# ESTIMATING TOPIC-BASED PUBLIC ANXIETY IN SOCIAL MEDIA USING FUZZY TREE

**Ravi Varshney, Sanidhya Varshney, Lalit Saini, Prabuddha Kumar**

Department of Computer Engineering & Application, GLA University

***Abstract-*** In this day and age everyone uses social media and some of these messages, posts or comments can cause anxiety or depression in the user. Although the assessment of personal stress has been well studied, there has not been much work to assess social stress in the right population, groups, that could be used to describe mental illness in the community. However, to measure social stress we cannot measure the stress level of an individual because the following must be taken into account:-

(1) The effect of correlations of individual stress level of each group (composite).

(2) Discussion-based content (content summary) showing the stress situation in the community. In this project, we set out to investigate social anxiety tests on social networking (TSNC)-based issues.

We developed an assessment system that converts TSNC stress scores to scores in the [0, 1] range. We developed a tiered model to measure individual stress scores using the Intervention model. We developed a fuzzy model to evaluate the stress score using users' phone calls and wrote a tree model (MC-Tree) to calculate the stress score for TSNC from the subjects.

Here we use fuzzy logic to predict issues of public concern. We focus on content and design using fuzzy logic.

Our main goal is to use fuzzy logic to predict content-based analysis. We see the results in the range [0, 1]. The performance of our model has been validated by real and repeated research on the real-world Twitter dataset.

**Index Terms :** Public Anxiety, Topic Based Public Anxiety, Anxiety Using Fuzzy Logic, Anxiety In Social Network, Social Network, Anxiety Evaluation, Topic-based Community.

## I. INTRODUCTION

Social network users form different communities by participating in discussions on different topics. Social media can affect people with social anxiety disorder[1]. In some cases, this can lead to an increased fear of being judged based on choices or appearance. It arises from comparing the emotional state of online communication, which can have a huge impact on society.

I am distracted by social media at work. The overwhelming need to share information on social media[2] sites. Feeling extremely stressed or anxious when you can't check your social media status for a period of time. Don't worry about embarrassing or humiliating yourself. Extreme fear of dating or talking to strangers. Fear that others will find you scared. Fear of physical symptoms that may embarrass you. B. Blush, sweating, trembling, trembling voice. Therefore, the study of the emotional state of topic-based social network communities (TSNC)[3], especially fear, has become an important task in social network analysis. Such a community consists of all users participating in discussion of a particular topic. Social network participation platform. Conventional methods use the Self-Rating Anxiety Scale (SAS) to assess individual anxiety levels by asking subjects to complete a questionnaire. In order to monitor users of social networking platforms, it is impractical to ask large numbers of users to complete SAS questionnaires to rate their individual anxiety levels.

Topic-based social network analysis tasks will be important social network analysis where the community of all users will participate in the discussion of a particular topic. Conventional methods are used to assess the Self-Rating Anxiety Scale (SAS) or to assess an individual's anxiety level. According to social influence theory, social relationships form the basis of one network's attitudes or behaviors towards another actor. Social anxiety disorder, also known as social phobia, is a long-lasting, overwhelming fear of social situations. This is a common problem that usually begins in the teenage years. It is very painful and can have a huge impact on your life. In some people, it gets better with age. According to the American Psychiatric Association, nearly 30% of adults suffer from anxiety disorders related to mental and physical strain at some point in their life. " Community public anxiety " refers to community reactions to stressful situations and is reflected in collective attitudes about community focus shaped by the mutual influence of all members. The Relationship Between Social Media and Social Anxiety The ever-expanding world of social media is a network of sites and websites designed to allow individuals to instantly share images and ideas with all of their friends. The popularity of these sites exploded quickly and they are used by billions of people every day. While such websites may seem harmless, these interactive platforms can significantly increase social unrest. are increasing. These social networks are actually harmful to the mental health of many users. In the United States, there are more than 15 million people diagnosed with social anxiety, and social media-related causes are on the rise in today's world.

## II. CONTRIBUTION

We initialize the study of the following question:

How to quantitatively evaluate the public anxiety of a Topic-based Social Network Community (TSNC)[3]?

To answer above question, we cannot simply add up or average all community members' individual anxiety levels tested by a general-purpose SAS. Some people tend to worry a lot while some don't, but all members adjust and obey community norms, and expose their feelings by discussing the topic of their community. For example, preliminary school parents with low individual anxiety levels may become extreme anxious in a community discussing parenting.And the community's public anxiety may even escalate,regardless of all members' general SAS scores. Therefore, we identify two key components contributingto the formation of a TSNC's public anxiety:

(1) The Structural component:-

A community's structure (i.e., relations between members) influences information diffusion and its members' behaviors , . The structural balance theory indicates that a person's sentiment status (positive or negative) is mainly determined by the relations involving this person. In general, negative emotions spread faster and deeper in social network than positive emotions. In a word, emotions are increasingly formed among connected users on social networks. Therefore, relationships among community members play an important role in forming the public anxiety of a TSNC.

(2) The Topical component:-

In a TSNC, members discuss with others to interchange their thoughts and feelings to the specific topic via social network messages. These messages reflect members' emotional states, esp. "anxiety" that we focus in this paper, and make up another key component to evaluate a community's anxiety. Given the interactions embedded in the message posting, forwarding, and commenting processes, simply averaging the anxiety scores of messages is inadequate, because the crowd is no simple summation or averaging of its members.

Our solution. We propose a framework to evaluate public anxiety of a TSNC. For each community, our framework computes its Structural and Topical anxiety scores separately and normalizes a linear combination of the two scores into range [0, 1] as the community's public anxiety score, where larger score corresponds to higher public anxiety.To obtain the Structural anxiety score, we first need to evaluate each member's anxiety level. Based on SAS questionnaire results filled by volunteer social network users, our model is trained to learn the correlation between a single user's anxiety score and his/her social network data (e.g., user profile). Then we devise an iterative cascading algorithm to gradually compute each member's anxiety score considering the influence from the kth connection via forwarding/commenting chains in the community, and aggregate all members' anxiety scores to generate the community's Structural anxiety score.

For the Topical score, we propose a probabilistic model to measure the topic-specific anxiety embedded in a TSNC's messages along forwarding/commenting chains. We design a tree structure (MC-Tree) and construct a forest of MC-Trees to organize original messages (as roots) and forwarding messages and comments (as child nodes). We then apply an influence-based aggregation to obtain the Topical anxiety score of every tree and summarize all scores in the forest to obtain the TSNC's Topical anxiety score. In practice, computing the anxiety score of social communities on a real-time basis can be expensive or impractical, especially for large communities. Thus we further develop an incremental sampling-based method to use a small subset of the whole community to quickly estimate the public anxiety within certain confidence interval.

## III. LITERATURE REVIEW

Literature studies are the results of reviews from national and international scientific journals that previous researchers have studied.The basic process in determining the decision support and literature studies before determining the selected parameters so that they can be investigated more deeply and further. Previous research using fuzzy ever learned, they analyzed the classification of supply chain risk (SCR) and put forward its index systems on the basis of its many results. As the final, they specified numerical example to analyze the model, the results indicate that the method is not only accommodates the existing fuzzy decision-making methods, but also successfully incorporates the decision preference into the optimization process.The purpose of this study is to define the relationship between technology use of students and their anxiety and aggression level using a fuzzy logic approach. The Technology use is defined by following dimensions :-Habits of technology use ,Social media ,Role of technology in daily life ,Educational use ,Communication use.

We explore attitudes and behaviors toward online social media based on whether one is depressed or not. [4]They conducted faceto-face, semi-structured interviews with 14 active Twitter users, half of whom suffered from depression and half of whom did not. Their result highlight the main differences between the two groups in their perceptions of online social media and their behavior within these systems. [5]In this article, they examine that traditional methods of predicting suicide attempts limit the accuracy and magnitude of risk detection for this risky behavior. So they sought to overcome these limitations by applying machine learning to electronic health records within a large medical database.

In this project [6], we found that many scales were created to measure the stress of relationship between students in various situations, but none of the research on measuring stress on social media platforms. This study explains the development and validity of the Multidimensional Social Anxiety Scale for Social Media Users (SAS-SMU), which can be used to measure high levels of social anxiety caused by the social media platform in students. Relationships This study was carried out in two stages. In the first stage, data collected from 174 students were used to provide evidence for the validity and reliability of the model and its dimensions. A four-dimensional model emerged: shared content stress, latent stress, emotional stress, and self-evaluating stress. In the second stage, the four-item model of the 21-item SAS-SMU was validated using data collected from 510 university students. Cronbach's alpha coefficient for a dimension between 0.80 and 0.92. In the second step, the 21-item SAS-SMU four-factor model was analyzed using data collected from 510 university students. Cronbach's Alpha coefficients for dimensions range from 0.80 to 0.92.The analysis on the topic “Knowing when you're wrong: building fast and reliable approximate query processing systems[7]”,In this project, they discovered that modern data analytics applications typically process large amounts of data across clusters of tens, hundreds, or thousands of machines to support near real-time decision making. Limitations in data volume and disk and memory bandwidth often make it impossible to deliver answers at interactive speeds. However, it is generally observed that many applications can tolerate some inaccuracies.

This research begins with observations of online course services widely available in the media and the internet[8]. Next, the researchers went directly to see the existing course materials and spread them on different platforms, and then continued with discussions in the form of interviews with experts. Through interviews and Sihotang and Utama, Journal of System and Management Sciences[9] observations, a very close relationship was obtained between the organizing company and the course service provider, course user, or so-called learner (student). So that the research study is continued with a literature study. Several studies have been conducted by previous researchers related to online learning.

This research is closely related to the effectiveness of education users (students), the results of service providers (companies) and the management of education. Research that refers to student performance with the help of machine learning. Research in detecting student performance. This research begins by completing the data in the form of a dataset of test results with a parameter-based algorithm. The results of this study showed that the student datasets produced different qualities. This causes shortcomings in this study, namely the scope of the analysis to determine absolute assumptions that still have to be eliminated again. Model is the determination of parameters, to later investigate the value of the decision value before entering the process of determining the basis in computational calculations using fuzzy logic. Selecting the online course material itself has parameters that are used as the basis for the formation of the model; for that, an observation process is used such as interviews, surveys.

Research Summary:

| Sr. No. | Authors | Year | Topic | Description |
|---|---|---|---|---|
| **1** | Minsu Park, KAIST, David McDonald, University of Washington, Meeyoung Cha, KAIST,[4] | 2013 | “Perception Differences between the Depressed and Non-Depressed Users in Twitter” | In this they studied active Twitter users with depression to compare their attitudes and behaviors towards online social media with non-depressed users. |
| **2** | Walsh, C. G.,Ribeiro, J. D., & Franklin, J. C.[5] | 2017 | “Predicting Risk of Suicide Attempts Over Time Through Machine Learning” | In this they used machine learning to improve the accuracy and scale of suicide attempt risk detection in a large medical database. |
| **3** | Alkis, Y., Kadirhan, Z., & Sat, M.[6] | 2017 | “Development and Validation of Social Anxiety Scale for Social Media Users” | In this they created and validated a Social Anxiety Scale for Social Media Users (SAS-SMU) with four dimensions: shared content, privacy, interaction, and self-evaluation. |
| **4** | Agarwal, S., Milner, H., Kleiner, A., Talwalkar, A., Jordan, M., Madden, S., ... & Stoica, I.[7] | 2014 | “Knowing when you're wrong: building fast and reliable approximate query processing systems” | Modern data analytics applications process large amounts of data on clusters of machines, but limitations can cause delays. Some inaccuracy is often acceptable. |

## IV. PUBLIC ANXIETY ESTIMATION FRAMEWORK

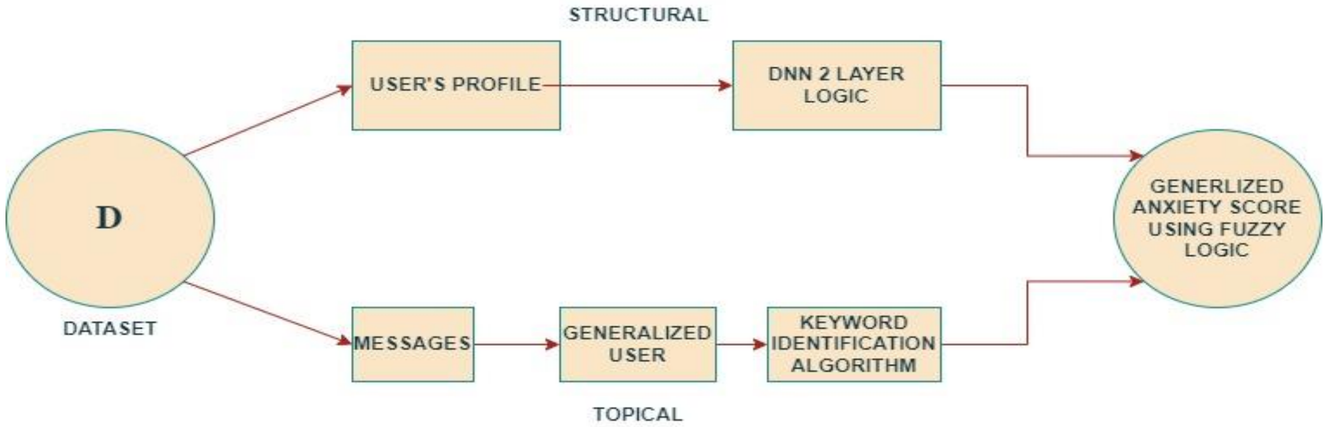


Fig.1. Framework to Estimate Public Anxiety

From the dataset, we plan to predict the design and stress in the region. There are user data for which we use some features such as age, location, gender for the stress process. Then we create a deep neural network for our model. A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers of input and output[10].The network we are going to build will use the sigmoid function. We will use this in test layer_2.

The only two possible outputs in the data are 0 and 1, and the sigmoid function limits the output to a range of 0 and 1.

Probability functions give you the probability of an event occurring. Of the data, only two possible values are 0 and 1, and the Bernoulli distribution is a distribution with two possible values.

The sigmoid function is a good choice if your problem fits the Bernoulli distribution, so we use it in the last step of the neural network. Since this function limits the output range from 0 to 1, we will use it to predict the result. If the output is greater than 0.5, we call the prediction 1.0 if it is less than 0.5, then we say the guess is 0.

For annoying content of the user participating in the discussion of a particular topic, by examining all the data and finding how many times a particular topic (for the user) occurs in all documents, counting the repetition of the topic and using it. It's a piece for setting the 0 and 1 logic. For the stress domain of the user participating in the discussion of a particular topic, we measure by examining all the data and finding how many times it occurs per unique (to the user) topic in all the data, we calculate the occurrence of the topic and use fuzzy logic to set the value between 0 and 1.

### i. ALGORITHM

Basically after calculating anxieties of particular topic given to communities , we applied DNN 2-Layer Logic[10]

1. Firstly we calculate Lower Range of U $\alpha = \frac{-1}{\sqrt{U}}$
2. We calculate Higher Range of U $\beta = \frac{1}{\sqrt{U}}$
3. We Choose a random Number N between 0 to 999
   $N = Rand(1000)$
4. We calculate the Scaled Value using Lower Range ,Higher Range and N
   $Scaled = \alpha + N * (\beta - \alpha)$
5. This is the Sigmoid Function that is Used in Calculating Layer-2 $\sigma(x) = \frac{1}{(1+e^{-x})}$
6. These are the Formulas to calculate Layer-1 & Layer-2
   $Layer\ 1 = u.W + bias$
   $Layer\ 2 = \sigma(Layer\ 1)$

Where,

$$\alpha = Lower\ Range$$

$$\beta = Higher\ Range$$

$$N = Random\ Numbers$$

$$U = Length\ of\ Input$$

$$u = List\ Anxiety\ Count$$

$$\sigma(x) = Sigmoid\ Function$$

$$W = Scaled.mean()$$

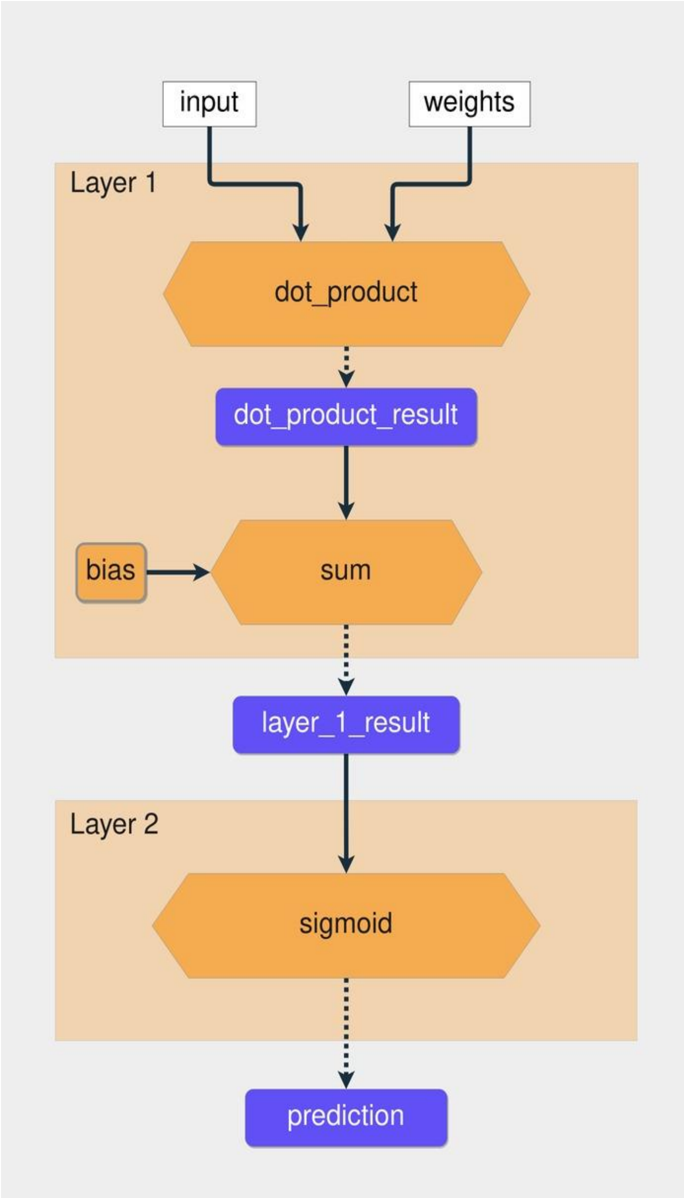


Fig.2. DNN 2 Layer Logic

ii. DataSet

We have picked the twitter dataset as our base to calculate the public anxiety of the whole community as well as the particular user anxiety. In the dataset there are three fields tweet, email and phone. When we provide our dataset to our DNN model it will give the structural as well as topical anxiety of the community and by this we can predict which group or community has greater anxiety as compared to other groups in terms of their messages and comments.

The link of the data-set is given in the references[11].

Methodology

For Topical Anxiety:-

Firstly we have to read the .csv file using pandas library of python. Then we take some empty lists and fill the list with values of the parameters of the .csv file. For eg.- Name, Comment, Id. Now in topical anxiety we have to check whether a particular keyword is commented by the user is how many times in discussion with other users. Then we calculate the percentage of entered keyword from the entire data-set. By this we can show the overall anxiety of a particular keyword entered by a user in a listed way so that the user can easily compare the anxiety with others anxiety.

For Structural Anxiety:-

In this we have to find the anxiety of a particular community for some predefined keywords. We have taken a set of predefined keywords which we will give to communities after discussing on particular keyword we find the overall anxiety of that community. After that we apply DNN-2 layer logic in which we calculate the higher and lower range of anxieties. Which is discussed above in the Formulation section. Create a weight array. Then we have define two functions . The first one is the sigmoid function and the second one is make prediction function. The value calculated from the sigmoid function pass to the make prediction function and the final overall anxiety of the group will be displayed though pie chart.

## V. RESULT AND ANALYSIS

Accuracy is also used as a measure of binary classification testing to accurately identify or exclude events. That is, accuracy is the proportion of correct predictions (true positives and negatives) in all patients studied [12]. Therefore, the formula for calculating accuracy is:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy measures the proportion of samples classified or the proportion of samples classified as positive [13].

Therefore, the formula for calculating the accuracy:

$$Precision = \frac{TP}{TP + FP}$$

Return is the number of relevant information provided by the research divided by the total number of relevant information, while accuracy is the number of relevant information provided by the research given divided by all the information provided by the research [14].

$$Recall = \frac{TP}{TP + FN}$$

The F score is a compromise between the accuracy of the system and the return values. can be calculated from the formula below:

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP = True positive; FP = False positive; TN = True negative; FN = False negative

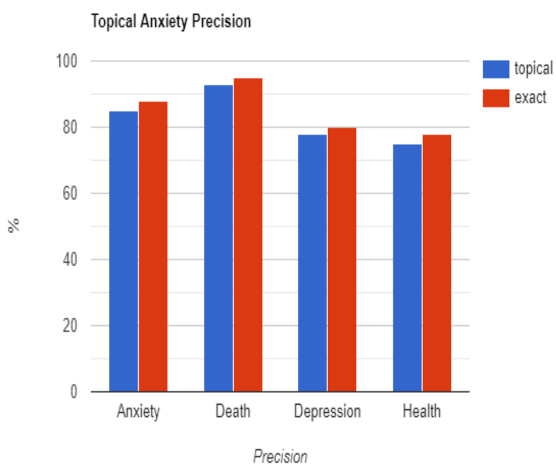


Fig.3. Precision of Topical Anxiety

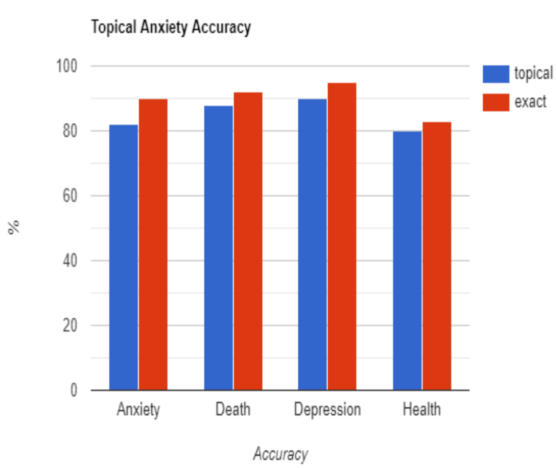


Fig.4. Accuracy of Topical Anxiety

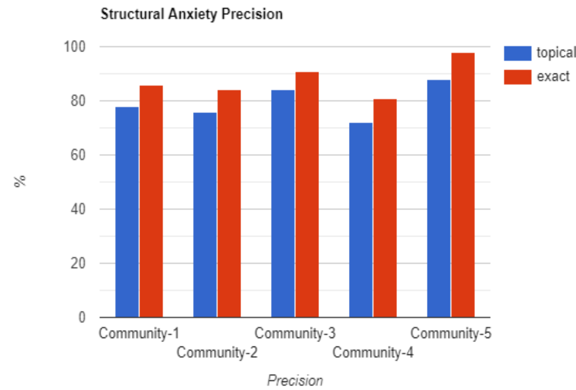


Fig.4.3. Precision of Structural Anxiety

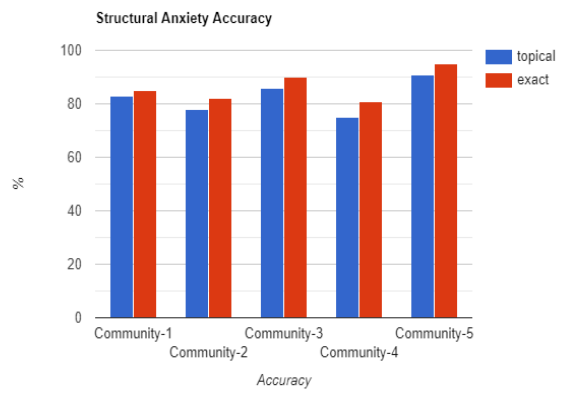


Fig.5. Accuracy of Structural Anxiety

We have tested our model in many test cases and also in topical and structural approach. From our analysis and testing we calculated the accuracy and precision of our model that is 88% accuracy and 92% precision and Recall is 84% and F1 score is 87%.

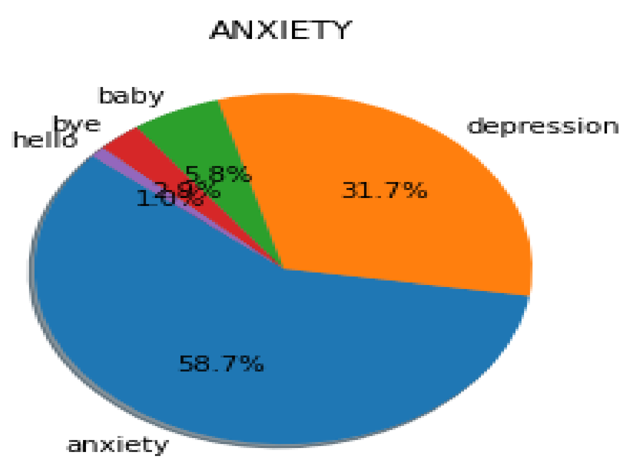


Fig.6. Pie chart Of Outcome

## VI. COMPARISION

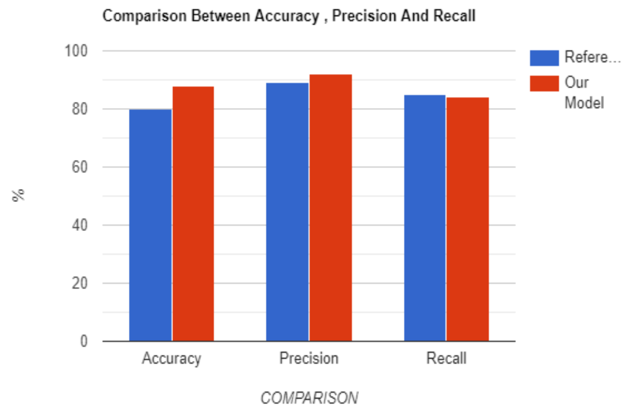


Fig.7. Comparison Bar Graph

Where Reference represents Evaluating Public Anxiety for Topic-based Communities in Social Networks[1] Accuracy and Precision.

By comparing their model with our model in many test cases we came to know that the accuracy of reference is come out to be 80% , their precision is come out to be 89% and their recall is come out to be 85% and our model's accuracy is came out to be 88% , our models precision is came out to be 92% and recall is 84%.

So from this comparison we came to know that our models precision and accuracy is more correct than other model.

## VII. CONCLUSION

In this research paper, we study and determine the problem of estimating topic - based public anxiety in social media communities using fuzzy tree. We design a valuable framework to estimate a topic – based public anxiety levels in social network communities using both Structural and Topical components. For the Structural component anxiety score, we calculate iteratively to evaluate community members anxiety scores. For the Topical components anxiety score, we prefer a problematic model to measure the comments and message anxiety score of the social communities. We design a message – comment tree structure (MC Tree) to estimate the public anxiety score in social media community to facilitate computation. Similarly, we estimate and evaluate the public anxiety of large communities by break- down into small segments. This model of estimating the anxiety score in social media exhibits high precision and accuracy in an actual study on real-world datasets.

## VIII. ACKNOWLEDGMENT

Thanks to GLA University for providing us with our mentor, Dr Pooja Pathak (Associate Professor at GLA University Mathura), by publishing a thesis statement in such a simple and legal form. He has helped us on this project since day one. He gave us a map explaining the main process of how the project would progress. He met regularly to review the progress of the project and provided us with project-related resources. Finally, we would like to thank our parents for helping us allocate this time for education, and my colleagues who helped me find resources during my studies.

## IX. REFERENCES

[1] Craske, Michelle G., Scott L. Rauch, Robert Ursano, Jason Prenoveau, Daniel S. Pine, and Richard E. Zinbarg. "What is an anxiety disorder?." Focus 9, no. 3 (2011): 369-388.

[2] Boulianne, Shelley. "Revolution in the making? Social media effects across the globe." Information, communication & society 22, no. 1 (2019): 39-54.

[3] Ta, N., Li, K., Yang, Y., Jiao, F., Tang, Z. and Li, G., 2020. Evaluating public anxiety for topic-based communities in social networks. IEEE Transactions on Knowledge and Data Engineering.

[4] Agarwal, S., Milner, H., Kleiner, A., Talwalkar, A., Jordan, M., Madden, S., ... & Stoica, I. (2014, June). Knowing when you're wrong: building fast and reliable approximate query processing systems. In Proceedings of the 2014 ACM SIGMOD international conference on Management of data (pp. 481-492).

[5] Alkis, Y., Kadirhan, Z., & Sat, M. (2017). Development and validation of social anxiety scale for social media users. Computers in Human Behavior, 72, 296-303.

[6] Walsh, C. G.,Ribeiro, J. D., & Franklin, J. C. (2017). Predicting risk of suicide attempts over time through machine learning. Clinical Psychological Science, 5(3), 457-469.

[7] Park, M., McDonald, D., & Cha, M. (2013). Perception differences between the depressed and non-depressed users in twitter. In Proceedings of the International AAAI Conference on Web and Social Media (Vol. 7, No. 1, pp. 476-485).

[8] Memmedova, Konul. "Fuzzy logic modelling of the impact of using technology on anxiety and aggression levels of students." *Procedia computer science* 120 (2017): 495-501.

[9] Almadi, Abdulla IM, Rabia Emhamed Al Mamlook, Yahya Almarhabi, Irfan Ullah, Arshad Jamal, and Nishantha Bandara. "A fuzzy-logic approach based on driver decision-making behavior modeling and simulation." *Sustainability* 14, no. 14 (2022): 8874.

[10] Zhang, Junbo, Yu Zheng, Dekang Qi, Ruiyuan Li, and Xiuwen Yi. "DNN-based prediction model for spatio-temporal data." In *Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems*, pp. 1-4. 2016.

[11] https://github.com/sanidhya12345/Evaluating-Public-Anxiety/blob/main/twitter_English.csv

[12] Metz, CE (October 1978). "Basic principles of ROC analysis" (PDF). Semin Nucl Med. 8 (4): 283–98. doi:10.1016/s0001-2998(78)80014-2. PMID 112681

[13] Bzdok, Danilo, and Andreas Meyer-Lindenberg. "Machine learning for precision psychiatry: opportunities and challenges." Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 3, no. 3 (2018): 223-230.

[14] Buckland, Michael, and Fredric Gey. "The relationship between recall and precision." *Journal of the American society for information science* 45, no. 1 (1994): 12-19.

**Project Mentor**

**Dr. Pooja Pathak** is an Associate Professor at GLA University Mathura.

**Our Team**

**Ravi Varshney** is a student at GLA University Mathura. He is pursuing B.Tech. in Computer Science & Engineering.

**Sanidhya Varshney** is a student at GLA University Mathura. He is pursuing B.Tech. in Computer Science & Engineering.

**Prabuddha Kumar** is a student at GLA University Mathura. He is pursuing B.Tech. in Computer Science & Engineering.

**Lalit Saini** is a student at GLA University Mathura. He is pursuing B.Tech. in Computer Science & Engineering.