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

A Project Report submitted in partial fulfilment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science and Engineering

by

RAVI VARSHNEY(191500642)
SANIDHYA VARSHNEY(191500714)
LALIT SAINI(191500428)
PRABUDDHA KUMAR(191500551)

Group No.:2

Under the Guidance of

Dr. Pooja Pathak ,Associate Professor

Department of Computer Engineering & Applications

Institute of Engineering & Technology

GLA University

Mathura- 281406, INDIA

May, 2023

Department of Computer Engineering and Applications
GLA University, 17 km Stone, NH#2, Mathura-Delhi Road, P.O. Chaumuhan, Mathura-281406 (U.P.)

Declaration

I hereby declare that the work which is being presented in the B.Tech. Project **"Estimating Topic Based Public Anxiety In Social Media Using Fuzzy Tree"**, in partial fulfillment of the requirements for the award of the ***Bachelor of Technology*** in Computer Science and Engineering and submitted to the Department of Computer Engineering and Applications of GLA University, Mathura, is an authentic record of my own work carried under the supervision of **Dr. Pooja Pathak, Associate Professor, Dept. of CEA, GLA University.**

The contents of this project report, in full or in parts, have not been submitted to any other Institute or University for the award of any degree.

Sign _______________________

Name of Student: Ravi Varshney
University Roll No.: 191500642

Sign _______________________

Name of Student: Sanidhya Varshney
University Roll No.: 191500714

Sign _______________________

Name of Student: Prabuddha Kumar
University Roll No.: 191500551

Sign _______________________

Name of Student: Lalit Saini
University Roll No.: 191500428

Certificate

This is to certify that the above statements made by the candidate are correct to the best of my/our knowledge and belief.

____

Supervisor
(Dr. Pooja Pathak)
Associate Professor
Dept. of Computer Engg, & App.

____

Project Co-ordinator
(Dr. Mayank Srivastava)
Associate Professor
Dept. of Computer Engg, & App.

____

Program Co-ordinator
(Dr, Rajesh Kumar Tripathi)
Associate Professor
Dept. of Computer Engg, & App.

Date:

ACKNOWLEDGEMENT

Presenting the ascribed project paper report in this very simple and official form, we would like to place my deep gratitude to GLA University for providing us the instructor Dr. Pooja Pathak, our supervisor.

She has been helping us since Day 1 in this project. She provided us with the roadmap, the basic guidelines explaining on how to work on the project. She has been conducting regular meeting to check the progress of the project and providing us with the resources related to the project. Without his help, we wouldn't have been able to complete this project.

And at last but not the least we would like to thank our dear parents for helping us to grab this opportunity to get trained and also my colleagues who helped me find resources during the training.

Sign: Sanidhya Varshney
Name: Sanidhya Varshney
University RollNo.:191500714

Sign: Lalit Saini
Name: Lalit Saini
University RollNo.:191500428

Sign: Prabuddha Kumar
Name: Prabuddha Kumar
University RollNo.:191500551

Sign:Ravi Varshney
Name: Ravi Varshney
University RollNo.:191500642

ABSTRACT

In this present era everyone is using social media in which there are some messages, posts or comments that creates anxiety or depression to users. Although individual anxiety evaluation has been well studied, there is still not much work on evaluating publicc anxiety of groups, especially in the form of communities on social networks, which can be leveraged to detect mental health of a society. However, we cannot simply average individual anxiety scores to evaluate a community's public anxiety, because following factors should be considered:-

(1) impacts from interpersonal relations on each individual group member's anxiety levels (the Structural component).
(2) topic-based based discussions which reflect a community's anxiety status (the Topical component).In this Project first of all we start the study of evaluating public anxiety of Topic-based Social Network Communities (TSNC).We make an evaluation framework to project the anxiety level of a TSNC into a score in the [0, 1] range.

We create a cascading model to dynamically compute the individual anxiety scores using the Structural influence. We design a fuzzy model to measure anxiety score of social network messages using a generalised user, and compose a tree structure (MC-Tree) to effectively compute the anxiety an score of a TSNC from the Topical aspect.

CONTENTS

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Motivation and Overview

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[2] at work. The overwhelming need to share information on social media 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. Intense fear of interacting 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.

1.2 Objective

Although individual anxiety assessments have been well researched, there is still not much work assessing public anxiety in groups that can be used to identify mental health in society, especially in the form of social networking communities. However, it is not possible to simply average individual anxiety scores to assess public anxiety in a community because the following factors need to be considered. (2) topic-based discussions (current component) that reflect the state of anxiety in the community; This article initiates study to assess public anxiety in topic-based social networking communities (TSNC). We propose a scoring framework for projecting TSNC anxiety levels into scores in the range [0, 1]. Develop a cascade model to dynamically calculate an individual's anxiety score using structural influences. We designed a probabilistic model to measure the anxiety scores of social network messages using generalized users and built a tree structure called fuzzr tree to effectively calculate the anxiety scores of TSNC from the topic aspects. increase. To avoid expensive real-time computations, for large communities we use small samples to compute public fears within a specific confidence interval. The validity of our model is validated by precision and recall in empirical studies on real Facebook,Twitter and Instagram datasets.

1.3 Issues and Challenges

We cannot simply add up the individual anxiety levels of all community members or the average of people tested by the generic SAS. Some are prone to worry, some are not, but all members conform and follow community norms and make their feelings known by discussing community issues. For example, her guardians on the interim, who have low personal anxiety levels in communities where they discuss parenting, can be very anxious. Public unrest in the community may even escalate, regardless of the overall SAS score of all members. Estimating public fear using surveys is a laborious task for calculating fear scores. A key challenge in our experiments is collecting sufficient training samples to train our cascading model. Estimate the exact value for each topic in the fuzzy tree.

1.4 Contribution

To our knowledge, it is the first study to formally define and quantitatively assess public fear of problem-based social networking communities. We propose a scoring framework and corresponding algorithm for estimating community fear using structural and thematic components.

Using generalized users, we propose a probabilistic model to measure topic public anxiety scores of social network messages in topic-based social network communities. We design a fuzzy tree structure that organizes the messages to speed up the calculation of the current fear score.

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.

1.5 Organization of the Project Report

In this project we are starting from collective enough data samples from social media networks to learn our model. After that we will calculate first anxiety on the basis of topic based in that calculation. We calculate particular topic is present in how much percent of the total dataset so that we came into conclusion that this keyword(anxiety containing word) is having high anxiety score by which ant-depression committee or anti-suicidal committee for the prevention of suicides on the basis of comments and post messages. After calculation of Topical Anxiety we calculate the Structural anxiety in which we use deep neural networking to train our model to calculate the anxiety on which how user relationship with others. At the end we have two types of anxiety scores first one is topical anxiety and second one is structural anxiety after applying fuzzy logic to both of these scores a level is decided on which anxiety evaluation will be done.

Chapter 2
Literature Review

2.1 Research Paper 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 face-to-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 several scales were developed with the aim of measuring social anxiety among students in various contexts, but none of the studies addressed the measurement of social anxiety on social media platforms. .This study describes the development and validation process of the Multidimensional Social Anxiety Scale for Social Media Users (SAS-SMU), which can be used to assess social anxiety in college students due to media platforms. social. The study was conducted in two phases. In the first phase, data collected from 174 students was used to provide evidence for the validity and reliability of the structure and its underlying dimensions. A four-dimensional structure emerged: shared content anxiety, privacy anxiety, interaction anxiety, and self-evaluation anxiety. In the second phase, data collected from 510 university students was used to confirm the four-factor structure of the 21-item SAS-SMU. Cronbach's Alpha coefficients for dimensions range from 0.80 to 0.92. In the second phase, data collected from 510 university students was used to confirm the four-factor structure of the 21-item SAS-SMU. 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 performance of learning users (students), the benefits of service providers (companies), and learning management. 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,

Although various scales have been developed with the aim of measuring students' social anxiety in a variety of settings, none of the studies has addressed the measurement of social anxiety in social media platforms. This study describes the process of developing and validating a multidimensional Social Anxiety Scale for Social Media Users (SAS-SMU) that can be used to assess college students' social anxiety arising from social media platforms. The study was conducted in two phases. In the first phase, data collected from 174 students were used to provide evidence for validity and reliability of the structure and its underlying dimensions. A four-dimensional structure emerged: shared content anxiety, privacy concern anxiety, interaction anxiety, and self-evaluation anxiety. In the second phase, data collected from 510 college students were used to confirm four-factor structure of the 21-item SAS-SMU. The Cronbach's Alpha coefficients for the dimensions ranged from 0.80 to 0.92, demonstrating a satisfactory level of reliability. Further validation studies were also conducted and their findings provided. This validated scale will be a useful tool for both researchers and instructors to assess college students' social anxiety as social media users.

One's state of mind will influence her perception of the world and people within it. In this paper, we explore attitudes and behaviors toward online social media based on whether one is depressed or not. We conducted semi-structured face-to-face interviews with 14 active Twitter users, half of whom were depressed and the other half non-depressed. Our results highlight key differences between the two groups in terms of perception towards online social media and behaviors within such systems. Non-depressed individuals perceived Twitter as an information consuming and sharing tool, while depressed individuals perceived it as a tool for social awareness and emotional interaction. We discuss several design implications for future social networks that could better accommodate users with depression and provide insights towards helping depressed users meet their needs through online social media.

Using technology such as mass media, mobile technology and the internet has positive and negative effects. In this research, the effect of using technology on aggression and anxiety levels of students was investigated using fuzzy logic approach. In order to evaluate the effect of technology use on students' aggression and anxiety levels, 100 students from the Faculty of Education of Near East University (North Cyprus) invited to participate in this study. Psychological variables were gathered using the "Technology use habits", "Effect of Social media", "Role of technology in daily life", "Educational use", "Communication use" and "Continuous - State Anxiety Inventory" Questionnaires. Application of fuzzy logic in this study allows researchers to handle the imprecision and vagueness inherence of input data and develop the more reliable model for computing input-output relations.

The present study proposes a decision-making model based on different models of driver behavior, aiming to ensure integration between road safety and crash reduction based on an examination of speed limitations under weather conditions. The present study investigated differences in road safety attitude, driver behavior, and weather conditions I-69 in Flint, Genesee County, Michigan, using the fuzzy logic approach. A questionnaire-based survey was conducted among a sample of Singaporean (n = 100) professional drivers. Safety level was assessed in relation to speed limits to determine whether the proposed speed limit contributed to a risky or safe situation. The experimental results show that the speed limits investigated on different roads/in different weather were based on the participants' responses. The participants could increase or keep their current speed limit or reduce their speed limit a little or significantly. The study results were used to determine the speed limits needed on different roads/in different weather to reduce the number of crashes and to implement safe driving conditions based on the weather. Changing the speed limit from 80 mph to 70 mph reduced the number of crashes occurring under wet road conditions. According to the results of the fuzzy logic study algorithm, a driver's emotions can predict outputs. For this study, the fuzzy logic algorithm evaluated drivers' emotions according to the relation between the weather/road condition and the speed limit. The fuzzy logic would contribute to assessing a powerful feature of human control. The fuzzy logic algorithm can explain smooth relationships between the input and output. The input–output relationship estimated by fuzzy logic was used to understand differences in drivers' feelings in varying road/weather conditions at different speed limits.

The nature of mental illness remains a conundrum. Traditional disease categories are increasingly suspected to mis-represent the causes underlying mental disturbance. Yet, psychiatrists and investigators now have an unprecedented opportunity to benefit from complex patterns in brain, behavior, and genes using methods from machine learning (e.g., support vector machines, modern neural-network algorithms, cross-validation procedures). Combining these analysis techniques with a wealth of data from consortia and repositories has the potential to advance a biologically grounded re-definition of major psychiatric disorders. Within the next 10-20 years, incoming patients could be stratified into distinct biological subgroups that cut across classical diagnostic boundaries.

In a new era of evidence-based psychiatry tailored to single patients, objectively measurable endophenotypes could allow for individualized prediction of early diagnosis, treatment selection, and dosage adjustment to reduce the burden of disease. This primer aims to introduce clinicians and researchers to the opportunities and challenges in bringing machine intelligence into psychiatric practice.

2.2 Research Summary

S.No.	Author	Year	Topic
1	Minsu Park, KAIST David McDonald University of Washington, Meeyoung Cha	2013	Perception Differences between the Depressed and Non-Depressed Users in Twitter
2	Walsh, C. G.,Ribeiro, J. D., & Franklin, J. C.	2017	Predicting Risk of Suicide Attempts Over Time Through Machine Learning
3	Alkis, Y., Kadirhan, Z., & Sat, M	2017	Development and Validation of Social Anxiety Scale for Social Media Users.
4	Agarwal, S., Milner, H., Kleiner, A., Talwalkar, A., Jordan, M., Madden, S., ... & Stoica, I	2014	Knowing when you're wrong: building fast and reliable approximate query processing systems

Table.2.1. Research Summary

2.3 Fuzzy Logic

The theory of fuzzy logic is based on the notion of relative graded membership, as inspired by the processes of human perception and cognition. Lotfi A. Zadeh published his first famous research paper on fuzzy sets in 1965[10]. Fuzzy logic can deal with information arising from computational perception and cognition, that is, uncertain, imprecise, vague, partially true, or without sharp boundaries. Fuzzy logic allows for the inclusion of vague human assessments in computing problems. Also, it provides an effective means for conflict resolution of multiple criteria and better assessment of options.

New computing methods based on fuzzy logic can be used in the development of intelligent systems for decision making, identification, pattern recognition, optimization, and control. Fuzzy logic is extremely useful for many people involved in research and development including engineers (electrical, mechanical, civil, chemical, aerospace, agricultural, biomedical, computer, environmental, geological, industrial, and mechatronics), mathematicians, computer software developers and researchers, natural scientists (biology, chemistry, earth science, and physics), medical researchers, social scientists (economics, management, political science, and psychology), public policy analysts, business analysts, and jurists. Indeed, the applications of fuzzy logic, once thought to be an obscure mathematical curiosity, can be found in many engineering and scientific works. Fuzzy logic has been used in numerous applications such as facial pattern recognition, air conditioners, washing machines, vacuum cleaners, antiskid braking systems, transmission systems, control of subway systems and unmanned helicopters, knowledge-based systems for multiobjective optimization of power systems, weather forecasting systems, models for new product pricing or project risk assessment, medical diagnosis and treatment plans, and stock trading. Fuzzy logic has been successfully used in numerous fields such as control systems engineering, image processing, power engineering, industrial automation, robotics, consumer electronics, and optimization. This branch of mathematics has instilled new life into scientific fields that have been dormant for a long time. Thousands of researchers are working with fuzzy logic and producing patents and research papers. According to Zadeh's report on the impact of fuzzy logic as of March 4, 2013, there are 26 research journals on theory or applications of fuzzy logic, there are 89,365 publications on theory or applications of fuzzy logic in the INSPEC database, there are 22,657 publications on theory or applications of fuzzy logic in the MathSciNet database, there are 16,898 patent applications and patents issued related to fuzzy logic in the USA, and there are 7149 patent applications and patents issued related to fuzzy logic in Japan. The number of research contributions is growing daily and is growing at an increasing rate. Zadeh started the Berkeley Initiative in Soft Computing (BISC), a famous research laboratory at University of California, Berkeley, to advance theory and applications of fuzzy logic and soft computing. The objective of this special issue is to explore the advances of fuzzy logic in a large number of real-life applications and commercial products in a variety of fields. Although fuzzy logic has applications in a number of different areas, it is not yet known to people unfamiliar with intelligent systems how it can be applied in different products that are currently available in the market. For many people, the engineering and scientific meaning of the word fuzzy is still fuzzy. It is important that these people understand where and how fuzzy logic can be used.

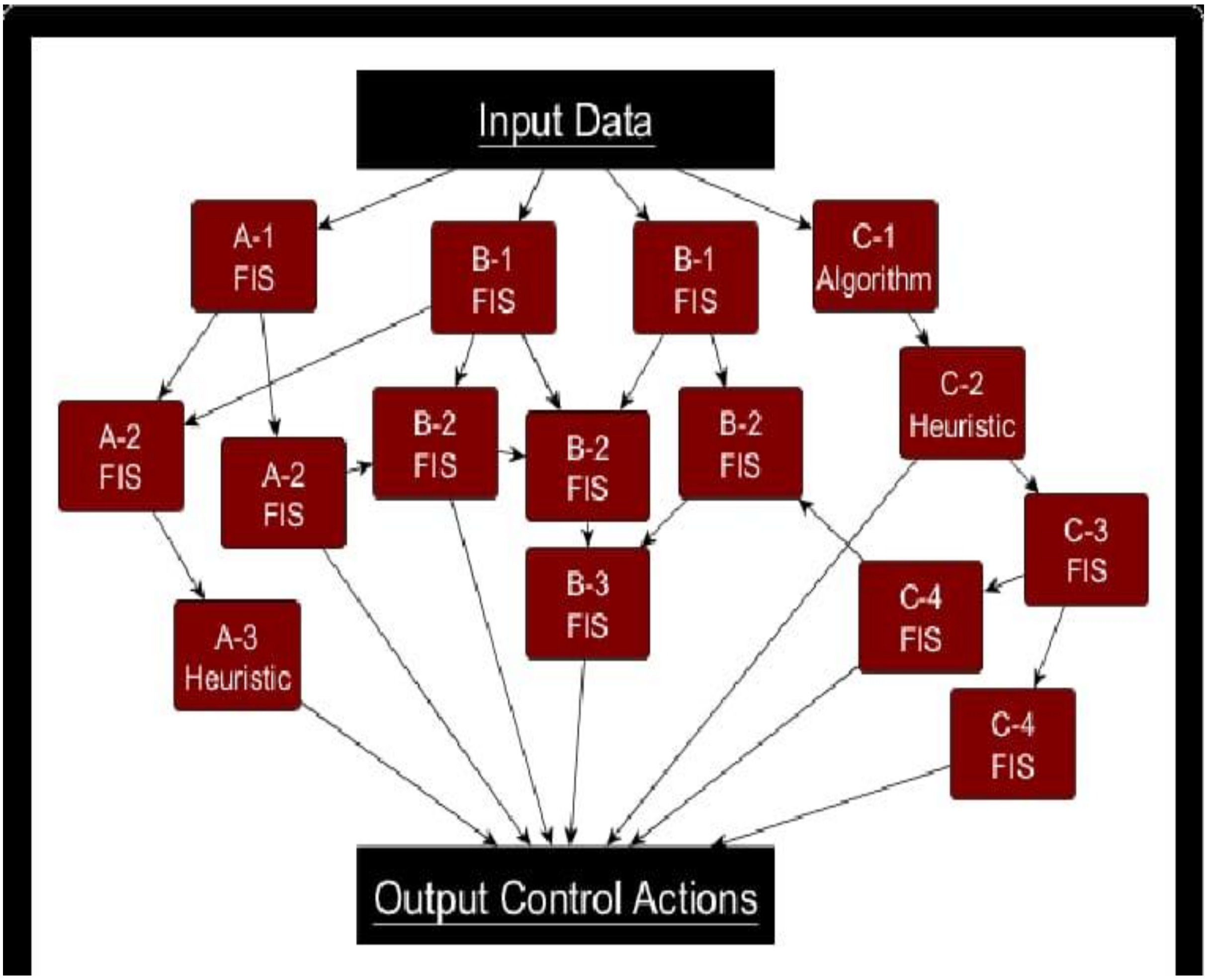

Fig.2.1. Fuzzy Tree Diagram

2.4 Fuzzy Logic Architecture

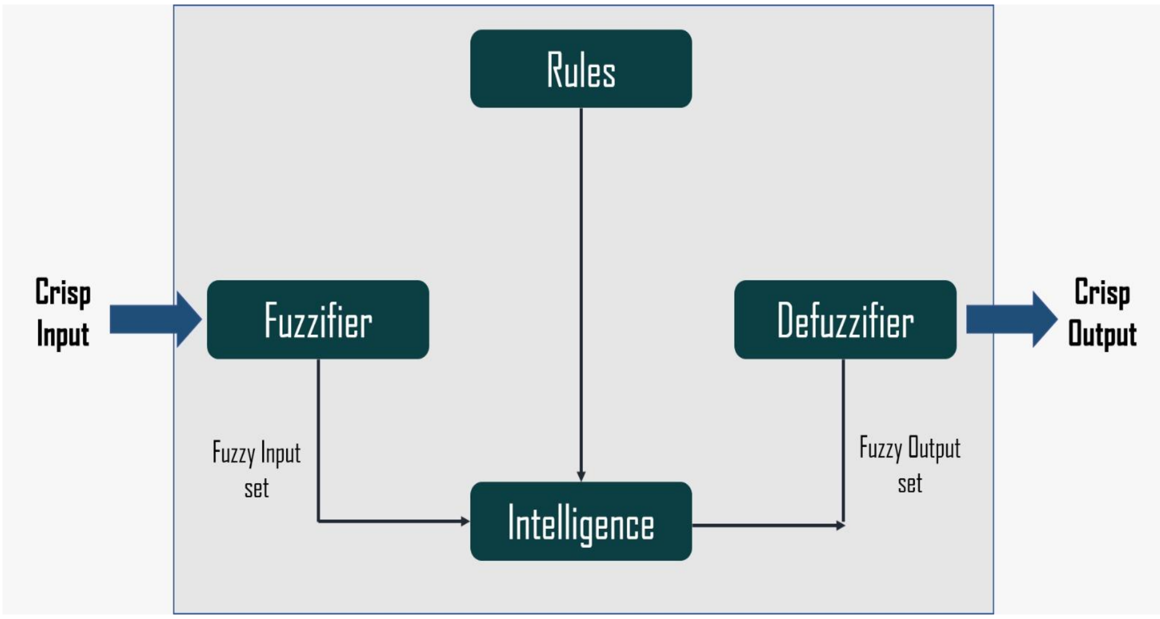

Fig.2.2. Fuzzy Logic Architecture

- **Rules** – It contains all the rules and the if-then conditions offered by the experts to control the decision-making system. The recent update in the fuzzy theory provides different effective methods for the design and tuning of **fuzzy controllers**. Usually, these developments reduce the number of fuzzy rules.
- **Fuzzification** – This step converts inputs or the crisp numbers into fuzzy sets. You can measure the crisp inputs by sensors and pass them into the **control system** for further processing. It splits the input signal into five steps such as-

LP	X is Large Positive
MP	X is Medium Positive
S	Small
MN	X is Medium Negative
LN	X is Large Negative

Table.2.2. Fuzzification Steps

- **Inference Engine** – It determines the degree of match between fuzzy input and the rules. According to the input field, it will decide the rules that are to be fired. Combining the fired rules, form the control actions.

- **Defuzzification** – The Defuzzification process converts the fuzzy sets into a crisp value. There are different types of techniques available, and you need to select the best-suited one with an expert system.
- Fuzzy Logic vs Probability

Fuzzy Logic	**Probability**
In fuzzy logic, we basically try to capture the essential concept of vagueness.	Probability is associated with events and not facts, and those events will either occur or not occur
Fuzzy Logic captures the meaning of partial truth	Probability theory captures partial knowledge
Fuzzy logic takes truth degrees as a mathematical basis	Probability is a mathematical model of ignorance

Table.2.3. Fuzzy Logic vs Probability

Chapter 3

Proposed Work

3.1 Work Break Down Structure

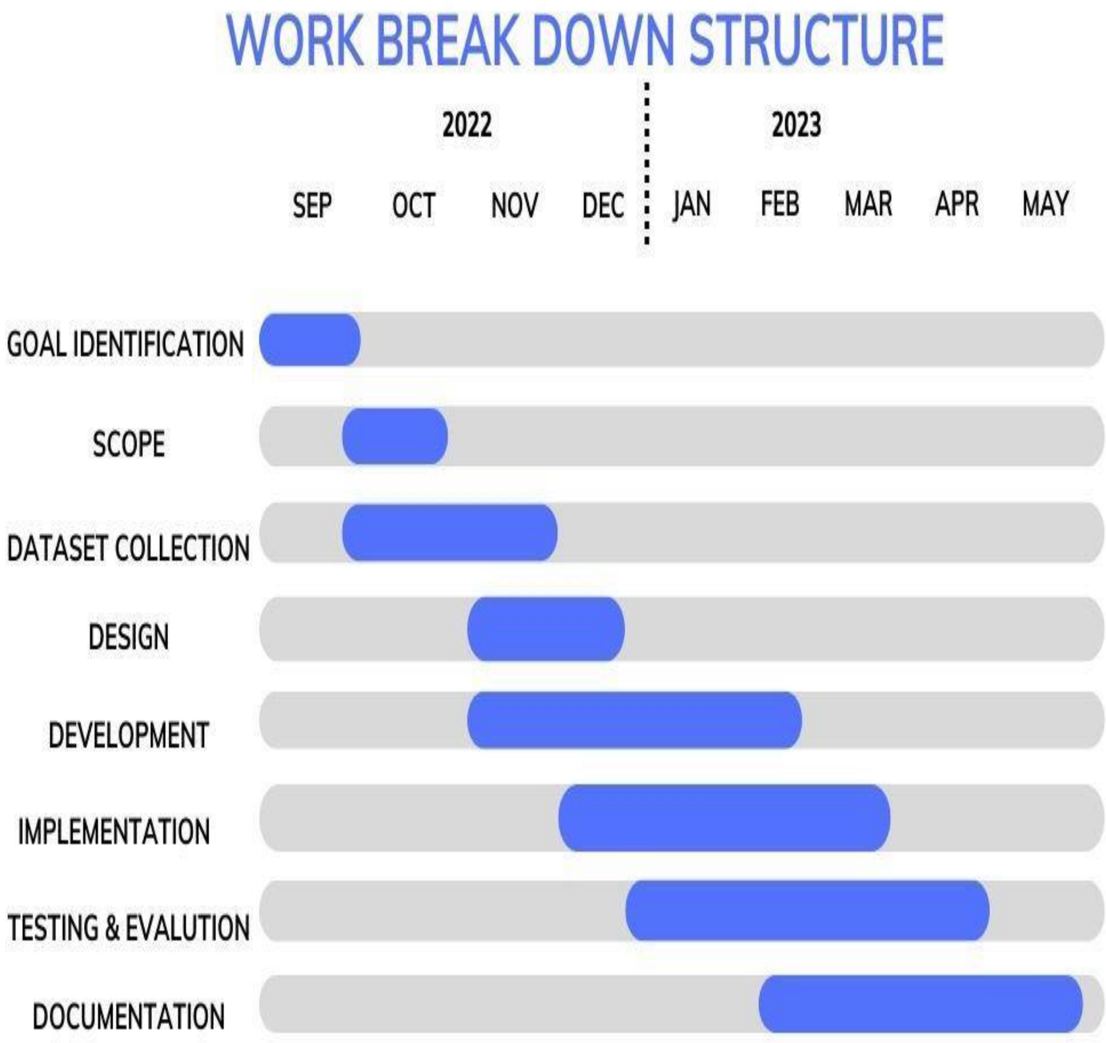

Fig.3.1. Work Break Down Structure

3.2 PUBLIC ANXIETY ESTIMATION FRAMEWORK

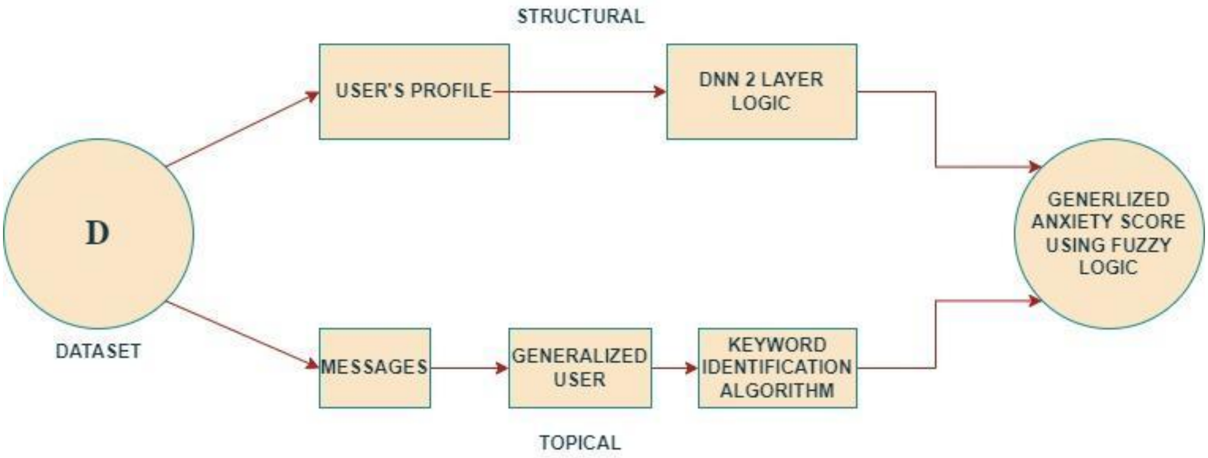

Fig.3.2. Framework to Estimate Public Anxiety

From dataset we start predicting the structural as well as topical anxiety.For structural anxiety, there are user's profiles in which we have taken some features such as age, location, gender etc. then after this we create deep neural network for our model.A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers[8]. A deep neural network (DNN) is an artificial neural network (ANN) with multiple layers between the input and output layers. The network we’re building will use the sigmoid activation function. We will use it in the last layer, layer_2. The only two possible outputs in the dataset are 0 and 1, and the sigmoid function limits the output to a range between 0 and 1.

Probability functions give you the probability of occurrence for possible outcomes of an event. The only two possible outputs of the dataset are 0 and 1, and the Bernoulli distribution is a distribution that has two possible outcomes as well. The sigmoid function is a good choice if your problem follows the Bernoulli distribution, so that’s why we’re using it in the last layer of your neural network. Since the function limits the output to a range of 0 to 1, we’ll use it to predict probabilities. If the output is greater than 0.5, then we’ll say the prediction is 1. If it’s below 0.5, then we’ll say the prediction is 0[9]. For topical anxiety in which a user participate in a discussion on the particular topic and we measure by traversing the whole dataset and find how many times a particular topic(given to user) occuring in whole dataset,calculate the probability of the topic and applying fuzzy logic to set the value between 0 and 1.

3.3 Formulation

Basically after calculating anxieties of particular topic given to communities , we applied DNN 2-Layer Logic[11]

i. Firstly we calculate Lower Range of U

$$\alpha = \frac{-1}{\sqrt{U}}$$

ii. We calculate Higher Range of U

$$\beta = \frac{1}{\sqrt{U}}$$

iii. We Choose a random Number N between 0 to 999

$$N = Rand(1000)$$

iv. We calculate the Scaled Value using Lower Range ,Higher Range and N

$$Scaled = \alpha + N * (\beta - \alpha)$$

v. This is the Sigmoid Function that is Used in Calculating Layer-2

$$\sigma(x) = \frac{1}{(1 + e^{-x})}$$

vi. These are the Formulas to calculate Layer-1 & Layer-2

$$Layer\ 1 = u.W + bias$$

$$Layer\ 2 = \sigma(Layer\ 1)$$

Where,

$$\alpha = Lower\ Range$$

$$\beta = Higher\ Range$$

$$N = Random\ Numbers$$

$$U = Length\ of\ Input$$

$$u = List\ Anxiety\ Count$$

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

$$W = Scaled.mean()$$

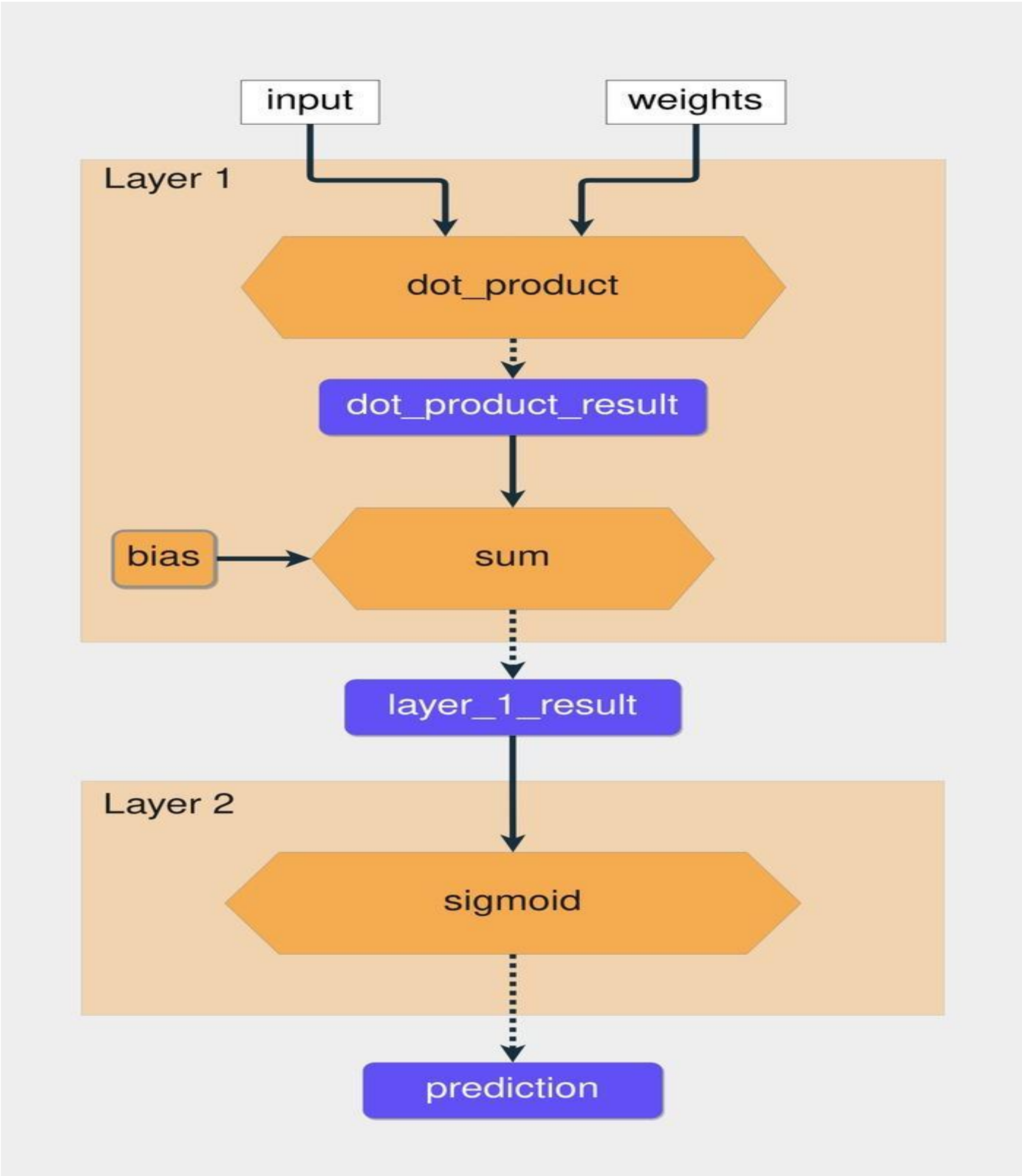

Fig.3.3. DNN 2 Layer Logic

3.4 Data Set

We have picked the twitter dataset as our base to calculate the public anxiety of whole community as well as the particular user anxiety. In 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 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.

```
                                                 tweet  label
0    I just don’t take my baby out because it gives...      1
1    Suicide figures are up 200% since lockdown. \n...      1
2    she has really been feeding us this week https...      0
3    I can confirm that yes, I do need to stay on m...      1
4    Today is pie day! Celebrate with your "I Ate P...      0
..                                                 ...    ...
995  Dementia, support for carers, electrical fault...      1
996  My view? \n\n'Africa needs more of @AbiyAhmedA...      0
997  This is the 3 year anniversary of adopting Ms....      0
998  “Our Young Addicts” story – How did we get HER...      0
999                         pain lol https://t.co/AKGL15Scy9      1

[1000 rows x 2 columns]
{'anxiety': 61, 'hello': 1, 'baby': 6, 'bye': 3, 'depression': 33}
['anxiety', 61, 0.5865384615384616, 'hello', 1, 0.009615384615384€
[{'topic': 'anxiety', 'count': 61, 'score': 0.5865384615384616}, {
        topic  count     score
0     anxiety     61  0.586538
1       hello      1  0.009615
2        baby      6  0.057692
3         bye      3  0.028846
4  depression     33  0.317308
        topic  count     score
0     anxiety     61  0.586538
4  depression     33  0.317308
2        baby      6  0.057692
3         bye      3  0.028846
1       hello      1  0.009615
```

Fig.3.4. Dataset

Chapter 4

Requirements

4.1 Hardware Requirements:

- Central Processing Unit (CPU) — Intel Core i3 8th Generation
- RAM-4 GB minimum.
- Graphics Processing Unit (GPU) — NVIDIA GeForce GTX 960 or higher.
- Operating System — Ubuntu or Microsoft Windows 10.

4.2 Software Requirements:

- Pycharm, Jupyter Notebook
- Python 3.7 (& libraries)

Chapter 5

Implementation and Result Analysis

4.1 Implementation

4.1.1 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 .csv file. For eg.- Name, Comment, Id. Now in topical anxiety we have to check a particular keyword 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 overall anxiety of particular keyword entered by user in a listed way so that the user can easily compare the anxiety with others anxiety.

4.1.2 For Structural Anxiety:-

In this we have to find the anxiety of 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 Formulation section. Create a weight array.

Then we have define two functions first one is sigmoid function and second one is make prediction function. The value calculated from sigmoid function pass to the make prediction function and final overall anxiety of group will displayed though pie chart.

4.2 Result And Analysis

The accuracy of a machine learning classification algorithm is one way to measure how often the algorithm classifies a data point correctly. Accuracy is the number of correctly predicted data points out of all the data points. More formally, it is defined as the number of true positives and true negatives divided by the number of true positives, true negatives, false positives, and false negatives. A true positive or true negative is a data point that the algorithm correctly classified as true or false, respectively. A false positive or false negative, on the other hand, is a data point that the algorithm incorrectly classified. For example, if the algorithm classified a false data point as true, it would be a false positive. Often, accuracy is used along with precision and recall, which are other metrics that use various ratios of true/false positives/negatives. Together, these metrics provide a detailed look at how the algorithm is classifying data points [12]. Thus, the formula to calculate the precision is given by:

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

Precision is one indicator of a machine learning model's performance – the quality of a positive prediction made by the model. Precision refers to the number of true positives divided by the total number of positive predictions (i.e., the number of true positives plus the number of false positives). In a customer attrition model, for example, precision measures the number of customers that the model correctly predicted would unsubscribe divided by the total number of customers the model predicted would unsubscribe. [13]. Thus, the formula to calculate the precision is given by:

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

Recall is the number of relevant documents retrieved by a search divided by the total number of existing relevant documents, while precision is the number of relevant documents retrieved by a search divided by the total number of documents retrieved by that search[14].
Recall places a high importance on reducing the number of false negatives, for example positive cases that are misclassified by the model as negatives. For that reason, it is important in mission-critical applications where a false negative could lead to loss of life or millions of dollars in damages. In such applications, it is essential to maximize recall.

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

The F-score, also called the F1-score, is a measure of a model's accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into 'positive' or 'negative'. The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall. The F-score is commonly used for evaluating information retrieval systems such as search engines, and also for many kinds of machine learning models, in particular in natural language processing.It is possible to adjust the F-score to give more importance to precision over recall, or vice-versa. Common adjusted F-scores are the F0.5-score and the F2-score, as well as the standard F1-score[15]. It can be calculated by the following formula:

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

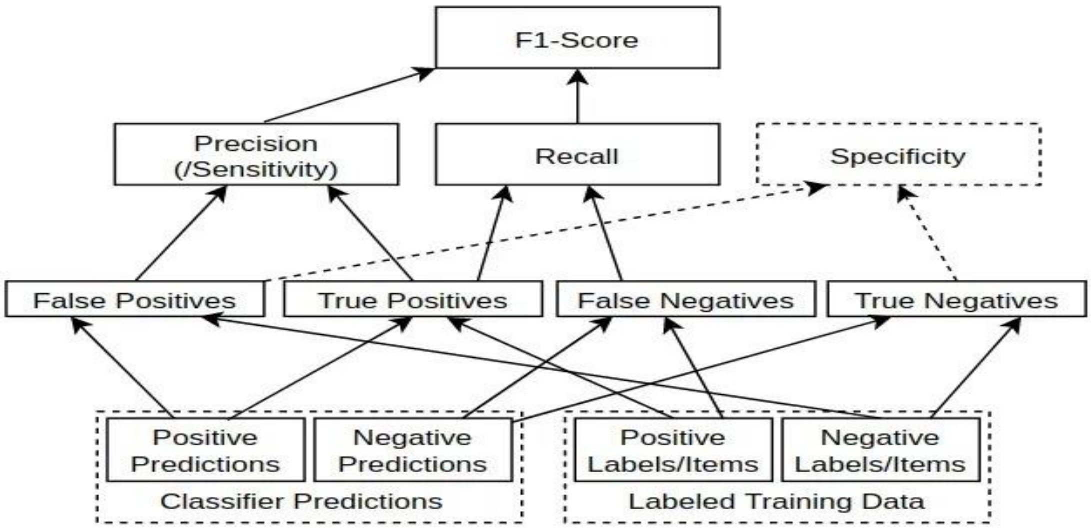

Fig.4.1. F1-Score

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

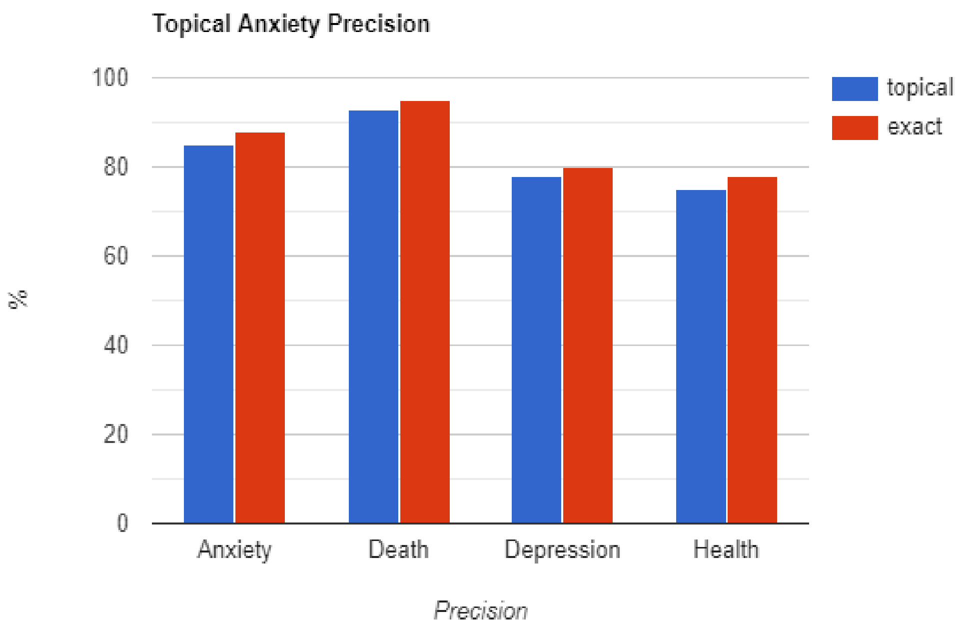

Fig.4.2. Precision of Topical Anxiety

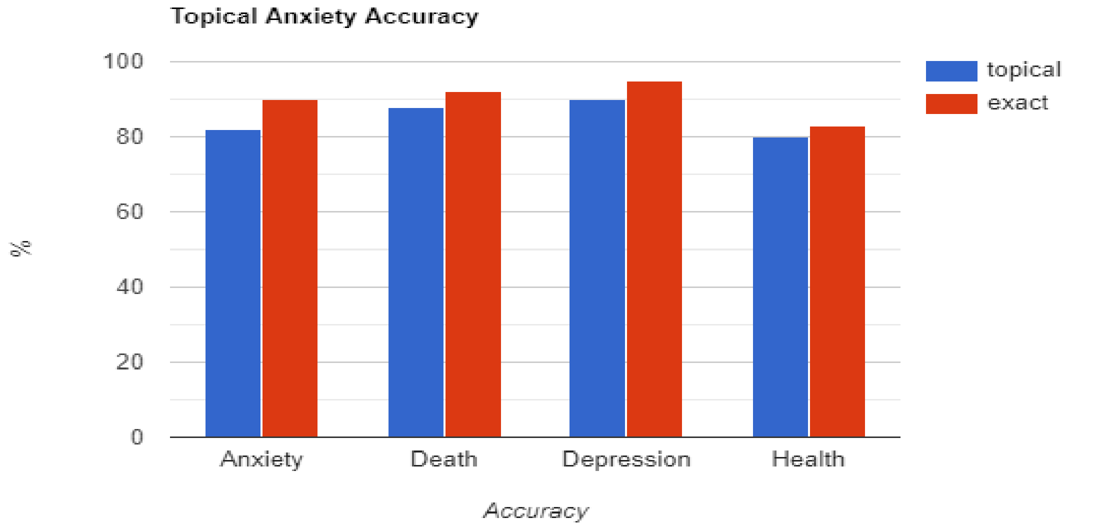

Fig.4.3. Accuracy of Topical Anxiety

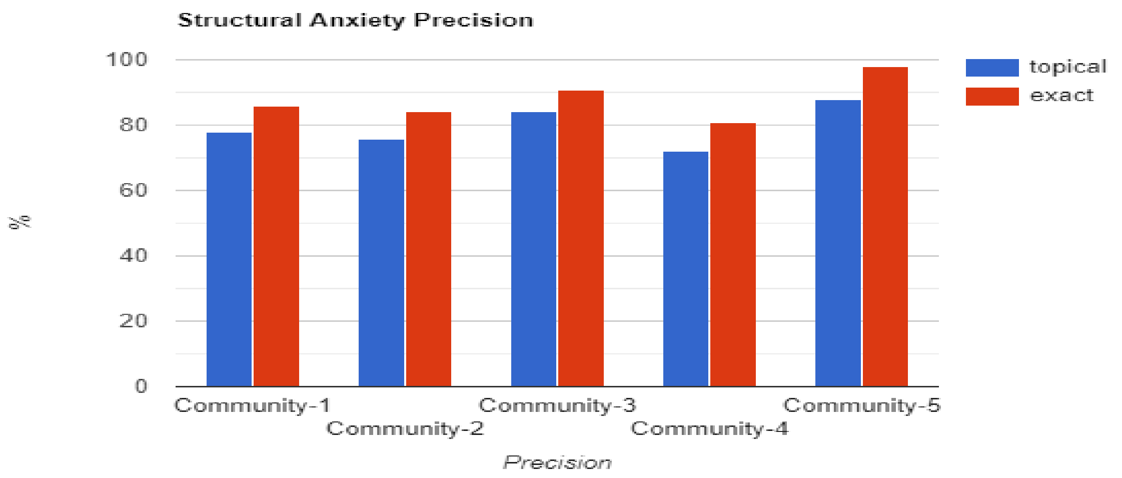

Fig.4.4. Precision of Structural Anxiety

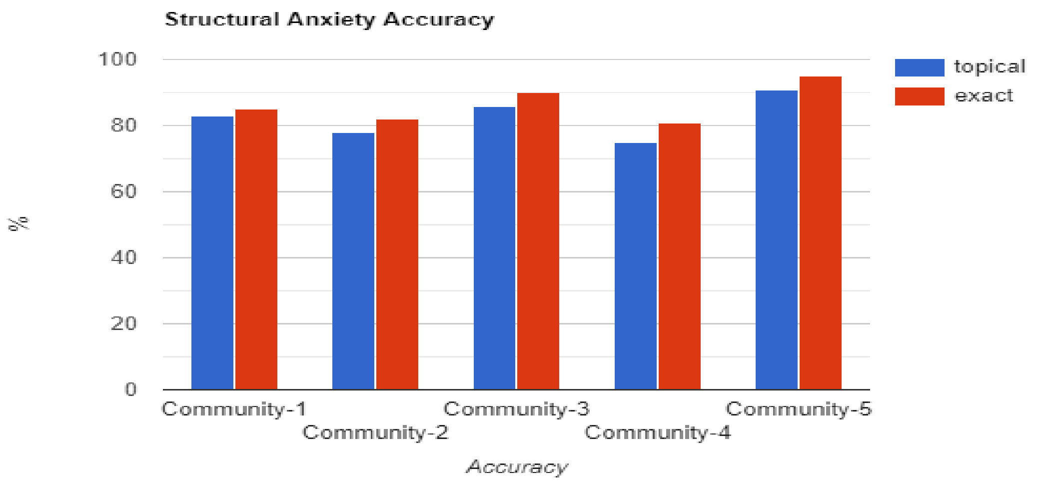

Fig.4.5. Accuracy of Structural Anxiety

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%.

4.3 Comparison

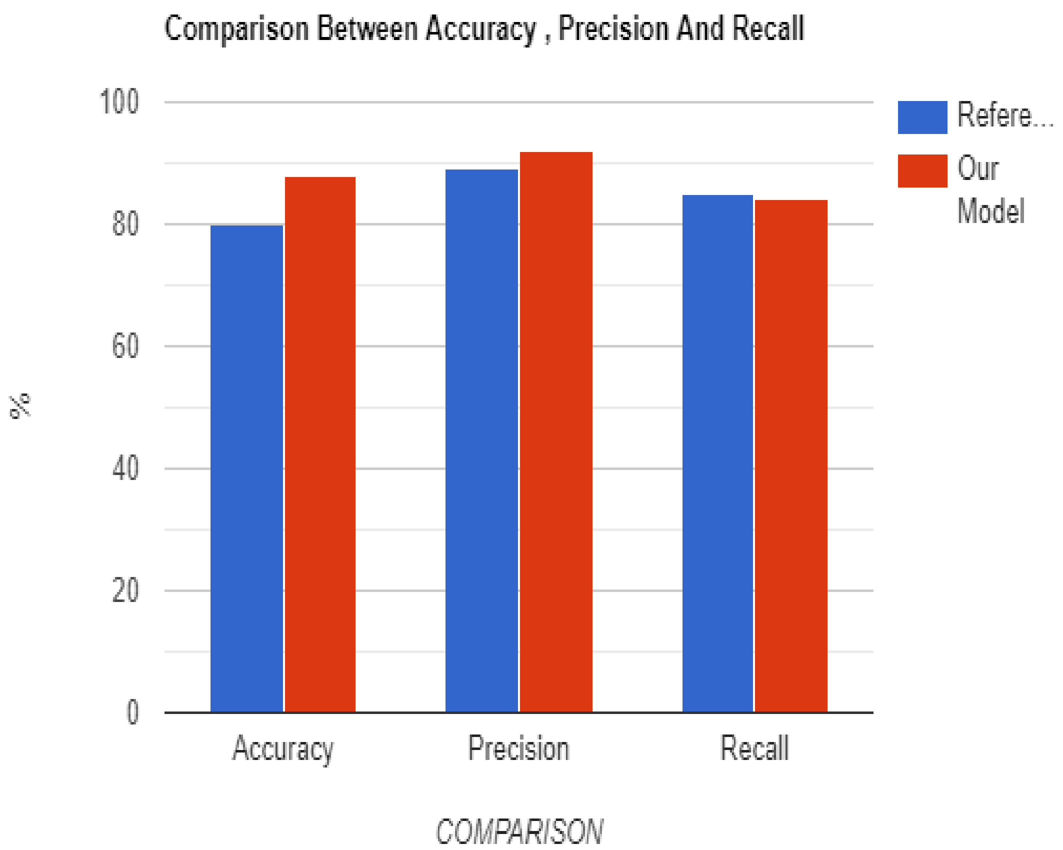

Fig.4.5. Comparison Bar Graph

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.

4.3 Outcome

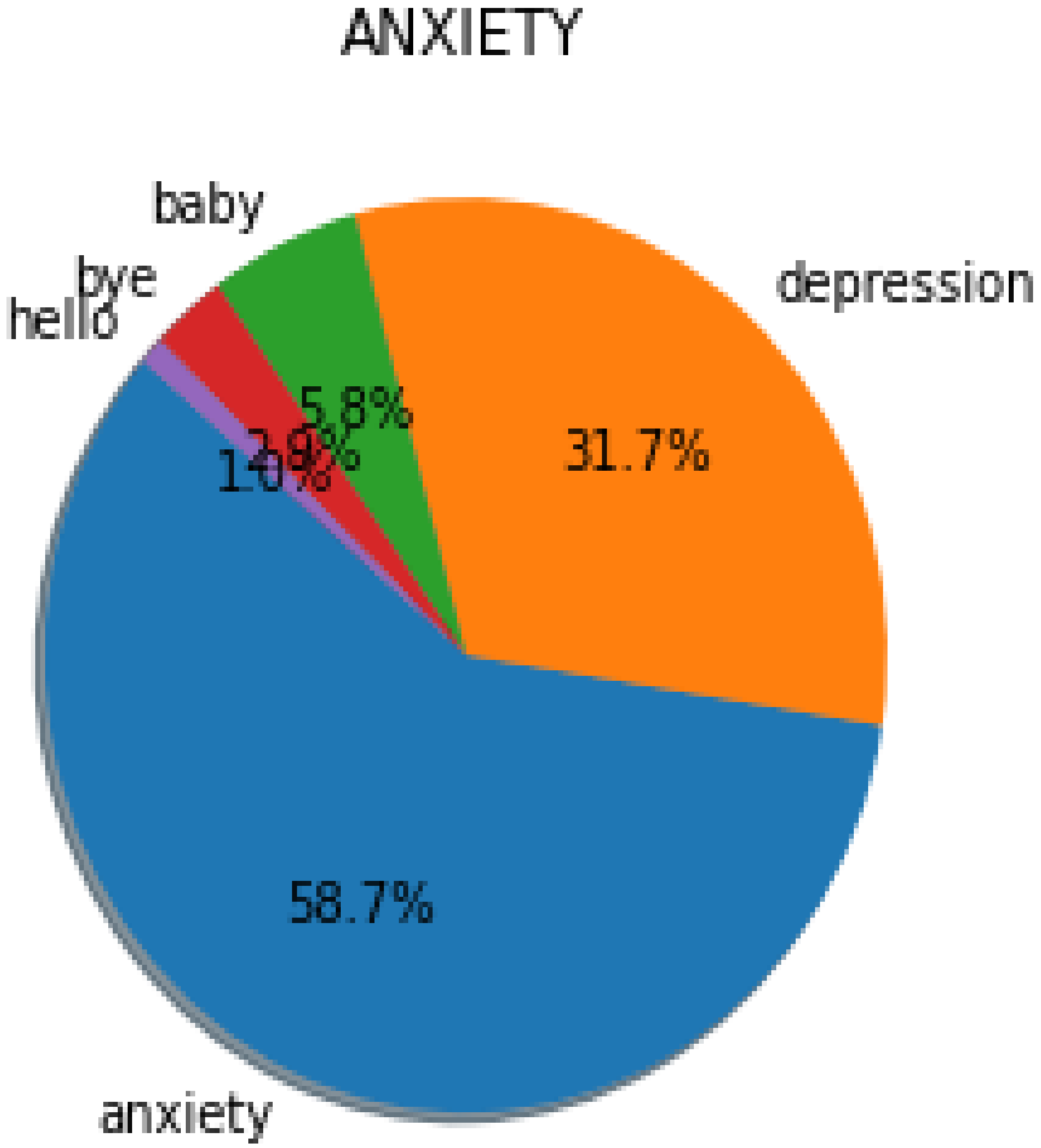

Fig.4.7. Pie chart Of Outcome

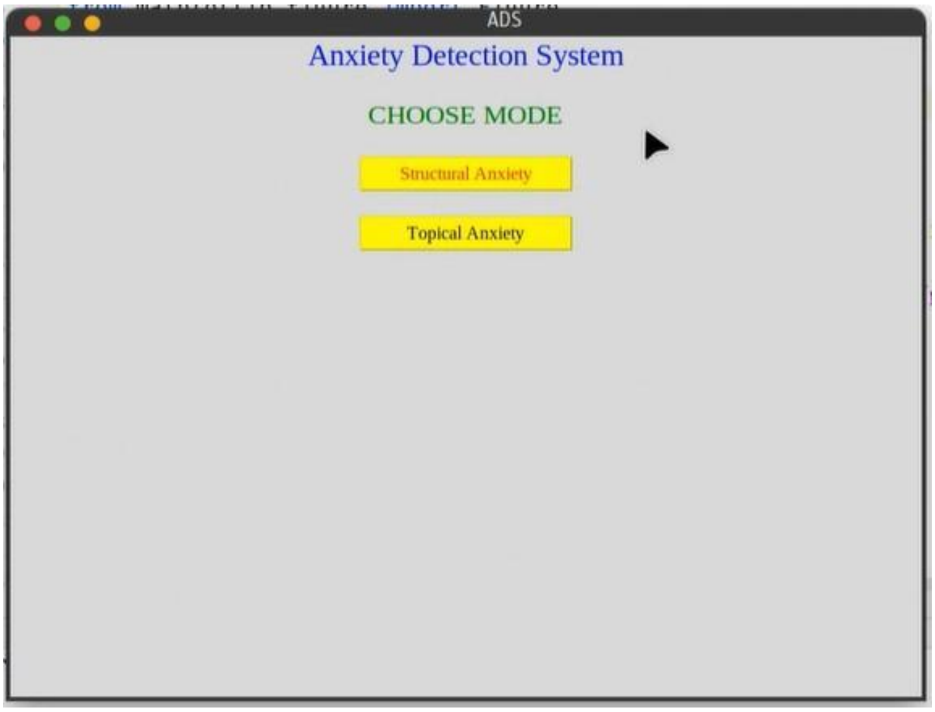

Fig.4.8. GUI of Front Page

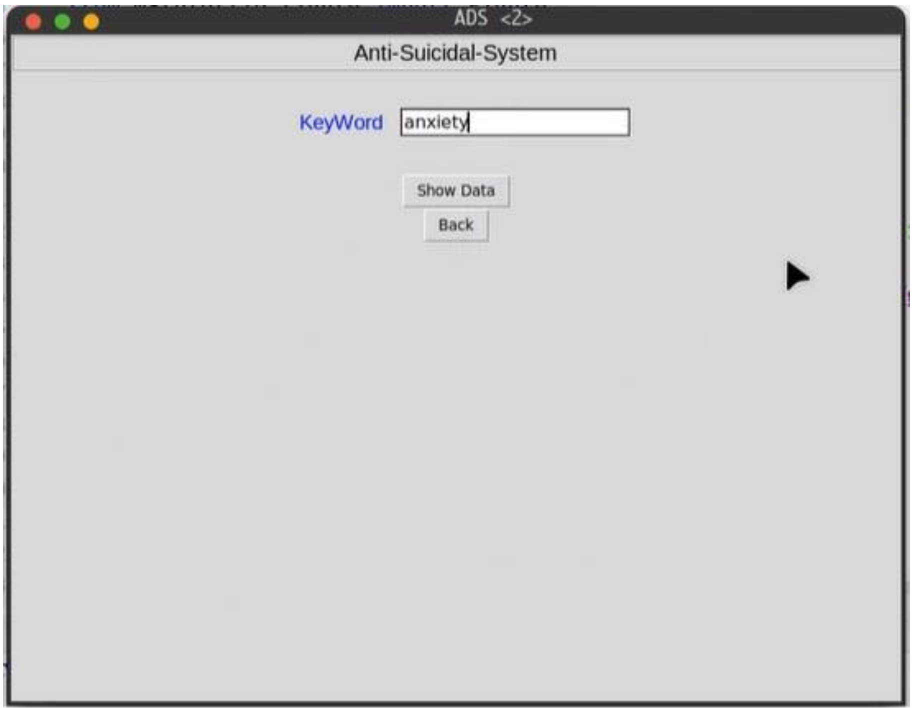

Fig.4.9. Topical Anxiety Front Page

Fig.4.10. Topical Anxiety Result

Fig.4.11. Topical Anxiety Sorted Result

Fig.4.12. Structural Anxiety Front Page

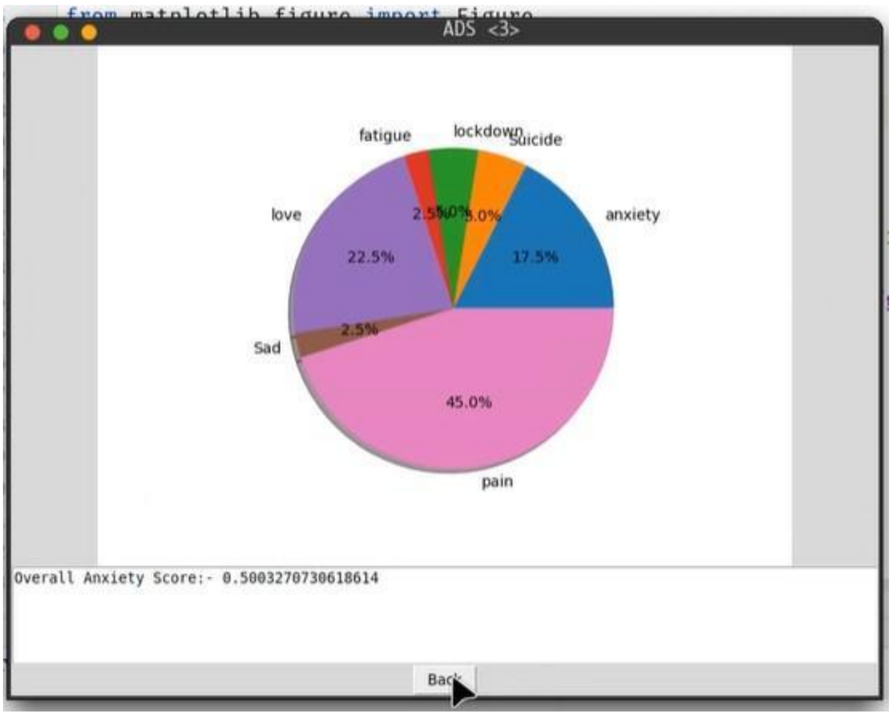

Fig.4.13. Structural Anxiety Result

Chapter 6
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.

APPENDICES

Appendix 1. 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] Singh, Harpreet, Madan M. Gupta, Thomas Meitzler, Zeng-Guang Hou, Kum Kum Garg, Ashu MG Solo, and Lotfi A. Zadeh. "Real-life applications of fuzzy logic." Advances in Fuzzy Systems 2013 (2013): 3-3.

[11] 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.

[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.

[15] Goutte, Cyril, and Eric Gaussier. "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation." In Advances in Information Retrieval: 27th European Conference on IR Research, ECIR 2005, Santiago de Compostela, Spain, March 21-23, 2005. Proceedings 27, pp. 345-359. Springer Berlin Heidelberg, 2005.