'14), pages 411–420, 2014.

[12] Golnoosh Farnadi, Geetha Sitaraman, Shanu Sushmita, Fabio Celli, Michal Kosinski, David Stillwell, Sergio Davalos, Marie Francine Moens, and Martine De Cock. Computational personality recognition in social media. User Modeling and UserAdapted Interaction, pages 1–34, 2016.

[13] Eileen Fischer and A. Rebecca Reuber. Social interaction via new social media: (how) can interactions on twitter affect effectual thinking and behavior? Journal of Business Venturing, 26(1):1–18 , 2011.

[14] Jerome H. Friedman. Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5):1189–1232, 1999.

[15] Rui Gao, Bibo Hao, He Li, Yusong Gao, and Tingshao Zhu. Developing simplified chinese psychological linguistic analysis dictionary for microblog. pages 359–368, 2013.

[16] Johannes Gettinger and Sabine T. Koeszegi. More Than Words: The Effect of Emoticons in Electronic Negotiations.

[17] Jennifer Golbeck, Cristina Robles, Michon Edmondson, and Karen Turner. Predicting personality from twitter. In Passat/socialcom 2011, Privacy, Security, Risk and Trust, pages 149–156, 2011.

[18] Mark S. Granovetter. The strength of weak ties. American Journal of Sociology, 1973.

[19] Quan Guo, Jia Jia, Guangyao Shen, Lei Zhang, Lianhong Cai, and Zhang Yi. Learning robust uniform features for cross-media social data by using cross autoencoders. Knowledge Based System, 102:64– 75, 2016.

[20] David W. Hosmer, Stanley Lemeshow, and Rodney X. Sturdivant. Applied logistic regression. Wiley series in probability and mathematical statistics, 2013.

[21] Sung Ju Hwang. Discriminative object categorization with external semantic knowledge. 2013.

[22] Sepandar D. Kamvar. We feel fine and searching the emotional web. In
In Proceedings of WSDM, pages 117–126, 2011.

[23] Herbert C. Kelman. Compliance, identification, and internalization: Three

processes of attitude change. general information, 1(1):51–60, 1958.
[24] Shigenobu Kobayashi. The aim and method of the color image scale.
Color research & application, 6(2):93–107, 1981.
[25] Novak P Kralj, J Smailovi, B Sluban, and I Mozeti. Sentiment of emojis. Plos One, 10(12), 2015.
[26] Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algorithm. Information Theory, IEEE Transactions on, 47(2):498–519, 2001.
[27] Yann LeCun and Yoshua Bengio. Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361, 1995.
[28] Kathy Lee, Ankit Agrawal, and Alok Choudhary. Real-time disease surveillance using twitter data: demonstration on flu and cancer. In Proceedings of ACM SIGKDD international conference on Knowledge discovery and data mining, pages 1474–1477, 2013.
[29] 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 proceedings of IEEE