ISSN 2319-2720
Volume 7, No.2, April – June 2018
International Journal of Computing, Communications and Networking
Available Online at http://www.warse.org/ijccn/static/pdf/file/ijccn17722018.pdf
https://doi.org/10.30534/ijccn/2018/17722018

Detecting and Analyzing Stress Based on Social Interactions in Social Networks

Namratha M M[1],Nethravathi R[2], Navya J[3],Nayana D J[4] , Kiran M[5]
[1]Student, EWIT, India, namratha.malur@gmail.com
[2]Student, EWIT, India, nethraranga.136@gmail.com
[3]Student, EWIT, India, navyaneha31@gmail.com
[4]Student, EWIT, India, nayanadj07@gmail.com
[5]Assistant Professor, EWIT, India, EWIT, India, kiranm@ewit.edu

ABSTRACT

Traditional mental health studies [3] rely on information primarily collected through personal contact with a health care professional. Recent work has shown the utility of social media data for studying depression, but there have been limited evaluations of other mental health conditions. We consider post traumatic stress disorder (PTSD), a serious condition that affects millions worldwide, with especially high rates in military veterans. We also present a novel method to obtain a PTSD classifier for social media using simple searches of available Twitter data ,a significant reduction in training data cost compared to previous work. We demonstrate its utility by examining differences in language use between PTSD and random individuals, building classifiers to separate these two groups and by detecting elevated rates of PTSD at and around U.S. military bases using our classifiers.

Key words : Stress detection, factor graph model, micro-blog, social media, healthcare, social interaction.

1. INTRODUCTION

With the success of many large-scale online social networks [4], such as Facebook and Twitter, the social networks are playing a very important role as a medium for the spread of information, ideas, and influences. In social networks, with the power of ―word of mouth‖, a new idea or innovation can influence a large population in a very short period, but may also die out quickly. To understand the underlying dynamics of the social networks, it is very important to know how people influence with each other.

Most algorithms attempt to identify the polarity of sentiment in text: positive, negative or neutral. Whilst for many applications this is sufficient, texts often contain a mix of positive and negative sentiment and for some applications it is necessary to detect both simultaneously and also to detect the strength of sentiment expressed. Programs to monitor sentiment in online communication, perhaps designed to identify and intervene when inappropriate emotions are used or to identify at-risk users would need to be sensitive to the strength of sentiment expressed and whether participants were appropriately balancing positive and negative sentiment.

Figure1. Sample tweets from Sina Weibo. In each tweet, the top part is tweet content with text and an image; the bottom part shows the social interactions of tweets where there are multiple indicators of stress: mentions of ’busy’ and ’stressed’, ’working overtime’, ’failed the exam’, ’money’ and a stressed emoticon.

2. RELATED WORK

2.1 Character recognition

The character recognition problem [2] has been studied extensively for decades. For the natural scene text, the methods based on feature descriptors has been widely used recently. Classical HOG based features outperform better those other features. Multi-scale features called stroke lets are learned to describe the structure of the characters and a low-dimensional attribute method is proposed to encode the characters. Despite being improved, the discriminative power of these feature representation is still limited. They proposed using the unsupervised learned Convolutional Neural Network (CNN) features to detect and recognize the texts.

2.2. Research on leveraging social interactions for social media analysis.

Social interaction is one of the most important features of social media platforms. Now many researchers are focusing on leveraging social interaction information to help improve the effectiveness of social media analysis. [2] Analyzed the relationships between social interactions and users’ thinking and behaviors, and found out that Twitter-based interaction can trigger effectual cognitions. Leveraged comments on Flickr to help predict emotions expressed by images posted on Flickr. However, these work mainly focused on the content of social interactions, e.g., textual comment content, while ignoring the inherent structural information like how users are connected.

2. PROPOSED WORK

Inspired by psychological theories, we first define a set of attributes for stress detection from tweet-level and user-level aspects respectively: 1) tweet-level attributes from content of user's single tweet, and 2) user-level attributes from user's weekly tweets. The tweet-level attributes are mainly composed of linguistic, visual, and social attention (i.e., being liked, retweeted, or commented) attributes extracted from a single-tweet's text, image, and attention list. The user-level attributes however are composed of: (a) posting behavior attributes as summarized from a user's weekly tweet postings; and (b) social interaction attributes extracted from a user's social interactions with friends. In particular, the social interaction attributes can further be broken into:(i)social interaction content attributes extracted from the content of users' social interactions with friends; and (ii) social interaction structure attributes extracted from the structures of users' social interactions with friends.

The contribution of this paper as follows:

• We propose a unified hybrid model integrating CNN with FGM to leverage both tweet content attributes and social interactions to enhance stress detection.

• We build several stressed-twitter-posting datasets by different ground-truth labeling methods from several popular social media platforms and thoroughly evaluate our proposed method on multiple aspects.

• We carry out in-depth studies on a real-world largescale dataset and gain insights on correlations between social interactions and stress, as well as social structures of stressed users.

3. PROBLEM FORMULATION

Let V be a set of users on a social network, and let |V| denote the total number of users. Each user vi∈ V posts a series of tweets, with each tweet containing text, image, or video content; the series of tweets contribute to users social interactions on the social network.

Definition1. Stress state: The stress state of user yvi ∈ V at time t is represented as a triple $(y,v_{i},^{t})$, or briefly y^{t}_{i}. In the study, a binary stress state y^{t}_{i}∈{ 0,1} is considered, where yti =1indicates that user vi is stressed at time t, and y_{i}^{t} =0 indicates that the user is non-stressed at time t, which can be identified from specific expressions in user tweets or clearly identified by user himself, as explained in the experiments. Let Y t be the set of stress states of all users at time t.

Definition 2. Time-varying user-level attribute matrix: Each user in V is associated with a set of attributes A. Let X^{t} be a |V|×|A| attribute matrix at time t, in which every row x^{t}_{i} corresponds to a user, each column corresponds to an attribute, and an element $x^{t}_{i,j}$ is the jth attribute value of user vi at time t.

Definition 3. Time-varying attribute-augmented network: An attribute-augmented network at time t is comprised of four elements, including 1) a user set V^{t}, 2) an edge set E^{t}, 3)a user-level attribute matrix set X^{t},and 4)a stress state set for all users Y^{t} at time t, denoted as Gt =$(V^{t},E^{t},X^{t},Y^{t})$.

4. ATTRIBUTE CATEGORIZATION AND DEFINITION

4.1 Tweet-level Attributes

Tweet-level attributes describe the linguistic and visual content, as well as social attention factors (being liked, commented, and retweeted) of a single tweet.

For linguistic attributes, we take the most commonly used linguistic features in sentiment analysis research. Specifically, we first adopt LTP — A Chinese Language Technology Platform — to perform lexical analysis, e.g., tokenize and lemmatize, and then explore the use of a Chinese LIWC dictionary — LIWC2007 , to map the words into positive/negative emotions. LIWC2007 is a dictionary which categorizes words based on their linguistic or psychological meanings, so we can classify words into different categories, e.g. positive/negative emotion words, Degree adverbs. We have also tested other linguistic resources including NRC5 and HowNet6, and found that the performances were relatively the same, so we adopted the commonly used LIWC2007 dictionary for experiments.

4.2 User-Level Attributes

Compared to tweet-level attributes extracted from a single tweet, user-level attributes are extracted from a list of user's tweets in a specific sampling period. We use one week as the sampling period in this paper. On one hand, psychological stress often results from cumulative events or mental states. On the other hand, users may express their chronic stress in a series of tweets rather than one. Besides, the aforementioned social interaction patterns of users in a period of time also contain useful information for stress detection. Moreover, as aforementioned, the information in tweets is limited and sparse, we need to integrate more complementary information around tweets, e.g., users' social interactions with friends. Thus, appropriately designed user-level attributes can provide a macro-scope of a user's stress states, and avoid noise or missing data. Here, we define user-level attributes from two aspects to measure the differences between stressed and non-stressed states based on users' weekly tweet postings: 1) user-level posting behavior attributes [29] from the user's weekly tweet postings; and 2) user-level social interaction attributes from the user's social interactions beneath his/her weekly tweet postings.

5. MODEL FRAMEWORK

Two challenges exist in psychological stress detection. 1) How to extract user-level attributes from user's tweeting series and deal with the problem of absence of modality in the tweets? 2) How to fully leverage social interaction, including interaction content and structure patterns, for stress detection?

5.1 Architecture

Figure shows the architecture of our model. There are three types of information that we can use as the initial inputs, i.e., tweet-level attributes, user-level posting behavior attributes, and user-level social interaction attributes, whose detailed computation will be described later. We address the solution through the following two key components:

• First, we design a CNN with cross autoencoders (CAE) to generate user-level interaction content attributes from tweet-level attributes. The CNN has been found to be effective in learning stationary local attributes for series like images and audios.

• Then, we design a partially-labeled factor graph (PFG) to incorporate all three aspects of user-level attributes for user stress detection. Factor graph model has been widely used in social network modeling. It is effective in leveraging social correlations for different prediction tasks.

The model consists of two parts. The first part is a CNN. The second part is a FGM. The CNN will generate user-level content attributes by convolution with CAE filters as input to the FGM. Take the user labeled with a red star as example. Tweet-level attributes of the user are processed through a convolution with CAE to form the user-level content attributes. The user-level attributes are denoted by x^t_i in the left box. Every x^t_i contains three aspects: user-level content attributes, user-level posting behavior attributes, and user-level social interaction attributes. Data of other users follows the same route. In the FGM, attribute factors connect user-level attributes to corresponding stress states. Social factors connect the stress state of different users. Dynamic factors connect stress state of a user over time. The output of the user's user-level stress state at time t is y^t_1 as highlighted in red, which actually denotes the stress state of the user in weekly period in this paper

5.2 LEARNING LATENT CORRELATIONS BETWEEN TWEET'S CONTENT AND SOCIAL INTERACTIONS

As the social correlation between users and time-dependent correlation are hard to be modeled using classic classifiers such as SVM, we use a partially-labeled factor graph model (PFG), which was first proposed, to incorporate social interactions and tweets' content for learning and detecting user-level stress states. We define an objective function by maximizing the conditional probability of users' stress states Y given a series of attribute-augmented networks.

Attribute factor: We use this factor $f(x^t_i, y^t_i)$ to represent the correlation between user vi's stress state at time t and her/his attributes x^t_i.

Dynamic factor: We use this factor $f(y^t_i, y^{t+1}_i)$ to represent the time correlation between user v_i's stress state at time t and t +1.

Social factor: We use social factor $g(y_c)$ (where $e = (v^t_i, v^t_i, c) \in Et$) to represent the correlation between userv_i and v_i's stress states according to c at time t.

Learning: Learning the predictive model is to estimate a parameters configuration $\theta = (\alpha, \{\beta c\}, \gamma)$ from the partially labeled dataset and to maximize the log-likelihood objective function Eq. 7, i.e., $\theta * = \mathrm{argmax}\theta O(\theta)$. For optimization, we adopt a gradient decent method. Specifically, we derive the gradients with respect to each parameter in our objective function.

ALGORITHM: Learning and inference by factor graph.

Input: a series of time-varying attribute augmented network G with stress states on some of the user nodes, learning rate n;

Output: parameter value $\theta = \{\alpha, \{\beta c\},$ and full stress state vector Y;

Randomly initialize Y;

Initialize model parameter ;

repeat

Compute gradient $\nabla\alpha, \nabla\beta c, \nabla$

Update

Update

Update

until convergence;

5.3 Dataset Collection

To conduct observations and evaluate our successive model, we first collect a set of datasets using different labeling methods, which are listed as following:

Dataset DB1: It is a challenge to construct a dataset with reliable ground truth labels from large-scale noisy social media data. The data crawled from social platforms is usually massive, thus manual labeling methods are not feasible due to the uncontrollable cost and quality. To solve this problem, we employed a sentence pattern labeling method to automatically extract labeled data from the crawled largescale social media data. We first crawled 350 million tweets data via Sina Weibo's REST APIs9 from Oct. 2009 to Oct. 2012. Sina weibo, as the biggest microblog website in China, provides users an open online platform for information sharing, communication and obtaining. Similar to Twitter and Facebook, users on Sina Weibo can post contents with multiple modalities, including text, image, social action(retweet, comment, favorite), video and etc.

In this way, we collected over 19,000 weeks of tweets that are labeled as stressed, and over 17,000 weeks of non stressed users' tweets. There are 492,676 tweets from 23,304 users in total. We use this dataset for experiments, analysis and further in-depth studies, which is represented by DB1 in this paper.

Dataset DB2: We verified the reliability of the above ground truth labeling method through dataset DB2 in Table 4. It is a small dataset collected from the users who have shared the score of a psychological stress scale PSTR10 designed by psychologists via Weibo. Guided by the rules of the PSTR scale, a user is taken as stressed when the score is larger than 80, otherwise non-stressed. We thus crawled the scores posted by users, and used the scores as ground truth label for the set of tweets in +-3-day window.

Dataset DB3 and DB4: To further test our method, we collected two more datasets from Tencent Weibo (DB3) and Twitter (DB4). They are again labeled using the sentence pattern labeling method as described above for DB1. In particular, as social platforms of different languages, Weibo and Twitter have many differences. Thus, experiments on Twitter can validate the universality of our method.

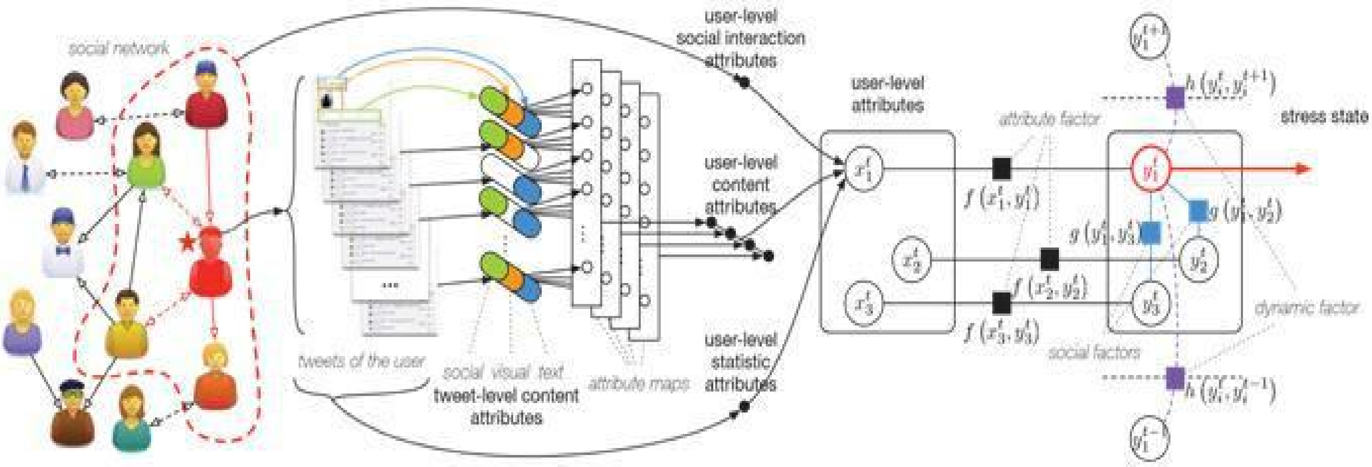

Figure 2. Architecture of the system

5.4 PROPOSED ALGORITHM

JCN (word1,word2)

Step 1: Start

Step 2: Construct T1 which is co related words to the word1 Construct T2 which is co related words to the word2 Step 3: Lowest Common Subsumer(s)

=argmax(IC(subsumer(T1,T2)))

LCS={subsumer(T1[1], T2[1]) } = { wordx };

Step 4: Find IC(Information Content) for wordx;

Find IC(Information Content) for word1;

Find IC(Information Content) for word2;

Step 5: Calculate Distance

Distance (word1, word2) = (IC(word1) + IC(word2) - 2 * IC(LCS))

Step 6: Score = 1 / Distance (word1, word2)

Step 7: return score;

Step 8: Stop

6. STUDIES OF SOCIAL INTERACTION

We have presented the experimental results on stress detection in the previous section, while in the setting of social networks, it would be helpful to further analyze how a user's stress status is developed and how they affect each other. To do so, we try to conduct several studies on DB1 to offer insights on how social interactions contribute to user stress and the task of stress detection from the following aspects:

(1) Content. How are users' social interaction contents (e.g., language used) related to users' stress states?

(2) Structure. Compared to non-stressed users, do stressed users show different structural diversity patterns when they behave in social networks? Do differences of social influence and strong/weak ties exist between stressed and non-stressed users?

6.1 Content

Content of social interaction refers to the content of tweets' comments and retweets, including text, emoticons, and punctuation marks. Based on a widely used psychological dictionary LIWC2007 [40], we extract emotional words from the interaction content of tweets, and categorize the extracted words into corresponding groups defined in LIWC2007. We compare the frequencies of different word categories between stressed and non-stressed users.

6.2 Structure

To examine structure properties (i.e., social influence and strong/weak tie) of (non)stressed users, we use risk ratio (RR) to measure the correlation between users' stress states and different structural attributes. Risk ratio is an effective measurement widely used in the statistical analysis and relevant fields. The risk ratio of a stressed state, associated with a structural attribute a, is calculated as follows:

$$\mathrm{RR}(a) = \frac{\mathrm{P(stressed\ user\ has\ attribute\ a)}}{P(stressed\ user\ does\ not\ have\ attribute\ a)}$$

6.2.1 Structural Diversity

We are interested in whether stressed and non-stressed users have any structural difference in respective friends' connection. In sociology, social structure refers to a society's framework, consisting of various relationships among people, as well as groups that direct and set limits on human behaviors. In social networks, direct connections (following or followed) of users that interact with each other via comments and retweets also form a kind of social structure. For this in-depth study, we select top four users with the most frequent interactions from users' weekly tweet postings, where four is adopted because this is the minimum number of nodes required to produce structural combinations (10 combinations), so as to calculate the probability of each combination, and incorporating more nodes would make the calculation combinatorial expensive.

6.2.2 Social Influence

Social influence is an important factor that governs the dynamics of social networks. The principle of social influence suggests that users tend to change their behaviors to match their friends' behaviors. In this study, we try to examine whether users' stress states will be influenced by their neighbors' states by looking at the probability of a user's stress state when he/she has different types of relationshipswithotherstressedusers.Asforthestressstate labeling, all users including friends are labeled using the sentence pattern method described in previous section.

6.2.3 Strong/Weak Tie

Strong/Weak Tie is one of the most basic principles in social network theories. We classify the constructed social relationships into strong or weak ties by the number of times that two users interact with each other via comment, @-mention, retweet, or like in a week. In our work, we tried different values for the threshold and finally chose three by cross-validation. If two users interact with each other more than three times, we call the relationship a strong tie, and otherwise a weak tie. This definition of user ties is adopted as the standard treatment in the research of social network analysis, so as to capture the most recent user relationships in a shifting environment. Figure 8(b) illustrates the results. We can see that strong ties indeed have strong influence on users' stress states, and the influence of weak ties is relatively weak. For example, when a user has three stressed strong-tie connections, the probability that the user will become stressed increases to 13%, more than twice as high as for a user with three stressed weak-tie connections.

7. EXPERIMENTAL RESULTS

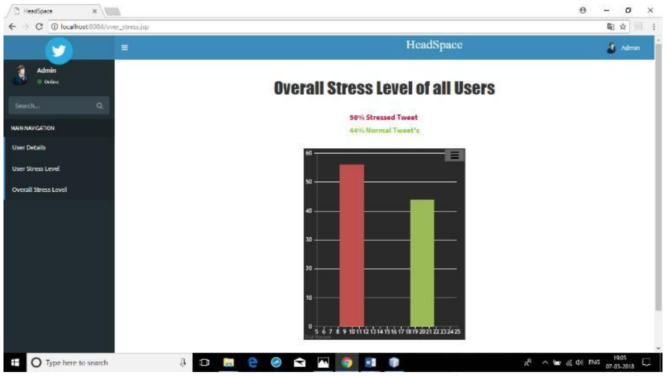

Figure 3. Overall stress level of all users

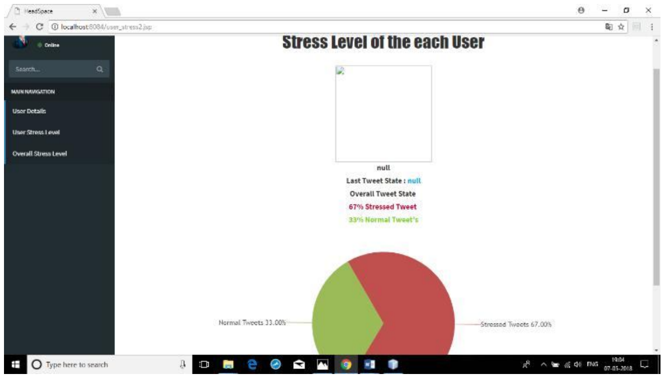

Figure 3. S tress level of each users

8. CONCLUSION

Based on the experimental results and analyses we know that:
1) users' stress states are not only revealed in their own tweets, but also affected by the contents of their social interactions, including commenting on and retweeting others' tweets; and 2) users' stress states are revealed by the structure of their social interactions, including structural diversity, social influence, and strong/weak ties. These insights quantitatively prove the necessity and effectiveness of combining social interactions for stress detection.

In this paper, we presented a framework for detecting users' psychological stress states from users' weekly social media data, leveraging tweets' content as well as users' social interactions. Employing real-world social media data as the basis, we studied the correlation between user' psychological stress states and their social interaction behaviors. To fully leverage both content and social interaction information of users' tweets, we proposed a hybrid model which combines the factor graph model (FGM) with a convolutional neural network (CNN).

In this work, we also discovered several intriguing phenomena of stress. We found that the number of social structures of sparse connection (i.e. with no delta connections) of stressed users is around 14% higher than that of non-stressed users, indicating that the social structure of stressed users' friends tend to be less connected and less complicated than that of non-stressed users. These phenomena could be useful references for future related studies.

REFERENCES

[1]Mike Thelwall, Kevan Buckley, Georgios Paltoglou and Di Cai. *Sentiment Strength Detection in Short Informal Text.* Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK

[2] Xinhao Liu, TakahitoKawanishi, Xiaomeng Wu and KunioKashino.*Scene Text Recognition with CNN Classifier and WFST-based Word Labeling.*NTT Corporation, Kanagawa, Japan 243-0198.

[3]GlenA.Coppersmith,CraigT.Harman andMarkH.Dredze. *Measuring Post Traumatic Stress Disorder in Twitter*. Human Language Technology Center of Excellence Johns Hopkins University Balitmore, MD, USA.
[4]Chi Wang, Jie Tang, Jimeng Sun and Jiawei Han. *Dynamic Social Influence Analysis through Time-dependent Factor Graphs.*University of Illinois at Urbana-Champaign, USA Tsinghua University, China IBM TJ Watson Research Center, USA.
[5]Chihchung Chang and Chih-Jen Lin. Libsvm: a library for supportvector machines. *ACM TRANSACTIONS ON INTELLIGENTSYSTEMS AND TECHNOLOGY*, 2(3):389–396, 2001.
[6] Dan C Ciresan, Ueli Meier, Jonathan Masci, Luca Maria Gambardella, and J¨urgenSchmidhuber. Flexible, high performanceconvolutional neural networks for image classification. In *Proceedingsof International Joint Conference onArtificial Intelligence*, pages1237–1242, 2011.
[7] Sheldon Cohen and Thomas A. W. Stress, social support, and thebuffering hypothesis. *Psychological Bulletin*, 98(2):310–357, 1985.
https://doi.org/10.1037/0033-2909.98.2.310
[8] Glen Coppersmith, Craig Harman, and Mark Dredze. Measuringpost traumatic stress disorder in twitter. In *Proceedings of theInternational Conference on Weblogs and Social Media*, pages 579–582,2014.
[9]Rui Fan, Jichang Zhao, Yan Chen, and Ke Xu. Anger is moreinfluential than joy: Sentiment correlation in weibo. *PLoS ONE*,2014.
https://doi.org/10.1371/journal.pone.0110184
[10]Zhanpeng Fang, Xinyu Zhou, Jie Tang, Wei Shao, A.C.M. Fong,Longjun Sun, Ying Ding, Ling Zhou, , and Jarder Luo. Modelingpaying behavior in game social networks. In *In Proceedings of theTwenty-Third Conference on Information and Knowledge Management(CIKM'14)*, pages 411–420, 2014.
[11]GolnooshFarnadi, GeethaSitaraman, ShanuSushmita, FabioCelli, Michal Kosinski, David Stillwell, Sergio Davalos,Marie Francine Moens, and Martine De Cock. Computationalpersonality recognition in social media. *User Modeling and User-Adapted Interaction*, pages 1–34, 2016.
[12] Eileen Fischer and A. Rebecca Reuber. Social interaction via newsocial media: (how) can interactions on twitter affect effectualthinking and behavior? *Journal of Business Venturing*, 26(1):1–18,2011.
https://doi.org/10.1016/j.jbusvent.2010.09.002

Ms.Namratha.M.Mis pursuing her 8thsem B.E in Computer Science and Engineering in East West Institute of Technology,Bangalore,India. Her area of Interest are Machine Learning and Data Mining

Ms.Nethravathi.Ris pursuing her 8thsem B.E in Computer Science and Engineering in East West Institute of Technology,Bangalore,India. Her area of Interest are Machine Learning and Data Mining

Ms.Navya.Jis pursuing her 8thsem B.E in Computer Science and Engineering in East West Institute of Technology,Bangalore,India. Her area of Interest are Machine Learning and Data Mining

Ms.Nayana.D.Jis pursuing her 8thsem B.E in Computer Science and Engineering in East West Institute of Technology,Bangalore,India. Her area of Interest are Machine Learning and Data Mining

Mr.Kiran.M is currently working as Asst.Professor in the department of Computer Science and Engineering, East West Institute of Technology(VTU), Bangalore, India. His Qualifications are B.E.,M.Tech. His area of research are Wireless Sensor Networs, Computer Networks,Cloud Computing, Internet of Things and WebTechnology.