Name: Anusha Devaram

Mental Health Analysis by Reviewing the Social Media data using Artificial Intelligence Algorithm

MSc Final Project Declaration

This report is submitted in partial fulfilment of the requirement for the Master of Science in Data Science and Analytics Masters Project with Sandwich Placement at the University of Hertfordshire (UH).

It is my own work except where indicated in the report.

I did not use human participants in my MSc Project.

I here permit the report to be made available on the university website, provided the source is acknowledged.

Acknowledgement

I sincerely express my thankful gratitude to my supervisor Dr. Helen Xiang who has supported me throughout my project journey and helped me constantly and cleared my all doubts. I am accepting this open door to recognize their help and I wish that they continue to help me like this later.

# Abstract

Artificial intelligence (AI) technological solutions are valuable in many areas of behavioural and mental health care, involving clinical decision-making, therapies, diagnostics, self-care, hospital administrators, research, and more. Recent technical advancements are emphasized to demonstrate new capabilities and prospects. The advantages of using artificial intelligence in mental health care are also explored. The significant prevalence of mental disorders and the need for efficient medical care, coupled with recent developments in AI, have prompted more research into how AI may help identify, diagnose, and treat mental health disorders. AI approaches can provide new opportunities for learning about human mental state, recognizing psychological symptoms, and health conditions, which is most helpful for making illness prognosis, and improving mental health therapies. With the advancement in the online platform, it can most effective solution to identify health-related data about people. Individuals' data on social media includes unstructured and unpredictably generated information. The artificial intelligence techniques can determine the semantic features of comments which are shared on social media by users. This study proposed an LSTM framework for identifying supportive, non-supportive, and enthusiastic people efficiently and decisively. The proposed AI-based LSTM model obtained 79% accuracy and 44% loss.

Keywords: Mental health analysis, LSTM, artificial intelligence, social media for mental health analysis, and mental health analysis using medical health data.

# Table of Contents

[Abstract 1](#_Toc108539584)

[Table of Contents 2](#_Toc108539585)

[List of figures 5](#_Toc108539586)

[1. Introduction 6](#_Toc108539587)

[1.1. Aim of the project 7](#_Toc108539588)

[1.2. Background 8](#_Toc108539589)

[1.3. Hypothesis 8](#_Toc108539590)

[1.4. Objectives 9](#_Toc108539591)

[1.5. Research question 9](#_Toc108539592)

[1.6. Research contribution 10](#_Toc108539593)

[1.7. AI in healthcare 11](#_Toc108539594)

[1.8. Motivation 11](#_Toc108539595)

[1.9. Benefits of AI in health care 12](#_Toc108539596)

[1.10. Challenges 13](#_Toc108539597)

[Project plan 13](#_Toc108539598)

[2. Literature review 15](#_Toc108539599)

[2.1. Artificial intelligence in mental health study 15](#_Toc108539600)

[2.2. Mental health care using social media 16](#_Toc108539601)

[2.3. Customize mental health therapies using AI 17](#_Toc108539602)

[2.4. ML techniques in mental health diagnosis 17](#_Toc108539603)

[2.5. Early treatment for mental health issue 18](#_Toc108539604)

[2.6. Health care diagnosis project for the clinical and medical staff 19](#_Toc108539605)

[2.7. Health analysis system using social media 21](#_Toc108539606)

[2.8. Detection and evaluation of mental health insights 22](#_Toc108539607)

[3. Methodology 23](#_Toc108539608)

[3.1. Dataset 24](#_Toc108539609)

[3.2. Data preprocessing 24](#_Toc108539610)

[3.3. Removing noise and null value, stop words, null character 26](#_Toc108539611)

[3.4. Split data 26](#_Toc108539612)

[3.5. Tokenization 26](#_Toc108539613)

[3.6. Word embedding 26](#_Toc108539614)

[3.7. Word2vec 27](#_Toc108539615)

[3.8. Planed work 28](#_Toc108539616)

[3.9. Data science in healthcare 29](#_Toc108539617)

[3.10. AI in mental healthcare 29](#_Toc108539618)

[3.11. LSTM algorithm 30](#_Toc108539619)

[3.12. Consideration of ethical, legal, and professional 31](#_Toc108539620)

[4. Implementation 32](#_Toc108539621)

[4.1. Coding 33](#_Toc108539622)

5. Result……………………………………………………………………………………………47

[5.1. Performance Metrix 48](#_Toc108539623)

[6.Conclusion 49](#_Toc108539624)

[6.1 Future work 50](#_Toc108539625)

[7.References 51](#_Toc108539626)

[Appendix 53](#_Toc108539627)

# List of figures

[Figure 1: Steps involves in sentiment analysis 7](#_Toc108539528)

[Figure 2: Steps involves in extracting and performing sentiment analysis on the twitter data 8](#_Toc108539529)

[Figure 3: Analysing people mental health 12](#_Toc108539530)

[Figure 4: Project plan (Gantt chart) 14](#_Toc108539531)

[Figure 5: Sentiment analysis architecture 16](#_Toc108539532)

[Figure 6: Proposed model 23](#_Toc108539533)

[Figure 7: Selected dataset 24](#_Toc108539534)

[Figure 8: Data pre-processing 25](#_Toc108539535)

[Figure 9: Mental health analysis steps (Kanaga, 2021) 28](#_Toc108539536)

[Figure 10: LSTM cell 31](#_Toc108539537)

# Introduction

The increased mental disorders and the need for mental health care system, combined with today’s developments in AI, has prompted a surge in research into how AI may aid in the identification, diagnosis, and management of mental health issues. Artificial intelligence approaches have the potential to open new avenues for studying human behavior patterns, recognizing mental health symptoms and risk factors, making illness progression estimations, and customizing and optimizing therapies. Despite the many advantages of AI in mental health, this is a new field of study, and developing successful AI-enabled systems that can be implemented is fraught with a slew of complicated, intertwined obstacles. This project serves as an introduction to, and a comprehensive overview of, current AI work involving psychosocially related mental health issues from the computer and literature, to guide future research and identify new directions for advancing growth in this vital topic. The research reflect on the compared with the state of AI work for mental health, (i) offer specific suggestions for an enable an organization of human-centered and interdisciplinary approaches in research and innovation, and (ii) invite as much consideration of the conceivably far-reaching individual, social, and ethical issues associated that AI algorithms and intervention strategies can have if those who are to find ubiquitous, effective implementation in a real-world mental health context (Ayesha Kamran Ul haq, 2020).

Social media is becoming an increasingly important source of information for businesses. Meanwhile, more than ever, people are willing to share facts about their lives, knowledge, experiences, and thoughts with the rest of the world through social media. Everyone is actively participated in events by expressing and commenting on what is happening in society. This way of sharing knowledge and emotions with society via social media allows businesses to gather more information about the people. The gathered information has been evaluated by the expert system/algorithms to analyze mental health.

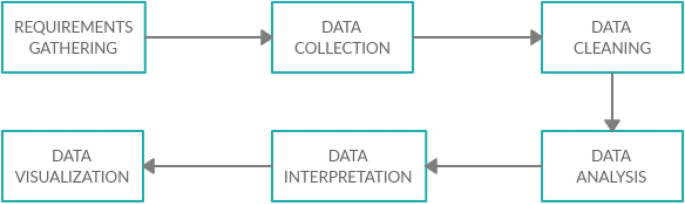


Figure 1: Steps involves in sentiment analysis

AI has made inroads into the healthcare sector in recent years. In this context, researchers have been examining large databases relating to patients' health to improve public health measures, scientific diagnostics, and patient experience. Data comes from a variety of sources, including both providers (pharmacy and patient records) and non-providers (smartphone, social media, and internet searches). The healthcare sector is one of the industries that has benefited from the use of big data. Healthcare companies have a large amount of data at their disposal, much of which is unstructured and potentially effective. The application of AI in medicine is projected to rise, and it will continue to present the profitable potential for solutions that is most prominent to improve patient lives (Sarah Graham, 2019).

To address this issue, researchers are developing AI algorithms that can analyze large amounts of raw data and extract usable information from it. A variety of AI algorithms are employed to forecast patient disease based on historical data. Wearable sensors have been created to cope with both individual and economic interactions in the real world. A high grade of affective illness, which results in significant depression and various anxiety disorders, is used to assess a person's mental health. Social anxiety, mood disorder, clinical depression, and borderline personality are just a few of the ailments that are classified as mental disorders. There are numerous mobile apps, smart devices such as smartwatches, and smart bands that expand healthcare options in mobile devices (Choudhury, 2020).

## Aim of the project

The main aim of this project is to collect data from Twitter to identify the mentality of the person. When a person posts a tweet about a particular event then manually evaluating that tweet to identify the person’s mood “Happy, Angry, or Neutral” is a challenging one. So, developing a machine-learning algorithm to take that tweet as an input to identify the person’s mentality saves time and gives the most accurate results in the classification.

## Background

When compared to face-to-face interactions with medical practitioners, social media platforms provide a readily accessible and saving more time on communication options for people with mental illnesses. Artificial intelligence-based mental health research employing massive social media data must have recently gotten a lot of attention. Because of the widespread use of social media, user communities have formed around mental health issues, particularly depression. Social media provides a rich setting for contextualizing and forecasting users' self-reported mental health burden (Kanaga, 2021). To analyze user-generated comments on social media, advanced artificial intelligence (AI) algorithms are routinely used. The research will assess the validity and reliability of these computer-based routine monitoring models about established diagnostic frameworks in an upcoming systematic review. The research will attempt to construct a normative judgment about the merits of these recent AI applications in the identification of depression from such a clinical standpoint (D.Luxton, 2016).



Figure 2: Steps involves in extracting and performing sentiment analysis on the twitter data

## Hypothesis

How does AI make mental health analysis more efficient? If I implement artificial intelligence in mental health analysis, I will increase the accuracy and research more efficiently. Artificial intelligence techniques have recently gained popularity as a unique way of creating outstanding unstructured information. Artificial intelligence is being used by psychiatry experts to gain a better understanding of the mental disease, to develop more successful and individualized treatment regimens. Artificial intelligence (AI) technology presents both immense promise and possible solutions in the field of mental health care. Researchers can use artificial intelligence to better characterize mental disorder categories and comprehend patient symptoms. The technique may one day aid in the diagnosis and treatment of psychiatric patients. Addressing the ethical concerns around AI in psychiatry could motivate physicians to use it. AI is being more commonly applied in fields such as oncology, radiology, and dermatology. On the other hand, AI's application in mental health and neuroscience studies has been limited. AI is desperately needed to help detect high-risk populations. It is most prominent to provide treatments to cure mental diseases, given the substantial morbidity and mortality in persons with psychological conditions, as well as a rising lack of mental health care professionals.

## Objectives

The main

objective of this research is to improve health by providing perfect treatment at right time. It helps to improve the quality of life for people who are affected by mental health conditions. The research has four key objectives in terms of evaluating healthcare paradigms.

* To obtain a great revolution in a mental healthcare solution that is best for present and future purposes.
* To ensure that everyone has access to basic mental health care in the foreseeable future.
* To obtain better accuracy in mental health care by implementing data science, with artificial intelligence.
* To Promote mental health information to be applied in general healthcare and education development.

## Research question

What effect does AI have on mental health?

Researchers can use artificial intelligence to better characterize mental disorder categories and comprehend mental illnesses. The technique may one day aid in the diagnosis and treatment of psychiatric patients. Addressing the ethical concerns around AI in psychiatry could motivate physicians to use it.

How does AI help with mental health analysis?

Modern artificial intelligence (AI) is being employed in the establishment of prediction, diagnosis, and treatment approaches for mental health care because of the transition to digital experiences of mental health.

## Research contribution

This research presents an effective and perspective artificial intelligence system founded on label encoder representation that can quickly detect mental health issues and concerns from user postings on social media. The research incorporates several information sources for an efficient examination of mental health-conscious data. They employed an information distillation strategy to transfer useful information from a large amount of pre-trained approaches to a smaller approach, while they examined sadness and stress data using the AI technique of LSTM. The results show this proposed approach accurately manages mixed data and improves psychological health classification ability. The significant contributions to this research project are classified here:

* A new approach is implemented for extracting a large amount of highly relevant mental health-related data using Twitter and other social media sources. Furthermore, to recognize the required mental medical issue data, they used a combination label encoder and tokenization technique based on the sequential model of emotion.
* LSTM is one of the effective classification algorithms proposed by an AI network that efficiently maximizes the quantity of knowledge accessible to the networking, increasing the material available to the proposed algorithm in recognizing what words naturally follow and precede a particular part of the sentence.
* The research is implemented with a strategy, which entails moving techniques from a large amount of pre-trained methods to a smaller approach to enhance effectiveness and overall accuracy.
* The research conducted thorough tests utilizing AI models, with both the results compared to those of other comparable models. This examination is crucial in regulating the changes and flaws in the methodologies and classification and detection techniques that have already been used. The experimental findings reveal that the developed model delivers better outcomes compared to the other approaches, LSTM obtained an accuracy of 79% after multiple hyper parameter tweaks.

## AI in healthcare

AI is now being utilized to help in early disease identification, greater comprehension of disease development, medicine dosage optimization, and the discovery of new remedies. Efficient pattern analysis of massive datasets is a major strength of AI. Ophthalmic, diagnostic tools and radiologists are among the branches of healthcare where AI technologies can execute or higher than trained physicians in reviewing images for problems or subtleties undetected to the naked eye. Although intelligent robots are uncertain of ever totally replacing clinicians, they are progressively being employed to assist in clinical decision-making. While the capacity to learn, accessibility to learning materials, and lived experience limit knowledge acquisition, AI-powered computers can rapidly synthesize information from an infinite quantity of medical data. Very huge datasets that could be studied computationally are perfect for maximizing AI's possibilities, highlighting trends and linkages about human behaviours and characters that are typically difficult for the average person to extract.

## Motivation

Artificial intelligence (AI) has infiltrated several areas that are directly relevant to daily life. As a result of technological advancement, AI for medical, which pertains to the application of artificial intelligence to authentic healthcare coverage, has emerged as one of the most pressing social challenges of the day. Numerous efforts have been made to apply AI as well as its implementations in medical services, with confidentiality as the foundation of AI-based health care. Several medical centres, researchers, and organizations have worked together to implement AI technology in the healthcare system to address concerns linked to physical wellbeing, as well as psychological health as a fast-developing social issue. Because mental health is a widespread and complex issue, detecting and exposing it is difficult. Given this trend, social media serves as a helpful tool for these folks, allowing them to create content, communicate and collaborate, and converse. Many academics have sought to evaluate hidden resources and experiences about mental health in huge consumer content on social media using artificial intelligence, which would be a powerful data-engineering tool. As a result, the following points provide us with a theoretical underpinning of linked investigations.

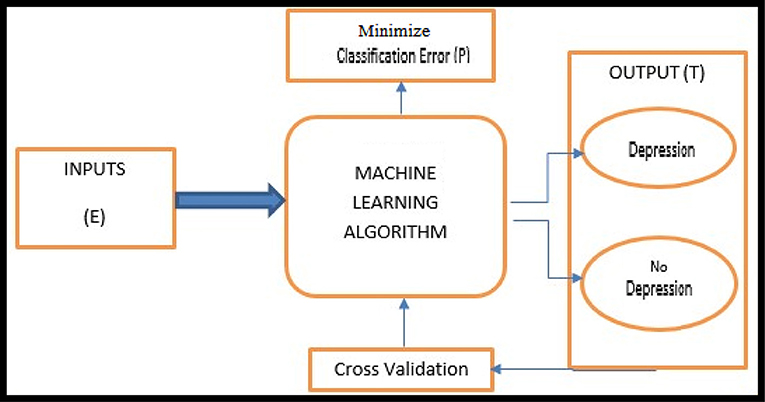


Figure 3: Analysing people mental health

## Benefits of AI in health care

* Obtaining reliable information at the right time is an important part of detecting and treating medical problems. Physicians and other medical specialists can use AI to use the real-time, important element to speed up and improve key healthcare decisions. Enhanced preventative measures, efficiency gains, and treatment wait times can all be achieved by producing more immediate and realistic outcomes. Real-time analysis is one of the major advantages to give the right treatment at the right time.
* Artificial intelligence in healthcare seems to have had a significant impact on healthcare practices around the world. Appointment systems, translation of clinical information, and patient history monitoring are among the innovations. Artificial intelligence is allowing healthcare facilities to automate more time-consuming and delicate activities. Intelligent radiology equipment, for example, can recognize important visual indicators, saving more hours of intensive analysis.
* AI helps the project save time, effort, and money while analysing mental health treatment. Medical personnel have more time to analyse patients and identify illness and ailment as more important processes are computerized. Artificial intelligence (AI) is speeding up procedures at medical facilities, allowing them to save valuable production hours. Time is money in any industry, so artificial intelligence can save a lot of money.
* AI allows researchers to collect vast amounts of data from a variety of sources. Greater analysis of this data of lethal diseases is possible thanks to the ability to draw on a large and growing data set. In terms of real-time data, research would benefit from the vast amount of data obtainable, as provided as it can be simply processed.

## Challenges

These problems include collecting large-scale, high-quality datasets that are representational of the majority's variety, as well as acquiring access to such datasets to construct more robust and equitable artificial intelligence models. For instance, mental health affects a diverse range of audiences across various demographics (ages, gender, and nationality), geographic areas, and socioeconomic levels, necessitating the participation of a diverse group of people to limit bias concerns in the collection of data. Data collecting, on the other hand, is expensive and hard, especially when the material is analysing personalized and confidential due to the social stigma frequently connected with mental illnesses. As a result, the problem of whether individuals should trust Artificial intelligence applications with their confidential data collection and information gathering, as well as pre-processing to what degree and by what channels individuals should authorize the gathering of such data, arises. These difficulties are worsened by errors, ambiguity, and bias, which make it difficult to implement "state-of-the-art" artificial intelligence in real-world intelligent platforms. Even when good accuracy is obtained, there is always the issue of generalization, where models that were trained with great accuracy in one situation may not generalize to instances from outside the training dataset's surroundings.

## Project plan

Analyze the user's mental health and get the correct polarity of the given mood. The data is got from the social media database as input to this proposed system. This system implements a variety of machine learning techniques to get the most accurate answer to a particular emotion via the appropriate classifier. The purpose of this classification task is to automatically classify mental health. Positive, negative, and neutral categories. This means choosing the appropriate class label for a particular input. Because we use monitored classifications, we need a labeled text corpus within the categories to train, test, and build classifiers.

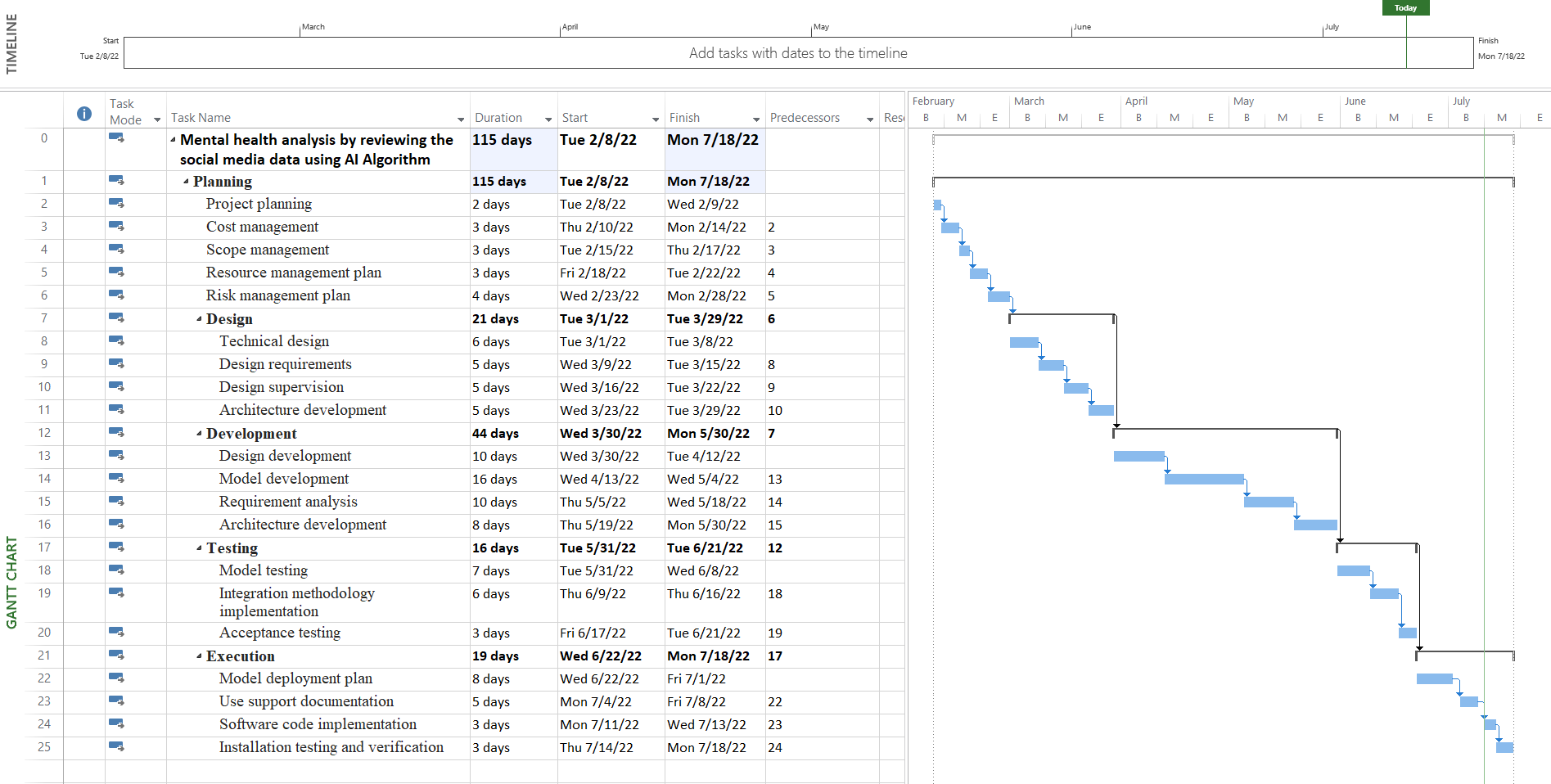


Figure 4: Project plan (Gantt chart)

# Literature review

## Artificial intelligence in mental health study

The researcher (Priscilla N. Owusu, 2021) used artificial intelligence to develop a content validity identification technique for analysing mental health. Because of the widespread use of social media, researchers have discovered that they may utilize it to collect data or information on mental health disorders, particularly depression. Users' self-reported depressed symptoms can be contextualized and forecasted via social media. Modern artificial intelligence (AI) algorithms are frequently employed to analyse user-generated comments on social media. In a forthcoming systematic review, they will analyse the reliability and validity of these computer-based routine health monitoring models about known diagnostic paradigms. From a clinical approach, they will seek to construct a normative judgment on the advantages of these current AI applications in the identification of depression. They offer a technique for a complete investigation that will consider all relevant literature from peer-reviewed sources, both direct and indirect consequences. The last review will explore depression as a self-reported health consequence in social media material. They will look at computational tools for online depressive surveillance, such as DL, ML, and AI. Additionally, recognized clinical assessments will be used as content validity indicators in the creation of the algorithms. The COSMIN framework will be utilized to evaluate the qualitative qualities of the algorithms' clinical construction. The study finishes with a normative review of the current use of AI on social media to screen for stress.

The author (Adrian B. R. Shatte, 2018) investigates the importance of mental health analysis using machine learning. This study seeks to synthesize the research on mental health by utilizing the ML technique, with a focus on the current study and practical implementations. The phrases "machine learning," and "mental health" were used to search 8 health and computer technology online databases. Two independent reviewers evaluated the papers, and information on the article's psychological health application, machine learning approach, data type and size, and research findings were retrieved. After that, a narrative review was used to synthesize the articles. Overall, the use of machine learning in mental health has yielded several improvements in the sectors of diagnosis, therapy, and support, as well as scientific and health management. With the bulk of research concentrating on the identification and diagnosis of mental health issues, there is an opportunity for ML to be used to improve other aspects of psychological performance. The benefits and drawbacks of adopting machine learning techniques are examined, as well as ways to enhance and progress the subject.

The researcher (David D.Luxton, 2018) developed an overview of mental health and behavioural identification using artificial intelligence. The smartest technique like Artificial intelligence approaches is valuable in many areas of detecting mental or physical health care, especially medical decision, therapies, evaluation, self-care, healthcare administration, research, as well as more. This research gives recent technical advancements emphasized to illustrate new capabilities and prospects.

The author (Alvaro Barrera, 2020) developed a quantitative study to analyse mental health by implementing the AI technique. Medical observations must be performed on all patients who are admitted ensuring their safe, peaceful life, and healthy. This approach assures the safety of patients, but it can also disrupt their sleep, which can have a severe influence on their rehabilitation. This research shows how artificial intelligence ('digitally aided nurse observations') was used in the analysis of mental health residential analyses while minimizing interruption to patients' sleep and ensuring their safety. This study illustrates that using digitally aided nurse monitoring in the mental ward help to keep patients safe while also improving the satisfaction of every soul who were involved in the health care system. This system obtained better results and findings in the mental health care system. These findings show that with this trending technology, the treatment delivered to a patient who is in psychiatric wards at night might be much improved. This necessitates a more in-depth and rigorous assessment.

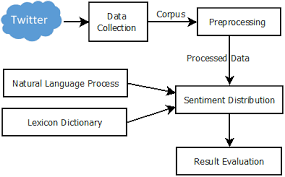


Figure 5: Sentiment analysis architecture

## Mental health care using social media

Mental health care on social media systems using AI technology is proposed by (D'Alfonso, 2020) the Simon. Early clinical improvements from mental health care may not last, necessitating the employment of longer-term preventive interventions. According to the high energy required for face-to-face primary prevention therapy, this may not be possible. Digital and advanced treatments tailored to children could be a cost-effective and entertaining approach to enhance the advantages of interventions. Traditional online therapies to provide advanced and effective treatments on mentally stressed people. More complicated models sensitive to user data are needed to guide personalized online therapy. Customer experience must be integrated with clever and proving the importance as well as the effectiveness of technology to distribute the material to rethink online interventions in teen mental health. The researcher gives an overview of the system's main characteristics and discusses the ongoing research into merging computational understanding and AI technologies to improve user interaction and therapeutic material identification and delivery.

## Customize mental health therapies using AI

Advanced AI approach was used to identify issues in the mental healthcare system (Ellen E. Lee, 2021). AI is being more commonly applied in fields such as oncology, radiology, and dermatology. On the other hand, AI's application in mental health and neuroscience studies has been limited. AI is desperately needed to help detect high-risk populations. It is most prominent to provide treatments to cure mental diseases, given the substantial morbidity and mortality in persons with psychological conditions, as well as a rising lack of mental health care professionals. While there is a lack of published investigation on AI in neuropsychiatry, there are an increasing number of great examples of AI's use including health records, neurobiological, sensor-based surveillance systems, and social networking sites to predict mental difficulties like suicidality. It presents an overview of AI techniques in mental health care, concentrating on various illustrative papers, to assist with clinical diagnosis, and therapy, as well as science and management difficulties. Although AI has the potential to help scientifically diagnose mental diseases, identify them early in their progression, customize therapies, and educate patients about their healthcare, it must address problems of bias, confidentiality, transparency, efficiency, and other ethical considerations. These ambitions represent human wisdom, which is more closely linked to individual and community well-being than IQ. As a result, future AI might deliver technology that allows varied groups of individuals to get more sensitive and morally sound treatment.

## ML techniques in mental health diagnosis

The researcher (Jina Kim, 2020) developed a system to predict mental state by analysing the social media data with the ML. When compared to questioner or interactions with medical practitioners, social media is such a great platform that provides readily accessible content for mental analysis. It is also brilliant work to save time on data collection and collaboration options for mental illness prediction. Mental health research employing ML techniques on massive social media analytics has recently gotten a lot of interest. The researcher wanted to give a study of ML technique-based mental health systems and recent developments in social media platforms. The study and Science Direct databases were used to find articles about the field of mental healthcare system and treatment. Earlier studies with a lot of citations had detailed research techniques that are documented in detail. The findings of this study show that this field of research is still evolving. Furthermore, they explored three important debate topics from a review of mental health care papers, which offer fresh in-depth possibilities for both scholars and practitioners.

## Early treatment for mental health issue

The article (Teo, 2022) was developed using a python language to diagnose mental health to prevent people by providing early treatment. Mental health issues are one of the rising issues and the demand for better medical health treatment has prompted researchers to investigate machine learning applications for mental health issues. A recent thorough assessment of machine learning methodologies in detecting mental health illness is presented in this publication. This research will also talk about the problems, constraints, and future and limitations of research for using computer vision in the mental health profession. They scan credible resources for articles and publications relating to ML methodologies in forecasting mental health disorders. Furthermore, in doing this systematic review, they use the PRISMA approach. Following the identification and assessment processes, they included a maximum of 30 research publications in this survey. This research will differentiate the collected scholarly articles depending on mental health issues. The researcher presents the implications and the obstacles and constraints that academics working on computer vision in mental health concerns confront. In addition, they make specific recommendations for future study and development in the medical field, especially in mental health with the advanced technique. The random forest method had a greater accuracy of 69 percent in this system. The Boost and SVM obtained 66 percent and 58 percent accuracy, respectively while diagnosing mental health.

The article (Konda Vaishnavi, 2022) developed early predictions of mental health to enhance the advancement and effectiveness of treatment. Rapid recognition of mental health concerns allows professionals to treat them more efficiently, enhancing the quality of life and care of patients. The psychological, physiological, and social well-being of an individual is referred to as mental health. It has an impact on how one believes, feels, and behaves. Mental health is crucial at all stages of life, including childhood to adolescence and maturity. The accuracy of five ML algorithms in detecting mental health concerns was examined using numerous accuracy measures in this research. Logistic Regression, K-NN Classifier, DT Classifier, RM classifier, and Stacking are indeed the five various ML approaches. The researcher evaluated and implemented different strategies, and the most accurate one was the Stacking strategy, which had a predictive performance of 81.75 percent.

The research (AnuPriya, 2020) to detect the mental health states using ML. Mental health concerns such as depression, anxiety, sadness, and stress are becoming quite frequent among the public in today's fast-paced environment. ML algorithms are most prominent to predict anxiousness, sadness, and anxiety in this research. The Sadness, Anxiety, and Pressure Scale interview questions were used to gather information from professional and unemployed people from various cultures and groups in possible to qualify those algorithms. Five separate machine learning techniques predicted nervousness, depression, and pressure on five degrees of severity - these would be particularly well adapted to diagnosing psychological issues due to their high accuracy. Following the application of the various approaches, the confusion matrix revealed that classes were unbalanced. As a result, the f1 score metric was added, which assisted in picking the appropriate control measures as the Random Forest classifier among some of the five used algorithms. The sensitivity parameter also demonstrated that the methods were particularly sensitive to unfavourable outcomes.

## Health care diagnosis project for the clinical and medical staff

The author (Xiaofeng Wang, 2021) developed a health care diagnosis project for the clinical and medical staff who volunteered in the pandemic situation of COVID. The most important aspect of minimizing the risk and challenges of mental disease is mental health diagnosis. Furthermore, mental health detection can serve as a theoretical foundation for public health care organizations to develop psychosocial intervention programs for medical personnel. The goal of this research is to use machine learning to analyse the medical professionals' state of mind based on 32 parameters. Through a structured questionnaire, they gathered 32 elements from 5,108 Chinese medical practitioners, and the findings of the Identity Inventory were used to assess mental health. In this paper, they offer a novel approach that relies on an optimization method and classification techniques that can identify and prioritize the most relevant elements affecting mental health. Furthermore, to forecast the psychological health of professionals, they employ sequential logistic regression, multimodal algorithm, enhanced hybrid algorithm, and the suggested classification algorithm. The findings demonstrate that the suggested model has a predictive performance of 92.55 percent, which is higher than existing techniques. This approach can be used to forecast worldwide medical employees' mental health. Furthermore, the strategy suggested in this study can aid in the enhancement of an acceptable work schedule for medical personnel.

This article (Deepali J.Joshi, 2018) explore a mental health care system to analyse the health problems. As demonstrated in the prior study, there is a critical requirement to examine and maintain a wide range of health problems. The proposed technique is a critical component of all data mining operations. They implement multi-class machine learning techniques with DL-based feature extraction techniques like sentence segmentation to diagnose people's mental health based on their social media postings and behavioural factors. Deep learning techniques have recently gained popularity as a unique way of creating outstanding unstructured information. They are not always good at encoding and decoding, but they also convey semantic information, which aids in better modelling. In comparison to their traditional equivalents, deep learning background subtraction assists in classifying normal people from abnormal people. Furthermore, the newest models have a very low percentage of false positives. The model's highest level of accuracy is 89 percent.

Mental diseases (Chang Su, 2020) are one of the most common and increased situations in the current days that may hurt one's physical wellbeing. Artificial intelligence approaches have already been launched to aid mental health clinicians, such as researchers and clinicians, in making decisions depending on mental health data that may be gathered from the clinics, social media, and other resources. Deep learning is the advanced and most popular AI technique which helps to obtain superior performance across a range of implementations, from machine learning to mental healthcare. The main aim of this developed research is to examine current research on the mental health use of DL methods. They begin by giving a brief overview of current deep learning techniques. The literature related to DL implementations in psychosocial functioning is then reviewed. The researcher categorizes these interesting publications into four categories depending on the application environments: clinical data identification and survival rate, data analysis for recognizing mental health problems, visual and audio expression analysis techniques for disease identification, and assessment of mental disorders by utilizing the data from various social media. Finally, they examine the difficulties of utilizing deep learning algorithms to increase the understanding of psychological disorders, as well as some promising directions for its use in seeking mental health evaluation and management.

The author (Stevie Chancellor, 2020) developed a model mental health analysis using various social media statuses. Nowadays, social networking sites are being employed to model emotional wellbeing and even to better understand health consequences. Quantitative research methods are increasingly being used by computer programmers to identify the presence of mental diseases and somatic symptoms such as sorrow, suicidal behaviour, and stress. This study has the prominence to improve monitoring, diagnosis, tasks, and techniques for many mental health states. There is, however, no defined approach for assessing the reliability of the research and indeed the methods used in its creation. The research undertakes a comprehensive literature analysis of the quality of using social networking site data to predict emotional wellbeing, concentrating on study design, methodologies, and research unique designs. Information characterization for mental health history, information gathering, and process improvement, pre-processing including feature selection, as well as model classification and verification are all included in the findings. Despite the increased interest in this topic, the research has noticed some troubling tendencies in content validity and an absence of introspection in the methodologies employed to quantify and define psychological status. The research offers some suggestions for addressing these issues, including a list of proposed publishing reporting requirements and collaboration prospects in this interdisciplinary field.

## Health analysis system using social media

The researcher (Patrick Robinson, 2019) explored a health analysis using social media. There are various efforts aimed at reducing the stigma associated with mental issues. However, determining how effective such are at changing attitudes is difficult. The technical use of various social media to monitor stigma amounts and highlight trends is possible. This research aims to investigate stigmatizing as well as trying to downplay attitudes about a variety of mental and physical health issues using Twitter as a venue. A maximum of 1,059,258 data were gathered, and 1300 data per grade were chosen at random for assessment. The mental-related problems were ended up finding to be more stigmatization (12.9%) and trivialized (14.3%). (8.1 And 6.8 percent, respectively). Schizophrenia is one of the most penalized mental health illnesses (41 percent), while antisocial personality disorder has been the most trivialized (33 %). The results demonstrate that there are lot of stigmas around the evaluation of the mental health of different people on social media. Trivialization is also common, implying that while the community is more open to considering mental health issues, caution should be exercised in doing so properly. This study also indicates the possibilities for social communication to be implemented to gather the general public's perceptions of mental health issues.

The rise in mental health (Teo, 2021) issues and the demand for better medical health treatment have prompted researchers to investigate machine learning applications for mental health issues. A recent thorough assessment of machine learning methodologies in predicting mental health disorders is presented in this publication. The research will also talk about the problems, restrictions, and future research directions for using machine learning throughout the mental health profession. The research scan credible resources for research reports and journals relating to machine learning methodologies in forecasting mental health disorders. Furthermore, in doing this comprehensive study, the research uses the PRISMA approach. Following the screening and selection processes, the research included a total of 30 research publications in this review. Then, depending on mental health issues such as mental disorders, depressed mood, stress disorder, and psychological issues in children, the research categorize the collected research articles. The researcher discusses the findings and the obstacles and constraints that academics working on machine learning in mental health concerns confront. In addition, the research makes specific recommendations for future study and advancements in the field of psychological health using computer vision.

## Detection and evaluation of mental health insights

The researcher (Shumaila Aleem, 2022) explored a system to detect mental health and evaluate the insights. Stress, worry, fastest and most quickly evolving lifestyles of the model day have enormous psychological consequences on people's minds all around the world throughout the years. The vast amount of data generated by worldwide technological advancements in healthcare is digitized, allowing for an even more accurate image of the diverse kinds of human physiology than old measuring procedures. Machine learning (ML) has indeed been recognized as a useful tool for processing large amounts of data in the healthcare industry. In mental health, machine learning approaches are being used to anticipate the likelihood of psychological illnesses and, as a result, to implement prospective treatment outcomes. The various machine learning techniques used to identify and evaluate depression are listed in this research article. There are three types of machine learning-based mental health detection techniques: categorization, deep learning, as well as and ensemble. A comprehensive framework for the diagnosis of mental health is described, which includes data gathering, pre-processing, ML training and exposure, detection categorization, and performance assessment. Furthermore, it provides an overview to help identify the goals and limitations of various research projects in the field of depression identification. In addition, future research opportunities about mental health diagnosis were considered.

# Methodology

The methodology and the sources used to develop mental health care system is described in this methodology part. Now, social networks are being utilized to model mental health and to better understand health consequences. Artificial intelligence approaches are being used by researchers to anticipate the presence of specific mental diseases and symptomatology including depression, suicidal behavior, and anxiety. People's electronic relationships and communication have been renewed thanks to the expansion of internet and wireless communications, particularly online social networks. In addition to providing written and video content, applications like social media allow users to share their ideas, emotions, and opinions about a matter, subject, or problem online. On the one hand, this is a fantastic way for participants of social networking websites to contribute and comment on any topic online in a public and honest manner. It permits health practitioners to learn more about what is going on in the thoughts of someone who has responded to a topic in a specific way. Artificial intelligence approaches could potentially give some unique characteristics that enable in studying the unique patterns contained in online communication and processing them to show the psychological condition to deliver such insight.

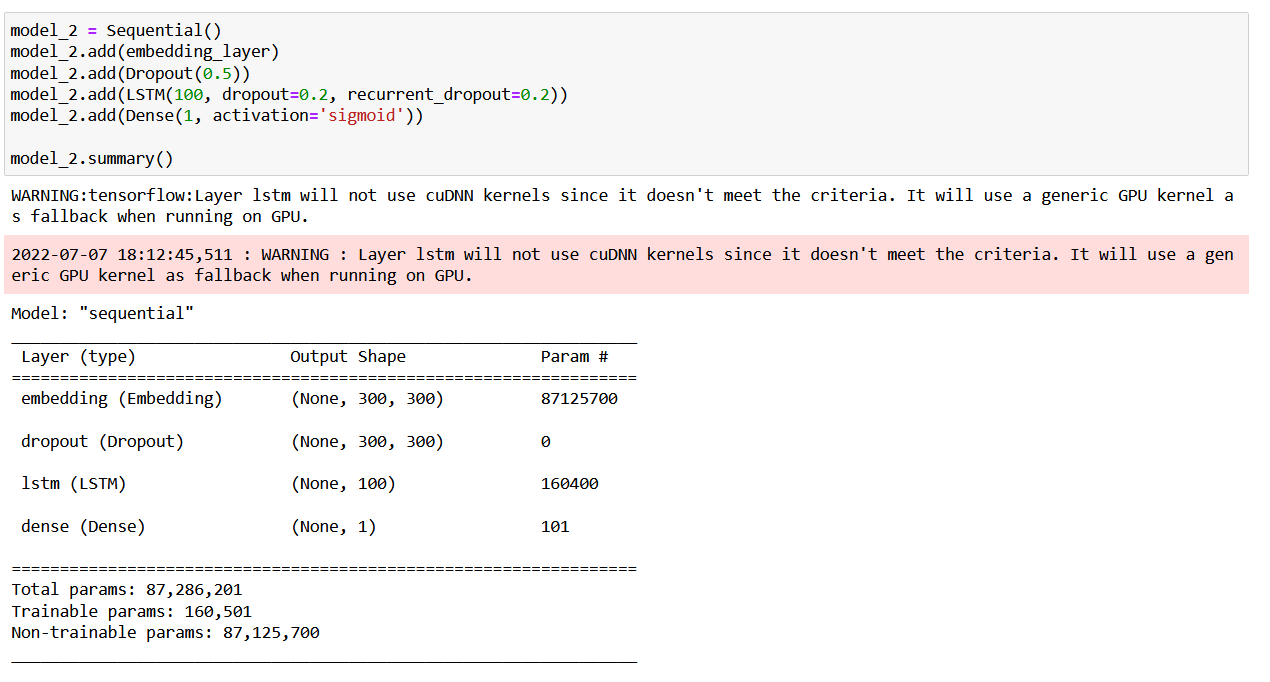


Figure 6: Proposed model

## Dataset

The selected dataset contains data about the people's states like anger, sadness, neutrality, and more. This dataset has 1600000 data which is gathered from the tweeter, so if you provide these data as input to the proposed model, you can easily learn from the dataset and get the best results. It can be displayed separately. The technique for gathering data from various sources is explored in this chapter. Using standard API wrappers, they gathered the unique data sets and stored them in a colab drive. The research chose relevant terms for Twitter that were accompanied by the hashtag’s indicator, which signifies the main idea of content for various themes. For the Twitter data, the research concentrated on certain subreddits which were relevant to the target themes, then ran a keyword search on those subreddits. Instance AI techniques are implemented that allow everyone to access the posts, tweets, and other information on these sites.

Link: <https://www.kaggle.com/datasets/kazanova/sentiment140>

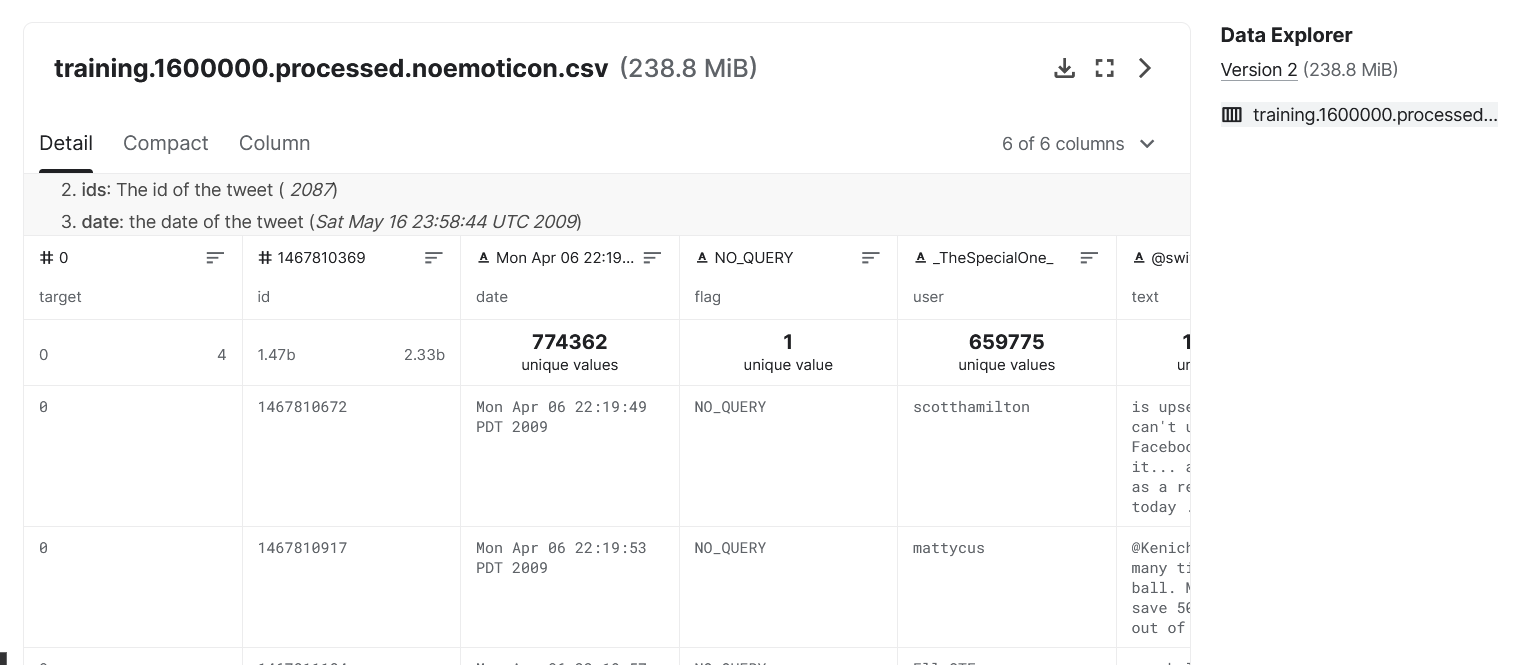


Figure 7: Selected dataset

## Data preprocessing

Different models have different variable specifications, and data training can change the generation of predictions, so it is essential to prepare the data before adopting the ensemble learning method. The goal of data preprocessing is to clean up and preprocess the data so that it is more accurate, has fewer missing values, and has a greater variance.

Data pre-processing is one of the crucial techniques for cleaning and classifying noisy and ambiguous data so that it can be used mostly for extracting the features. People transmit content on social media informally in the real world, using hashtags, symbols, special characters, and unnecessary words in their texts. Before submitting one such text document to every classification algorithm, the research must use the concepts of AI to extract and explore some sense from it. Furthermore, leveraging Emoji to replace jargon and emoji with the proper explanatory text is the focus of the efforts. The Online world is a well-known source to build a communications channel where people commonly utilize colloquial English and emoticons to express themselves. Understanding the context and importance of mental health-related data the social media posts can be accomplished by extracting significant language from the posts. The research used various preprocessing approaches to help us detect sadness and anxiousness in the gathered text, including much technical jargon including informal terms.

Original document

Removing noise and null value, stop words, null character

Split data

Word to vocabulary

Tokenization

Label encoder

Figure 8: Data pre-processing

## Removing noise and null value, stop words, null character

The elimination of words that have no important knowledge about mental problems, as a result, is discussed in this part. Pronouns, adverbs, symbols (e.g., dates, #), punctuations, and adjectives are the most frequently deemed irrelevant terms (a, an, and the). Furthermore, uniform resource locator locators (URLs) in any textual information should indeed be filtered since they include no crucial data for data preparation. To consider removing stop words from a phrase, the proposed system separates the text information and then look to see if the keyword is in the NLTK's collection of stop words, as explained above. Using Emoji to replace jargon and emoticons with underlying factual language is a crucial step since they include useful social and emotional support.

## Split data

Splitting data is one of the prominent steps where the research aims to obtain a better accuracy score and a strong solution for any problem. It also enhances the performance of training and evaluation. It is also very easy to evaluate the accuracy and loss of significant models’ training, testing, and overall accuracy score. Typically, the dataset is divided into two categories to improve the performance of prediction. Likewise, the people sentiment prediction dataset is split as a test and train dataset. The training dataset contains 1280000 data, and the test dataset contains 320000.

## Tokenization

Tokenization is a technique for dividing a massive amount of information into smaller chunks termed tokens. As seen in job 1, these tokens are used to uncover some patterns and are then used as feed during the next common processes in the NLP process, such as separating and tokenization. A huge text is made up of hash symbols, capitalization, and letters that do not even text in particular. This step is carried out in the proposed system by employing the Tokenizer from the natural language toolkit (NLTK) to clean the words known as tokens. Tokenization is a technique for breaking down phrases into words and reducing nonalphanumeric letters. Finally, a bunch of text represents all the text in the provided file.

## Word embedding

Word embedding or embedding layer is a methodology for representing words in something like a document collection with a real-valued array that captures the definition of the word, only with the expectation that words will have the same meaning. The research used a variety of word-embedding approaches to convert the collected text content (bag of words) into matrices. Word vectors are far superior to prior methods of representing words, including validity and construct validity, in which the index associated with each sentence has no semantic information. Furthermore, word vectors take up far less memory than one-hot encoding vectors while still maintaining a semantic vector representation of words. The research has outlined several real-valued vectors encoding approaches, the majority of which are supported in Python language. The word embedding layer comes from the classic context-free embedding approaches. To determine the context of a specific word in the gathered document collection, the researchers recommend using the recent, extremely smart, effective method LSTM.

## Word2vec

The word2vec method applies a text document and constructs a vector that corresponds to the input. The resulting vector will also have a large dimensionality, and each word inside the text would have its vector in feature space. The vectors are arranged in space in such a way that phrases with validated measures meanings are clustered together. Nevertheless, if a particular term was not included in training examples, the word2vec method has been unable to describe it with a matrix. The word2vec structure has two well-known implementations: continuously bag-of-words as well as skip-gram, with one convolution layer in the training process.

Word2Vec is a collection of linked models for producing word representations. These algorithms are shallow two-layered networks that represent word linguistic settings. The research must provide a corpus of words to the classifier, which generates vector representation for each word with typically huge dimensions, allowing words with similar situations to be clustered together in the vector space. The research utilized the Word embedding technique to train and then get a feature vector for each keyword in the tweets in this study. The research started the training by feeding it many tweets and creating context. The vectors are created by the Word2Vec model contextual information wherein the different terms are being used in the texts. This makes it easy to locate terms that are connected, opposite terms, and phrases and are used in comparable contexts. The research also implemented an extra filter to the word2voc Prediction model that removed any words that appeared multiple times throughout all tweets. As a result, the research disregard words for which the research lacks sufficient information or background. This accomplishes the Word2Vec methods training. As a result, when the research examined a tweet, the research divided it into a list of words, each with its vector. They calculated the mean vector using all the word representations. This vector is a representation of the entire tweet. The mean vector was calculated.

## Planed work

Data analysis is the collection, cleansing, interpretation, and compilation of data to capture crucial information about various processes. The primary purpose is to gather information from raw data. It includes steps such as acquiring data requirements, collecting data, cleaning data, analyzing data, interpreting data, and visualizing data. The requirement for data pre-processing should be discovered first. The data for the study should then be gathered from various data sources. Data cleansing is the next critical step. All undesired elements such as duplicate records, line spacing, and typos will be deleted from the obtained data to ensure that it is error-free. The cleaned and prepared data will be subjected to serious analysis in the Analysis process. The next phase is data visualization, which involves visualizing the results. The relevant data analysis procedures are shown in Figure 2.

Data collection

Data pre-processing

Feature extraction

Mental health analysis

Figure 9: Mental health analysis steps (Kanaga, 2021)

Analyze the user's mental health and get the correct polarity of the given mood. The data is got from the social media database as input to this proposed system. This system implements a variety of machine learning techniques to get the most accurate answer to a particular emotion via the appropriate classifier. The purpose of this classification task is to automatically classify mental health. Positive, negative, and neutral categories. This means choosing the appropriate class label for a particular input. Because the research use monitored classifications, it needs a labeled text corpus within the categories to train, test and build classifiers.

## Data science in healthcare

While the discipline of data science in healthcare has been expanding quickly for some time, through use of data to comprehend and treat mental health issues has trailed behind. Furthermore, in many aspects, mental health is designed specifically for data science methodologies: the stigma associated with mental illness, which frequently goes ignored and is poorly understood, offers great potential for data-driven study and solutions. With the development of new data sources and analytical methods, data science has become more and more prevalent in the healthcare industry. The utilization of data science in healthcare is undeniable. Data science in healthcare has a revolution and boomed. The following are three of the most well-known uses of data science in healthcare:

* Image recognition for medical diagnosis.
* Advancements in precision medicine that use genetic data to identify chronic diseases such as diabetes and heart problems.
* Assessments of hospital data to provide the treatments more effective and earlier.

## AI in mental healthcare

Whereas AI becomes the most prominent trending technology in medicine and hospitals for physiological healthcare applications, psychological health seems to have been a little late to accept the technologies. Most non-psychiatric professionals and healthcare practitioners are much more interactive and patient-cantered in conducting clinical practice, relying on "softer" capabilities such as developing connections with treatment and patients by personally observing patient behaviours and emotions. Medical evidence in the field of mental health is one of the highly evolving technology-based solutions. It concentrates on technical data instead of written data. Nevertheless, AI technology can enhance its impact in a different field. Likewise, it has a maximum potential to find the solution for mental health. The interactions between these physiological, psychological, and societal structures are best explained by an individual who has a unique bio-psycho-social character; nevertheless, the research has a rather narrow knowledge of the interactions between these physiological, psychological, and welfare structures. The physiology of mental disease or illness is highly spreading heterogeneous, and the discovery of indicators may enable more accurate and improved classifications of many illnesses. Work pressure, financial pressure, and most other criteria are causes of mental health issues. Using AI techniques, the research can create better pre-diagnosis assessment instruments and effective models to predict a person's proclivity for, or probability of acquiring the mental illness. They need to use computational methodologies best suited to large data to accomplish individualized mental health support as a long-term goal.

## LSTM algorithm

A mapping input object, as well as its output, is required for the categorization procedure. There are several methods for categorization, the most common of which are various classification algorithms, and Artificial Neural Networks. Each instance must be treated as a single shot in these categorization algorithms (fixed size of a vector). To put it another way, there is no internal state that is updated in real-time. As a result, if there are sequence variables, the research must figure out a means to describe time-varying information in a form that the algorithm accepts. Descriptive and inferential statistics like total, mean, sample variance, and so on are commonly used to translate time-series information into a single value. Furthermore, several tools exist for specialized areas to represent period data differently. Neither one of those techniques, however, effectively completely represent genetic sequences. Positive constants, for example, are always lacking from the data.

The research may train a classification method without using pre-processing to reduce the shape of sequential features to a single shot vector because of the architecture of the Recurrent Neural Network (RNN). Throughout model training, the fundamental state of RNN modelling is updated in real-time so that it can determine the associations between each series. RNNs can be thought of as several copies of a network in a loop, with information passing from one network to the next. As a result, it can move data from one step to the next in succession. RNN was used to achieve excellent results in natural language processing and some other types of technology with sequence data. However, due to RNN's disappearing or expanding convolution layers, it should only be used for short-term dependencies. As a result, Long Short-Term Memory was developed to tackle the issues of short-term reliance and enable the network to recall longer connections. LSTM has an internal state in contrast to the RNN structure, which allows data transformation across LSTM cells with a little modification in the information. As a result, the default behaviour of the networking is to remember long-term interactions. In addition, the LSTM structure comprises three gates that determine whether the information is sent towards the next cell.

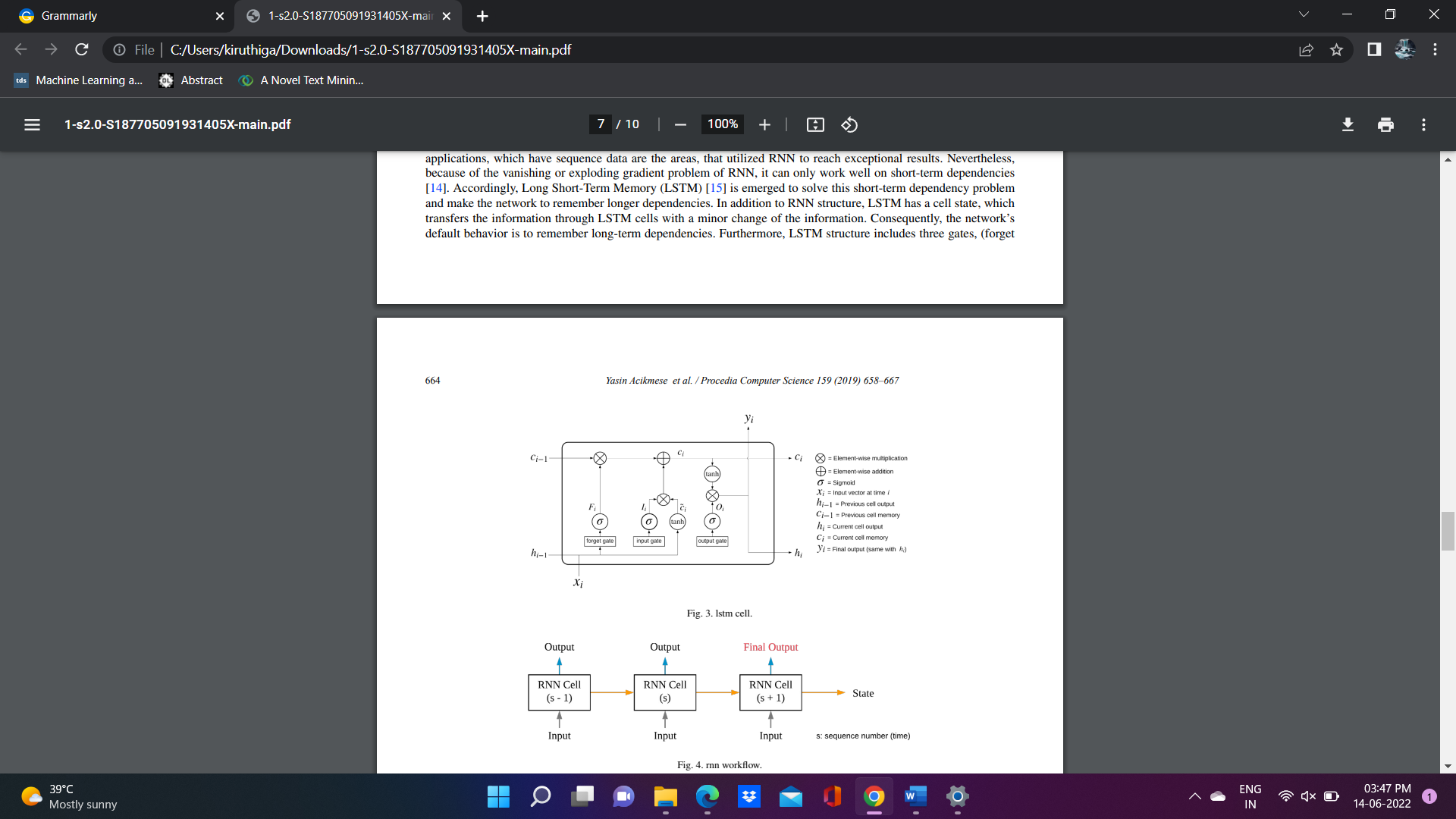


Figure 10: LSTM cell

## Consideration of ethical, legal, and professional

Ethical considerations are the backbone of the research or thesis. It is known as a rule or principle that helps the researcher to design the project in the right manner. The main criteria of research ethics are integrity, confidentiality, and following the ethics or rights of creating a thesis. Research has more concerned about the dignity of using public data. It also respects the participants and confirms the dignity and safety of using this system. This system also ensures the safety and privacy of the participant. It is highly protected and secured with confidential data.

# Implementation

Internet is the most prominent trend in this smart world. With the present improvement in the online platform, there is a thriving market for intelligent, digital, and smart technologies that can effectively address the identification of health-related difficulties on social media, including depressive symptomatology detection. These systems, which rely mostly on AI approaches, must be capable of determining the semantic features and meaning of writings shared on social media by individuals. Users' information on social media involves unstructured and unpredictably generated information. Numerous AI and social media platform-based algorithms for detecting health-related issues have recently been implemented. Nevertheless, the text encoding and intelligent systems used only supply a limited amount of information and understanding about the various texts that users have written. Word2voc and LSTM techniques were implemented to detect mental healthcare from the social media data.

Python

Python programming has been chosen as the main programming language because it contains libraries to help you develop the proposed model quickly. Python programming has several packages and libraries available to help you develop code to train and test the model.

Google colab

Google colab is the programming environment of choice for developing proposed projects. In addition, the code developed for this project runs on Jupiter Notebook, Google Colab, and many platforms.

Deliverables

Implementation of Artificial intelligence in people sentiment analysis helps to achieve a better result. Developing sentiment analysis prediction using python is added advantage to improving the speed of execution and evaluation.

## Coding

Importing libraries

|  |
| --- |
| # DataFrame  import pandas as pd  # Matplot  import matplotlib.pyplot as plt  get\_ipython().run\_line\_magic('matplotlib', 'inline')  # Scikit-learn  from sklearn.model\_selection import train\_test\_split  from sklearn.preprocessing import LabelEncoder  from sklearn.metrics import confusion\_matrix, classification\_report, accuracy\_score  from sklearn.manifold import TSNE  from sklearn.feature\_extraction.text import TfiVectorizer  # Keras  from keras.preprocessing.text import Tokenizer  from keras.preprocessing.sequence import pad\_sequences  from keras.models import Sequential  from keras.layers import Activation, Dense, Dropout, Embedding, Flatten, Conv1D, MaxPooling1D, LSTM  from keras import utils  from keras.callbacks import ReduceLROnPlateau, EarlyStopping  # nltk  import nltk  from nltk.corpus import stopwords  from nltk.stem import SnowballStemmer  # Word2vec  import gensim  # Utility  import re  import numpy as np  import os  from collections import Counter  import logging  import time  import pickle  import itertools  # Set log  logging.basicConfig(format='%(asctime)s : %(levelname)s : %(message)s', level=logging.INFO)  # In[4]:  nltk.download('stopwords') |

Before evaluating and predicting people's sentiment, the research must implement prominent libraries such as train\_test\_split, confusion\_matrix, classification\_report, accuracy\_score, and more.

Dataset

|  |
| --- |
| # DATASET  col\_names = ["Anxiety", "EntryID", "Published", "flag", "user", "Tweet"]  enc = "ISO-8859-1"  train\_size = 0.8  # TEXT CLENAING  regex\_txt = "@\S+|https?:\S+|http?:\S|[^A-Za-z0-9]+"  # WORD2VEC  word2vec\_s = 300  word2vec\_w = 7  word2vec\_e = 32  word2vec\_c = 10  # KERAS  seq\_len = 300  epoch\_num = 8  batch\_s = 1024  # SENTIMENT  pos = "Supportive"  neg = "Angry and Not supportive"  net = "Enthusiastic and Informative"  thresh = (0.4, 0.7)  # In the dataset:  # 0 -> Supportive 2 -> Enthusiastic and Informative 4 -> Angry and Not supportive  #  # As like above the dataset has been changed  # In[7]:  #Read the dataset  data = pd.read\_csv('./training.1600000.processed.noemoticon.csv', encoding =enc , names=col\_names)  data = data.dropna(subset=['Anxiety'])  data['Anxiety'] = data['Anxiety'].astype(int)  data |

Let us declare the dataset by creating the dataset columns, dataset encoding, and train size. Then the dataset attempts cleaning to remove the special characters, symbols, and other unwanted characters. The word2voc model

They needed to employ a classification approach to achieve the system objective of stress recognition due to the discontinuous outputs. The resulting label is available during the classification stage, and the algorithm attempts to divide data into discrete categories. To develop a mathematical problem, the research created three classifications ("supportive," "angry," and "excited"). "Supportive" is represented by the label "positive," whereas "angry" is represented by the label "negative." In the proposed task, "neutral" stands for "excited."

Read dataset

|  |
| --- |
| #Read the dataset  data = pd.read\_csv('./training.1600000.processed.noemoticon.csv', encoding =enc , names=col\_names)  data = data.dropna(subset=['Anxiety'])  data['Anxiety'] = data['Anxiety'].astype(int)  data |

Let us read the data from the dataset which helps to encode and evaluate the count of data in the dataset. The selected dataset contains data about the people's states like anger, sadness, neutrality, and more. This dataset has nearly 1600000 data which is gathered from the tweeter, so if you provide these data as input to the proposed model, you can easily learn from the dataset and get the best results. It can be displayed separately.

|  |
| --- |
| data['Anxiety'].value\_counts()  # In[10]:  print("df size:", len(data))  # In[11]:  maping = {0: "Supportive", 2: "Enthusiastic and Informative", 4: "Angry and Not supportive"}  def decode\_sentiment(label):  return maping[int(label)]  # In[12]:  data.Anxiety = data.Anxiety.apply(lambda x: decode\_sentiment(x))  # In[13]:  a\_cnt = Counter(data.Anxiety)  plt.figure(figsize=(16,8))  plt.bar(a\_cnt.keys(), a\_cnt.values())  plt.title("Label") |

The above code is used to decode the map which means 0 represents supportive, 2 represents enthusiastic and informative, and 4 represents angry and not supportive. Then the figure is plotted depending on the people's responses who are supportive and not supportive.

Data pre-processing

|  |
| --- |
| sw\_list = stopwords.words("english")  stem = SnowballStemmer("english")  # In[18]:  def process(t, stem=False):  t = re.sub(regex\_txt, ' ', str(t).lower()).strip()  tk = []  for token in t.split():  if token not in sw\_list:  if stem:  tk.append(stem.stem(token))  else:  tk.append(token)  return " ".join(tk)  # In[19]:  data.Tweet = data.Tweet.apply(lambda x: process(x)) |

Different models have different variable specifications, and data training can change the generation of predictions, so it is essential to prepare the data before adopting the ensemble learning method. The goal of data pre-processing is to clean up and pre-process the data so that it is more accurate, has fewer missing values, and has a greater variance.

Splitting data

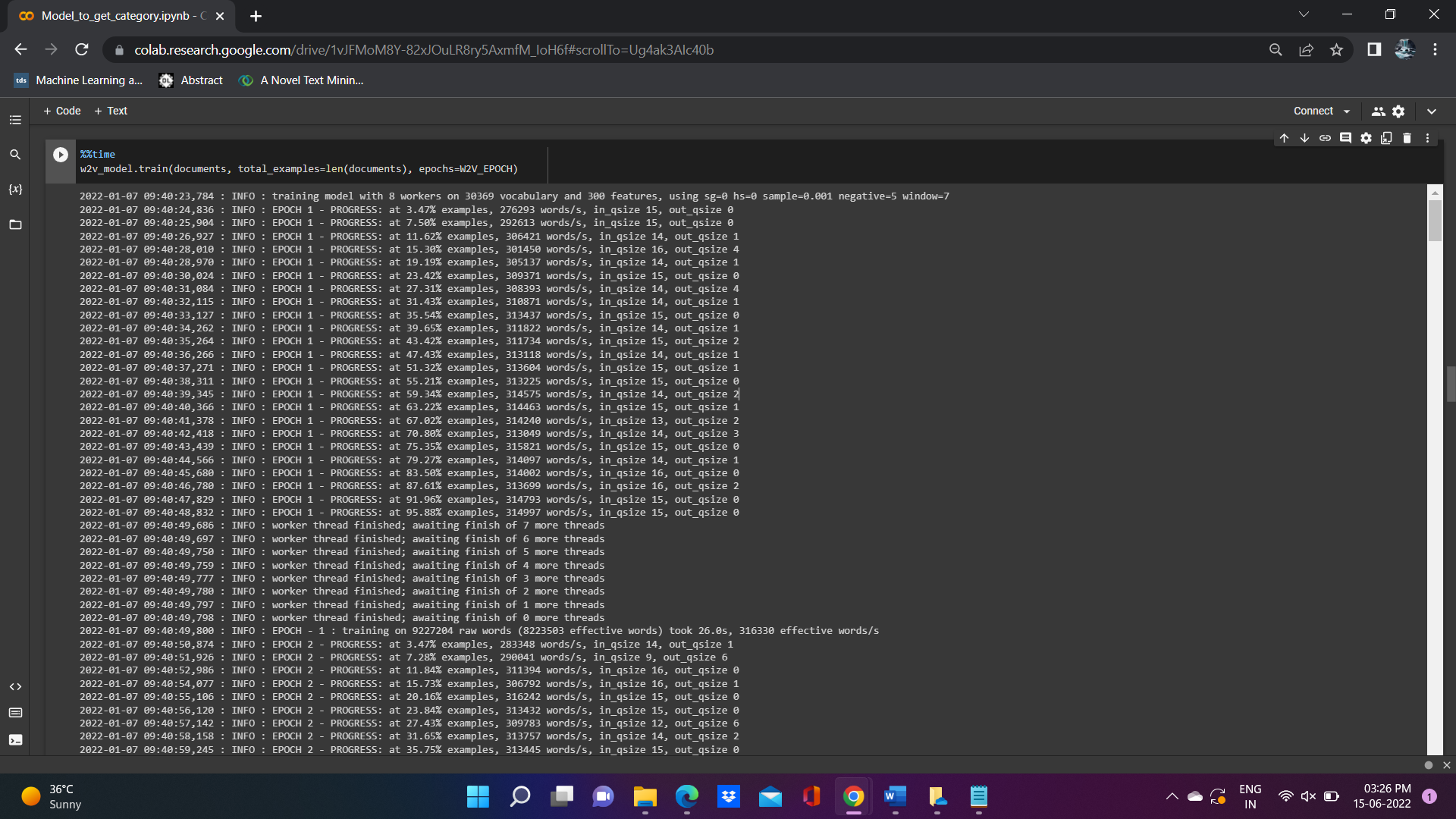
|  |
| --- |
| # ## Split Train and Test data  # In[20]:  train\_df, test\_df = train\_test\_split(data, test\_size = 1 - train\_size, random\_state = 42)  # ## word2vec  # In[21]:  doc = [txt.split() for txt in train\_df.Tweet] |

Typically, the dataset is divided into two categories to improve the performance of prediction. Likewise, the people sentiment prediction dataset is split as a test and train dataset. The training dataset contains 1280000 data, and the test dataset contains 320000.

Word2vec

|  |
| --- |
| model\_1 = gensim.models.word2vec.Word2Vec(size=word2vec\_s, window=word2vec\_w, min\_count=word2vec\_c, workers=8)  # In[23]:  model\_1.build\_vocab(doc)  # In[18]:  w = model\_1.wv.vocab.keys()  v\_size = len(w)  # In[27]:  model\_1.train(doc, total\_examples = len(doc), epochs = 32)  # In[28]:  model\_1.most\_similar("election")  # # Tokenize text  # In[29]:  tock = Tokenizer()  tock.fit\_on\_texts(train\_df.Tweet)  v\_size = len(tock.word\_index) + 1 |

Let us check the length of the words which is available in the vocab. 30369 data have been created from the vocab.



In this project Word2vocab is used to train word embedding. Word2Vec is a collection of linked models for producing word representations. These algorithms are shallow two-layered networks that represent word linguistic settings. The research must provide a corpus of words to the classifier, which generates vector representation for each word with typically huge dimensions, allowing words with similar situations to be clustered together in the vector space. The research utilized the Word embedding technique to train and then get a feature vector for each keyword in the tweets in this study. The proposed system started the training by feeding it many tweets and creating context. The vectors are created by the Word2Vec model contextual information wherein the different terms are being used in the texts. This makes it easy to locate terms that are connected, opposite terms, and phrases and are used in comparable contexts. The proposed system also implemented an extra filter to the word2voc Prediction model that removed any words that appeared multiple times throughout all tweets. As a result, the system disregard words for which the system lacks sufficient information or background. This accomplishes the Word2Vec methods training. As a result, when the system examined a tweet, the data divided it into a list of words, each with its vector. They calculated the mean vector using all the word representations. This vector is a representation of the entire tweet. The mean vector was calculated.

|  |
| --- |
| train\_x = pad\_sequences(tock.texts\_to\_sequences(train\_df.Tweet), maxlen = seq\_len)  test\_x = pad\_sequences(tock.texts\_to\_sequences(test\_df.Tweet), maxlen = seq\_len)  # # Label encoder  # In[32]:  cat = train\_df.Anxiety.unique().tolist()  cat.append(net)  cat  # In[33]:  enc = LabelEncoder()  enc.fit(train\_df.Anxiety.tolist())  train\_y = enc.transform(train\_df.Anxiety.tolist())  test\_y = enc.transform(test\_df.Anxiety.tolist())  train\_y = train\_y.reshape(-1,1)  test\_y = test\_y.reshape(-1,1) |

Tokenization is a technique for dividing a massive amount of information into smaller chunks termed tokens. As seen in job 1, these tokens are used to uncover some patterns and are then used as feed during the next common processes in the NLP process, such as separating and tokenization. A huge text is made up of hash symbols, capitalization, and letters that do not even text in particular. This step is carried out in the proposed system by employing the Tokenizer from the natural language toolkit (NLTK) to clean the words known as tokens. Tokenization is a technique for breaking down phrases into words and reducing nonalphanumeric letters. Finally, a bunch of text represents all the text in the provided file.

|  |
| --- |
| enc = LabelEncoder()  enc.fit(train\_df.Anxiety.tolist())  train\_y = enc.transform(train\_df.Anxiety.tolist())  test\_y = enc.transform(test\_df.Anxiety.tolist())  train\_y = train\_y.reshape(-1,1)  test\_y = test\_y.reshape(-1,1)  # # Embedding layer  # In[34]:  e\_mat = np.zeros((v\_size, word2vec\_s))  for word, i in tock.word\_index.items():  if word in model\_1.wv:  e\_mat[i] = model\_1.wv[word]  print(e\_mat.shape) |

**Embedding layer**

|  |
| --- |
| embedding\_layer = Embedding(v\_size, word2vec\_s, weights=[e\_mat], input\_length = seq\_len, trainable=False) |

Word embedding or Embedding layer is a methodology for representing words in something like a document collection with a real-valued array that captures the definition of the word, only with the expectation that words will have the same meaning. The research used a variety of word-embedding approaches to convert the collected text content (bag of words) into matrices. Word vectors are far superior to prior methods of representing words, including validity and construct validity, in which the index associated with each sentence has no semantic information. Furthermore, word vectors take up far less memory than one-hot encoding vectors while still maintaining a semantic vector representation of words.

Model building

|  |
| --- |
| model\_2 = Sequential()  model\_2.add(embedding\_layer)  model\_2.add(Dropout(0.5))  model\_2.add(LSTM(100, dropout=0.2, recurrent\_dropout=0.2))  model\_2.add(Dense(1, activation='sigmoid'))  model\_2.summary() |

The sequential model has been built to analysis the mental health by using social media. As a result, Long Short-Term Memory was developed to tackle the issues of short-term reliance and enable the network to recall longer connections. LSTM has an internal state in contrast to the RNN structure, which allows data transformation across LSTM cells with a little modification in the information. As a result, the default behaviour of the networking is to remember long-term interactions. In addition, the LSTM structure comprises three gates that determine whether the information is sent towards the next cell.

In this model, an embedding layer has been added. An embedding layer is used to map the sequence of images which may be used to learn the work while training the model. It is also used to convert the word to a fixed length. Likewise, dropout, LSTM, and Dense layers are also implemented in this model. The dropout layer is used to drop or strop the process before going to the next cell while the random elements occurred. Dense is the final stage of the process which helps to generate output. It is also used to implement translation, rotation, and other dimensions on the vector.

|  |
| --- |
| model\_2.compile(loss='binary\_crossentropy',  optimizer="adam",  metrics=['accuracy'])  # # Train  # In[43]:  history = model\_2.fit(train\_x, train\_y,  batch\_size = 1225,  epochs = 1,  validation\_split=0.1,  verbose=1) |

Adam optimizer is used to compile the model. In this section, the accuracy and loss value of the proposed model is evaluated. Adam optimizer is better for LSTM model development compared to another optimizer. Because it has a faster computation time than other optimizers.

|  |
| --- |
| %%time  history = model.fit(x\_train, y\_train,  batch\_size=BATCH\_SIZE,  epochs=EPOCHS,  validation\_split=0.1,  verbose=1,  callbacks=callbacks) |

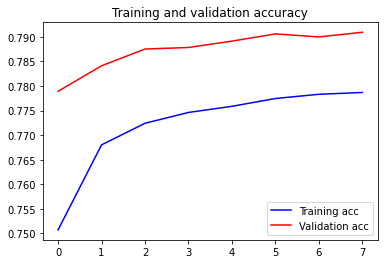
The LSTM model has been trained with the 8 epoch counts. It obtains 77% accuracy and 46% loss while training the model. It achieves 79% validation accuracy and 44% validation loss.

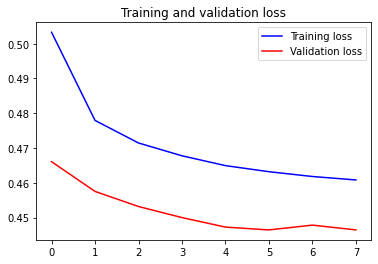
|  |
| --- |
| plt.clf()  plt.plot(history.history['accuracy'])  plt.plot(history.history['val\_accuracy'])  plt.title('model1 accuracy')  plt.ylabel('accuracy')  plt.xlabel('epoch')  plt.legend(['train', 'test'], loc='upper left')  plt.show()  plt.savefig('acc.png')  plt.clf()  # In[47]:  plt.clf()  plt.plot(history.history['accuracy'])  plt.plot(history.history['val\_accuracy'])  plt.title('model2 accuracy')  plt.ylabel('accuracy')  plt.xlabel('epoch')  plt.legend(['train', 'test'], loc='upper left')  plt.show()  plt.savefig('acc.png')  plt.clf() |

The evaluation part evaluates the model accuracy and loss. It obtained 79% accuracy and 44% loss in the proposed model.

5.Result

The research investigated the proposed model characteristics as well as effectiveness, information distillation-based approach, and state-of-the-art artificial intelligence techniques in the proposed research.



The above figure demonstrates the training and validation accuracy of the proposed model. The above figure demonstrates the training and validation loss of the proposed model.

## 5.1. Performance Metrix

The proposed research employed generally used measures like precision, recall, as well as accuracy to test the AI algorithm. A confusion matrix, also known as an error matrix, is a matrix used to evaluate classification efficiency. It depicts the number of inaccurate guesses compared to the total number of accurate predictions in a tabular fashion. The research may compute the accurateness, precision, as well as recall using the confusion matrix as follows:

The following are the most regularly used terminologies for computing the confusion matrix:

P: an immediate valid case, which would in the proposed model corresponds to the depression, anger, anxiety, enthusiastic, supportive, and non-supportive-related class.

N: a real negative instance in the proposed framework, is a depression, anger, anxiety, enthusiastic, supportive, and non-supportive-related class.

True positive (TP): a situation in which the personal information point's predicted data (actual text data) is positive (1) and the proposed model's anticipated class is also true (1).

True negative (TN): when the data point's predicted class is false, and the anticipated class is also false.

False-positive (FP): when the data point's actual class is 0 (false) yet the anticipated class is 1. (true)

False-negative (FN): when the data point's actual class is positive, but the projected class is false.

# 6.Conclusion

The proposed system is AI-based LSTM to build a robust framework for analysing mental health-related issues using social media information. The suggested framework improves the effectiveness of sophisticated healthcare systems in analysing and predicting mental health-related, such as supportive, non-supportive, and enthusiastic. These findings can be used to create a real-time technique for monitoring early mental health-related problems to find the treatment using social media. The proposed system discusses about a pre-processing component that specializes in converting large datasets into a useful form utilizing various data filtering approaches, as well as psychological textual data gathering from social media sites using application software. On the acquired text dataset, they also used a vocabulary model-based text categorization method to extract important aspects linked to mental health. Moreover, the mental medical condition identification platform employs the much more recent AI-based text quantization, which ensures that the sentimental and situational words in user postings are captured.

To increase the performance of analysing mental health utilizing an AI technique, the suggested word2voc text representation system transforms gathered words into vectors representing the semantic information in the gathered text dataset. Furthermore, the proposed system gathered data modules to prepare the dataset from social media platforms by extracting the most relevant textual information that can be analysed to create an intelligent model for advanced healthcare systems. Furthermore, the system ran an extensive experiment, demonstrating that the developed LSTM model enhances text sentiment analysis accuracy using user postings. The model beats previous machine learning or DL techniques models in part because it leverages the capabilities of both LSTM frameworks and artificial intelligence to interpret each word's structural and qualitative data. Furthermore, the system used artificial intelligence to fine-tune the problem of analysing supportive, non-supportive, and excited people, and this system achieved incredibly high accuracy.

## Future work

A multi-model mental health analysis system could be developed in the future to use more detailed information including text, visual, and behavioural aspects to get good findings. Artificial intelligence is becoming a bigger aspect of digital healthcare, and it will enhance the accuracy of mental health analysis and practice. To realize the full promise of artificial intelligence, a varied community of specialists is involved in mental health analysis and care. Artificial intelligence’s potential in mental health care seems bright. They should indeed take a more active role in trying to inform the emergence of AI into patient practice as scholars and practitioners interested in enhancing mental health care by providing financial medical knowledge and working collaboratively with data and analytical scientists, as well as some other specialists, to completely change mental health practice and better patient outcomes.

# 7.References

Adrian B. R. Shatte, D. H. a. S. T., 2018. Machine learning in mental health: A systematic scoping review of methods and applications. *Google scholar,* pp. 24 - 48.

Alvaro Barrera, C. G. A. W. O. G. D. B. a. J. G., 2020. Introducing artificial intelligence in acute psychiatric inpatient care: qualitative study of its use to conduct nursing observations. *University of Oxford,* 28(6), pp. 1 - 37.

AnuPriya, S. a. N. P., 2020. Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms. *Procedia Computer Science,* 167(27), pp. 34 - 67.

Ayesha Kamran Ul haq, A. K. N. J. M. A. N. a. F. M., 2020. Data Analytics in Mental Healthcare. *Healthcare Big Data Management and Analytics in Scientific Programming,* 4(3), pp. 1 - 323.

Chang Su, Z. X. J. P. &. F. W., 2020. Deep learning in mental health outcome research: a scoping review. *translational psychiatry ,* 16(3), pp. 1 - 37.

Choudhury, S. C. &. M. D., 2020. Methods in predictive techniques for mental health status on social media: a critical review. *NPJ digital medicine,* 6(4), pp. 34 - 89.

D.Luxton, D., 2016. An Introduction to Artificial Intelligence in Behavioral and Mental Health Care. *Artificial Intelligence in Behavioral and Mental Health Care,* 9(3), pp. 1 - 26.

D'Alfonso, S., 2020. AI in mental health. *Research Gate,* 27(34), pp. 24 - 48.

David D.Luxton, J. D. A. a. T. B., 2018. Intelligent Mobile, Wearable, and Ambient Technologies for Behavioral Health Care. *Artificial Intelligence in Behavioral and Mental Health Care,* 48(86), pp. 137-162.

Deepali J.Joshi, M. M. N. N. D. a. S., 2018. Mental health analysis using deep learning for feature extraction. *ACM digital library,* 38(18), pp. 1 - 36.

Ellen E. Lee, M. J. T. M. M. D. C. P. C. A. D. P. a. S. A. G., 2021. Artificial Intelligence for Mental Healthcare. *UC San Diego,* 16(8), pp. 1 - 87.

Jina Kim, J. L. E. P. a. J. H., 2020. A deep learning model for detecting mental illness from user content on social media. *Scientific reports,* 9(4), pp. 1 - 48.

Kanaga, N. V. B. &. E. G. M., 2021. Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review. *Springer link,* 8(6), pp. 1 - 67.

Konda Vaishnavi, U. N. K. B. A. R. a. N. V. S. R., 2022. Predicting Mental Health Illness using Machine learning algorithm. *IOP science,* 9(7), pp. 1 - 48.

Patrick Robinson, D. T. S. J. &. M. C., 2019. Measuring attitudes towards mental health using social media: investigating stigma and trivialisation. *Springer link,* 8(9), pp. 14 - 28.

Priscilla N. Owusu, U. R. G. K. I. D.-M. a. T. B., 2021. Artificial intelligence applications in social media for depression screening: A systematic review protocol for content validity processes. *Plos one,* 28(4), pp. 32 - 38.

Sarah Graham, C. D. E. E. L. C. N. X. T. H.-C. K. &. D. V. J., 2019. Artificial Intelligence for Mental Health and Mental Illnesses: an Overview. *Springer link,* 7(4), pp. 45 - 78.

Shumaila Aleem, N. u. H. R. A. S. K. S. S. A. a. A. A., 2022. Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions. *MDPI,* 28(3), pp. 1 - 20.

Stevie Chancellor, a. M. D. C., 2020. Methods in predictive techniques for mental health status on social media: a critical review. *Digital medicine,* 8(3), pp. 1 - 36.

Teo, J. C. a. J., 2021. Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. *Applied computational intelligence and soft computing.*

Teo, J. C. a. J., 2022. Mental Health Prediction Using Machine Learning: Taxonomy, Applications, and Challenges. *Applied Computational Intelligence and Soft Computing ,* 28(3), pp. 24 - 87.

Xiaofeng Wang, H. L. C. S. X. Z. T. W. C. D. a. D. G., 2021. Prediction of Mental Health in Medical Workers During COVID-19 Based on Machine Learning. *Frontiers in public health,* 37(9), pp. 1 - 87.

# Appendix

#!/usr/bin/env python

# Coding: utf-8

# In[1]:

get\_ipython().system('wget "https://storage.googleapis.com/kaggle-data-sets/2477/4140/bundle/archive.zip?X-Goog-Algorithm=GOOG4-RSA-SHA256&X-Goog-Credential=gcp-kaggle-com%40kaggle-161607.iam.gserviceaccount.com%2F20220707%2Fauto%2Fstorage%2Fgoog4\_request&X-Goog-Date=20220707T082549Z&X-Goog-Expires=259199&X-Goog-SignedHeaders=host&X-Goog-Signature=" -O data.zip')

get\_ipython().system('unzip data.zip')

# In[3]:

# DataFrame

import pandas as pd

# Matplot

import matplotlib.pyplot as plt

get\_ipython().run\_line\_magic('matplotlib', 'inline')

# Scikit-learn

from sklearn.model\_selection import train\_test\_split

from sklearn.preprocessing import LabelEncoder

from sklearn.metrics import confusion\_matrix, classification\_report, accuracy\_score

from sklearn.manifold import TSNE

from sklearn.feature\_extraction.text import TfiVectorizer

# Keras

from keras.preprocessing.text import Tokenizer

from keras.preprocessing.sequence import pad\_sequences

from keras.models import Sequential

from keras.layers import Activation, Dense, Dropout, Embedding, Flatten, Conv1D, MaxPooling1D, LSTM

from keras import utils

from keras.callbacks import ReduceLROnPlateau, EarlyStopping

# nltk

import nltk

from nltk.corpus import stopwords

from nltk.stem import SnowballStemmer

# Word2vec

import gensim

# Utility

import re

import numpy as np

import os

from collections import Counter

import logging

import time

import pickle

import itertools

# Set log

logging.basicConfig(format='%(asctime)s : %(levelname)s : %(message)s', level=logging.INFO)

# In[4]:

nltk.download('stopwords')

# In[6]:

# DATASET

col\_names = ["Anxiety", "EntryID", "Published", "flag", "user", "Tweet"]

enc = "ISO-8859-1"

train\_size = 0.8

# TEXT CLENAING

regex\_txt = "@\S+|https?:\S+|http?:\S|[^A-Za-z0-9]+"

# WORD2VEC

word2vec\_s = 300

word2vec\_w = 7

word2vec\_e = 32

word2vec\_c = 10

# KERAS

seq\_len = 300

epoch\_num = 8

batch\_s = 1024

# SENTIMENT

pos = "Supportive"

neg = "Angry and Not supportive"

net = "Enthusiastic and Informative"

thresh = (0.4, 0.7)

# In the dataset:

# 0 -> Supportive 2 -> Enthusiastic and Informative 4 -> Angry and Not supportive

#

# As like above the dataset has been changed

# In[7]:

#Read the dataset

data = pd.read\_csv('./training.1600000.processed.noemoticon.csv', encoding =enc , names=col\_names)

data = data.dropna(subset=['Anxiety'])

data['Anxiety'] = data['Anxiety'].astype(int)

data

# In[8]:

data['Anxiety'].value\_counts()

# In[10]:

print("df size:", len(data))

# In[11]:

maping = {0: "Supportive", 2: "Enthusiastic and Informative", 4: "Angry and Not supportive"}

def decode\_sentiment(label):

return maping[int(label)]

# In[12]:

data.Anxiety = data.Anxiety.apply(lambda x: decode\_sentiment(x))

# In[13]:

a\_cnt = Counter(data.Anxiety)

plt.figure(figsize=(16,8))

plt.bar(a\_cnt.keys(), a\_cnt.values())

plt.title("Label")

# # Data preprocessing:

# In[15]:

sw\_list = stopwords.words("english")

stem = SnowballStemmer("english")

# In[18]:

def process(t, stem=False):

t = re.sub(regex\_txt, ' ', str(t).lower()).strip()

tk = []

for token in t.split():

if token not in sw\_list:

if stem:

tk.append(stem.stem(token))

else:

tk.append(token)

return " ".join(tk)

# In[19]:

data.Tweet = data.Tweet.apply(lambda x: process(x))

# ## Split Train and Test data

# In[20]:

train\_df, test\_df = train\_test\_split(data, test\_size = 1 - train\_size, random\_state = 42)

# ## word2vec

# In[21]:

doc = [txt.split() for txt in train\_df.Tweet]

# In[22]:

model\_1 = gensim.models.word2vec.Word2Vec(size=word2vec\_s, window=word2vec\_w, min\_count=word2vec\_c, workers=8)

# In[23]:

model\_1.build\_vocab(doc)

# In[18]:

w = model\_1.wv.vocab.keys()

v\_size = len(w)

# In[27]:

model\_1.train(doc, total\_examples = len(doc), epochs = 32)

# In[28]:

model\_1.most\_similar("election")

# # Tokenize text

# In[29]:

tock = Tokenizer()

tock.fit\_on\_texts(train\_df.Tweet)

v\_size = len(tock.word\_index) + 1

# In[30]:

train\_x = pad\_sequences(tock.texts\_to\_sequences(train\_df.Tweet), maxlen = seq\_len)

test\_x = pad\_sequences(tock.texts\_to\_sequences(test\_df.Tweet), maxlen = seq\_len)

# # Label encoder

# In[32]:

cat = train\_df.Anxiety.unique().tolist()

cat.append(net)

cat

# In[33]:

enc = LabelEncoder()

enc.fit(train\_df.Anxiety.tolist())

train\_y = enc.transform(train\_df.Anxiety.tolist())

test\_y = enc.transform(test\_df.Anxiety.tolist())

train\_y = train\_y.reshape(-1,1)

test\_y = test\_y.reshape(-1,1)

# # Embedding layer

# In[34]:

e\_mat = np.zeros((v\_size, word2vec\_s))

for word, i in tock.word\_index.items():

if word in model\_1.wv:

e\_mat[i] = model\_1.wv[word]

print(e\_mat.shape)

# In[35]:

embedding\_layer = Embedding(v\_size, word2vec\_s, weights=[e\_mat], input\_length = seq\_len, trainable=False)

# # Build the model

# In[36]:

model\_2 = Sequential()

model\_2.add(embedding\_layer)

model\_2.add(Dropout(0.5))

model\_2.add(LSTM(100, dropout=0.2, recurrent\_dropout=0.2))

model\_2.add(Dense(1, activation='sigmoid'))

model\_2.summary()

# # Compile model

# In[37]:

model\_2.compile(loss='binary\_crossentropy',

optimizer="adam",

metrics=['accuracy'])

# # Train

# In[43]:

history = model\_2.fit(train\_x, train\_y,

batch\_size = 1225,

epochs = 1,

validation\_split=0.1,

verbose=1)

# # Evaluate

# In[44]:

r = model\_2.evaluate(test\_x, test\_y, batch\_size = batch\_s)

# In[45]:

r

# In[46]:

plt.clf()

plt.plot(history.history['accuracy'])

plt.plot(history.history['val\_accuracy'])

plt.title('model1 accuracy')

plt.ylabel('accuracy')

plt.xlabel('epoch')

plt.legend(['train', 'test'], loc='upper left')

plt.show()

plt.savefig('acc.png')

plt.clf()

# In[47]:

plt.clf()

plt.plot(history.history['accuracy'])

plt.plot(history.history['val\_accuracy'])

plt.title('model2 accuracy')

plt.ylabel('accuracy')

plt.xlabel('epoch')

plt.legend(['train', 'test'], loc='upper left')

plt.show()

plt.savefig('acc.png')

plt.clf()

# In[48]:

def ds(s, i=True):

if i:

l = net

if s <= thresh[0]:

l = pos

elif s >= thresh[1]:

l = neg

return l

else:

return pos if s < 0.5 else neg

# In[52]:

def category(text, include\_net=True):

x = pad\_sequences(tock.texts\_to\_sequences([text]), maxlen = seq\_len)

score = model\_2.predict([x])

l = ds(score)

return l

# # Testing the model with single values:

# In[53]:

category("Congratulations to Nawaz Sharif. Now he can vote in 2023 election directly from the UK")

# In[54]:

category("Transparency, scrutiny and electoral accountability are central characteristics of democracy. So why is your government trying to demolish them?")

# In[56]:

category('know jacob rees mogg became politician worked offshore investment firm chose run election put charge emerging markets created 8 billion funds bet uk lobbied brexit sackreesmogg')

# In[58]:

category('less minute vote welshyouthparliament election starts vote today')

# In[ ]: