# MED EASY

## Project report in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology

**In**

## Computer Science & Engineering

**Submitted By**

Pallab Banerjee Enrollment No. 12020009001083

Tamal Das Enrollment No. 12020009001079

Trishita Biswas Enrollment No. 12020009001133

Shayanika Das Enrollment No. 12020009001099

Sayak Mondal Enrollment No. 12020009001066

Rajib Parbat Enrollment No. 12020009001178

Soumoja Gupta Enrollment No. 12020009001103

Srijeet Roy Enrollment No. 12020009001064

Debadyuti Paul Enrollment No. 12020009001063

Swastik Dhar Enrollment No. 12020009001107

Suman Halder Enrollment No. 12020009001224

**Under the guidance of**

Prof. Varsha Poddar

&

Prof. Bipasha Mukhopadhyay

Department of Computer Science & Engineering



UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA

University Area, Plot No. III – B/5, New Town, Action Area – III, Kolkata – 700160.

## CERTIFICATE

This is to certify that the project titled **MEDEASY** submitted by Pallab Banerjee **(University Roll No.** 12020009001083**),** Tamal Das **(University Roll No.** 12020009001079**)**, Trishita Biswas **(University Roll No.** 12020009001133**),** Shayanika Das **(University Roll No.** 12020009001099**)**,Sayak Mondal **(University Roll No.** 12020009001066), Rajib Parbat **(University Roll No.** 12020009001178**),**  Soumoja Gupta **(University Roll No.** 12020009001103**),** Srijeet Roy **(University Roll No.** 12020009001064**),** Debadyuti Paul **(University Roll No.** 12020009001063**),** Swastik Dhar **(University Roll No.** 12020009001107**)** andSuman Halder **(University Roll No.** 12020009001224**)** students of **UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA**, in partial fulfillment of requirement for the degree of Bachelor of **Computer Science Engineering**, is a bonafide work carried out by them under the supervision and guidance of Prof. Varsha Poddar & Prof. Bipasha Mukhopadhyay during 7th Semester of academic session of 2023- 2024. The content of this report has not been submitted to any other university or institute. I am glad to inform that the work is entirely original and its performance is found to be quite satisfactory.

Signature of Guide Signature of Guide

Signature of Head of the Department

## ACKNOWLEDGEMENT

We would like to take this opportunity to thank everyone whose cooperation and encouragement throughout the ongoing course of this project remains invaluable to us.

We are sincerely grateful to our guide Prof. Varsha Poddar and Prof. Bipasha Mukhopadhyay of the Department of Computer Science Engineering, UEM, Kolkata, for his wisdom, guidance and inspiration that helped us to go through with this project and take it to where it stands now.

Last but not the least, we would like to extend our warm regards to our families and peers who have kept supporting us and always had faith in our work.

Pallab Banerjee

Tamal Das

Trishita Biswas

Shayanika Das

Sayak Mondal

Rajib Parbat

Soumoja Gupta

Srijeet Roy

Debadyuti Paul

Swastik Dhar

Suman Halder

# TABLE OF CONTENTS

**ABSTRACT.............................................................................................1**

## CHAPTER – 1: INTRODUCTION……………………………………………..2 - 3

**CHAPTER – 2: LITERATURE SURVEY……………………………………..4 - 5**

## CHAPTER – 3: PROBLEM STATEMENT……………………………………6

**CHAPTER – 4: PROPOSED SOLUTION……………………………………...7**

## CHAPTER – 5 : EXPERIMENTAL SETUP AND RESULT ANALYSIS…….8 - 15

**CHAPTER – 6 : CONCLUSION & FUTURE SCOPE…………………………16**

## REFERENCES…………………………………………………………………16 - 17

**ABSTRACT**

Issues related to mental health are increasingly becoming recognized as legitimate medical concerns in the workplace. To mitigate mental health concerns among employees, companies need to identify the viewpoints that bear the majority of the blame. Classification techniques are therefore needed to determine whether or not a representative requires treatment for mental health issues. We performed some pre-processing on an open-source survey dataset before selecting features of NLP. We then classified the dataset using a few machine learning algorithms. The activities were modeled as a bag of words using a probabilistic Naive Bayes (NB) Classifier with an accuracy of 94.07% and Random forest with accuracy of 82.2% to predict mood outcomes. We divide the analysis into two parts: a personalized model and a generic model in which we combined the data from all participants.Finally, we used a voting classifier that adds together all of the findings from every classifier that is fed into it and forecasts the result based on the most notable majority of votes. Other machine learning models used were Logistic Regression with accuracy of 72.67%. By combining the aforementioned techniques, Voting Classifier improves our accuracy to 90.48%.

**INTRODUCTION**

Our lives revolve around our work. Most of our time is spent at work, where we also receive our pay and frequently meet new people. Having a rewarding career is essential for maintaining excellent mental health and overall wellbeing. Sometimes life gets to us too much, mostly because of work-related issues like having to put in extra hours or taking time off owing to deadlines. Other problems, such as those involving our circumstances, relationships, or family, occasionally arise. We must protect that respect by attending to the mental health of those who are already struggling, those who are at risk, and the workforce at large while they are at work. Negative work environments have a negative impact on mental health. Good managers who look out for people and bring out the best in them are essential to a good workplace. Research demonstrates that people with sound mental health are very productive.

Mental health illnesses are currently among the most burdensome health issues. For example, in late June 2020, forty percent of American adults reported struggling with mental health issues or substance abuse. Every year, one in six American youth aged 6 to 17 struggles with psychological wellbeing issues. By the age of 14, and by the age of 24, 75% of all lifetime dysfunctional conduct has begun. The nation loses over $210.5 billion a year due to despair alone [1]. Diabetes, heart disease, and other physical health issues are linked to mental health illnesses. Patients who have both physical and mental health problems incur far higher treatment costs. Opening out about mental illness often makes people uncomfortable, especially in the workplace. Celebrities and other people have started sharing their stories these days, which has increased social awareness. Employers have an obligation to establish a work environment that genuinely supports employees' mental health. To encourage their employees to enhance their mental health, employers must create a work environment that stresses psychological components and offer pertinent resources.

Globally, an estimated 264 million people experience depression. These folks have anxiety symptoms, which impede their capacity to function. According to a recent study conducted by the World Health Organization (WHO), lost productivity from mental health illnesses costs the world economy $1 trillion USD annually [2]. Unemployment is another significant element in mental health. This could result in drug or alcohol abuse, absenteeism, etc. People who are employed also need to think about their workplace. A positive work environment boosts output, which generates revenue. Approximately 13% of children, 46% of teenagers, and 19% of adults suffer from mental illness each year [3]. However, it is concerning to note that only 50% of individuals afflicted obtain therapy, as most people find it awkward to disclose. Improper treatment of mental illness can result in increased medical costs, an increased risk of suicide, subpar work output, and other issues.

This initiative aims to determine and quantify the human elements that influence the probability of seeking mental health assistance. The feelings that people experience when they seek assistance are categorized according to the aspects of human nature that either favorably or unfavorably impact their mental health.

The issue at hand in this study is mental health, which may be examined utilizing sentiment analysis on our platform and natural language processing. Our platform contains

user-generated material with comments from other users that are gathered for sentiment

analysis. Text mining, which may extract data from our platform using data crawling techniques, will be utilized to gather comments in order to gain a high number of comments [14], [15]. The data acquired from our platform is quite unstructured, in this regard, text preparation techniques need to be employed. Unstructured data that still contains unnecessary letter and number symbols can be altered at any of these stages [16], [17]. When performing data analysis for sentiment analysis, data processing is a crucial step. In order to perform sentiment analysis, algorithms must be used. One of these is using the naïve Bayes algorithm, which has been shown to be capable of classifying comments with positive, negative, and neutral sentiments. This allows the algorithm to see the opinions included in the comments on our site. Sentiment analysis-generated opinions can reveal information about the mental health of users on our platform [18], [19].

## LITERATURE SURVEY

Use of Machine Learning Models

1. Naive Bayes

The Naïve Bayes technique applies the Bayes theorem to data classification. Although this may not always be the case in practice, this method operates under the assumption that every feature (or attribute) in the data is independent of one another [22], [23]. As a result, this approach is referred to as "naive," which is straightforward or simple. In essence, Naïve Bayes uses the Bayes theorem to determine the likelihood of every class (or target) in the data, taking into account the features (or characteristics) present in the data. Stated otherwise, Naïve Bayes determines the likelihood that the input data falls into a particular class by evaluating each attribute's value. Applications requiring data classification, like document classification, sentiment analysis, and email categorization as spam or not, frequently use Naïve Bayes. Naïve Bayes has the advantage of being fast and simple to build, and it can handle a large amount of feature-rich data [24].

1. Random Forest

A subset of the bagging approach, the RF is insensitive to parameter initialization and performs better when there is noise and weak discrimination data [68]. Every iteration, an RF uses bootstrap to randomly select many samples for the purpose of developing a decision tree (without pruning), and it builds multiple decision trees to construct the entire RF. Next, every tree "votes" for the output, which is the most popular class.

Widely employed in numerous fields, RF-based feature selection (permutation) is reliable for variables involving high-dimension and high-order correlation [30]. For a given variable xi, for instance, the permutation technique breaks the initial correlation between xi and the outcome Y by replacing every xi with a random value and classifying the permutation as noise. In the meantime, the RF establishes which factors should be employed to reduce a predictor's purity decline utilizing the Gini coefficient. A function from the add-in package RF called random forest cross-validation (rfcv) for feature selection is frequently used to determine the variable number in order to reduce the dimension of the data. Through the use of nested cross-validation, this function gradually lowers the number of predictors (ranked according to variable relevance). The RF model is generally less sensitive to parameter settings than some other predictors. The optimization of an RF is mostly based on two parameters: m-try and n-trees, where m-try is the number of variables used in splitting a node and n-trees is the number of trees in an RF. The optimal value of m-try is determined by traversing all possible values.

According to the research of Breiman [22], the generalization error of the RF converges as the number of trees increases, a characteristic absent in most other classifiers. In other words, the model performs better and better as the value of n-trees increases [29]. Therefore, the optimization of n-trees is to balance the classification accuracy with computational effectiveness.

1. Sentiment Analysis

Sentiment analysis is a technique in natural language processing that is used to recognize and evaluate viewpoints or feelings that specific people or groups of people convey in written or spoken materials. Understanding how an individual or group reacts favorably or unfavorably to a subject, good, brand, or service on social media or other online platforms is the aim [29]. A variety of approaches, such as machine learning, natural language processing techniques, and classification algorithms, are available for sentiment analysis [30]. Sentiment analysis is typically applied in a business setting to ascertain a customer's viewpoint or impression of a specific good or service. By doing this, companies are able to better understand the requirements and issues of their customers and improve the caliber of their goods and services. Sentiment analysis is also utilized in the social and political spheres to comprehend public opinion on significant social topics, leaders, and policies. In the domains of business, politics, and society, sentiment analysis results can aid in decision-making and the development of stronger strategies [31], [32].

## PROBLEM STATEMENT

Our lives are intricately intertwined with our work, where we spend a significant portion of our time, earn our livelihoods, and often forge new connections. A fulfilling career is not just about financial stability; it's also crucial for maintaining optimal mental health and overall well-being. However, the demands of work can sometimes become overwhelming, whether it's dealing with long hours, tight deadlines, or balancing personal and professional responsibilities. These challenges, compounded by factors such as relationship issues or family dynamics, can significantly impact our mental health.

Recognizing the importance of mental health in the workplace is essential, as negative work environments can exacerbate existing mental health issues and contribute to a decline in overall well-being. Conversely, supportive and positive work environments fostered by empathetic and effective managers can enhance mental health and boost productivity. Research consistently shows that individuals with good mental health tend to be more productive, underscoring the importance of prioritizing mental well-being in the workplace.

The prevalence of mental health disorders is a significant public health concern, with millions of individuals worldwide experiencing conditions such as depression and anxiety. Despite the widespread impact of these disorders, stigma and discomfort often prevent people from seeking help, particularly in professional settings. However, there is a growing movement toward destigmatizing mental health issues, thanks in part to public figures and celebrities sharing their own experiences.

Employers have a crucial role to play in promoting mental health awareness and support in the workplace. By creating environments that prioritize psychological well-being and offering relevant resources and support, employers can encourage employees to prioritize their mental health. This not only benefits individuals but also contributes to a more productive and successful workforce.

Unemployment is also a significant factor affecting mental health, leading to issues such as substance abuse and absenteeism. Even those who are employed must consider the impact of their workplace environment on their mental well-being. Unfortunately, despite the prevalence of mental health disorders, many individuals do not seek treatment due to barriers such as stigma and lack of access to resources.

Addressing these barriers requires a comprehensive understanding of the human factors that influence individuals' willingness to seek mental health assistance. By exploring the emotions and perceptions associated with seeking help, we can better identify strategies to promote mental health awareness and support in both professional and personal spheres. Ultimately, prioritizing mental health not only improves individual well-being but also has far-reaching benefits for society as a whole.

## PROPOSED SOLUTION

In efforts to address mental health concerns among employees, companies are increasingly turning to data-driven approaches, leveraging classification techniques to identify key factors contributing to these issues. To achieve this, extensive preprocessing of open-source survey data is conducted, followed by the selection of relevant features using natural language processing (NLP) techniques. This process involves transforming textual responses into a structured format suitable for analysis.

Once the data is prepared, various machine learning algorithms are employed for classification tasks. Among these, probabilistic Naive Bayes (NB) Classifier and Random Forest algorithms have demonstrated promising accuracy rates. The NB Classifier achieves an impressive accuracy of 94.07%, while Random Forest achieves 82.2% accuracy in predicting mood outcomes based on the input features.

To further enhance the analysis, the approach is divided into two main parts: a personalized model and a generic model. In the personalized model, data from individual participants is analyzed independently, allowing for tailored insights into each individual's mental health status.

Conversely, the generic model combines data from all participants, enabling broader trends and patterns to be identified across the entire workforce.

In addition to these individual models, a voting classifier is employed to consolidate findings from multiple classifiers. This ensemble technique aggregates predictions from each classifier and generates a final prediction based on the most commonly agreed upon outcome. By incorporating insights from various models, including Logistic Regression with an accuracy of 72.67%, the voting classifier achieves an improved accuracy rate of 90.48%.

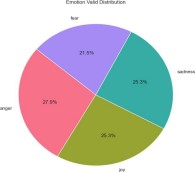
Overall, this comprehensive approach to mental health analysis leverages a combination of preprocessing techniques, feature selection, and machine learning algorithms to identify and address mental health concerns among employees effectively. By harnessing the power of

data-driven insights, companies can take proactive steps to support the mental well-being of their workforce, ultimately fostering a healthier and more productive work environment.

## ENVIRONMENT SET UP

Model 1 - Emotion Detection from Text

One of the challenging challenges in natural language processing is textual emotion recognition. The problem's multi-class nature and the lack of a labeled dataset are the causes. Due to the wide range of emotions that people experience and the challenge of collecting sufficient data for each emotion, class imbalance becomes an issue. The aim is to create an effective model for recognizing emotions using the labeled data that we have here. There are 40000 records total in our collection, representing 13 distinct emotions. Thus, developing an effective multiclass classification model is difficult. In order to create an effective model, we might need to employ certain cutting-edge techniques and rationally minimize the number of classes.



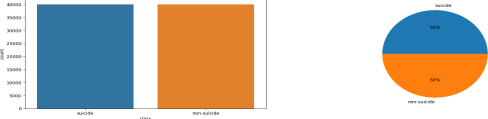
Model 2 - Stress Detection from text Dataset:

Texts about individuals dealing with anxiety, depression, and other mental health conditions are gathered in the Mental Health Corpus. The corpus is divided into two columns: the comments are in one, and labels indicating whether or not the remarks are poisonous are in the other. The corpus can be used for many different things, including sentiment analysis, mental health language analysis, and the detection of hazardous language. Researchers, mental health practitioners, and others interested in comprehending the language and attitude surrounding mental health concerns may find the data in the corpus interesting.

Model 3 - Suicide Text Detection Dataset:

There were no publicly available datasets when I was thinking about developing a text classifier to identify suicide ideation. I hope this will save time and be helpful to anyone searching for datasets on suicide detection.The dataset consists of posts from the Reddit platform's "depression" and "SuicideWatch" subreddits. The Pushshift API is used to gather

the posts. postings on "SuicideWatch" from its founding on December 16, 2008, to January 2, 2021, were all gathered, while postings on "depression" were gathered between January 1, 2009, and January 2, 2021. Posts from the depression subreddit are categorized as depression, whilst all posts gathered from SuicideWatch are labeled as suicide. We gather non-suicide posts from r/teenager.



Text processing

* 1. Data Processing –

Tokenization - The process of tokenizing a stream of text involves dividing it up into words, phrases, symbols, or other significant components, which are referred to as tokens. Tokenization is an essential step in natural language processing since it enables robots to comprehend and analyze human language. The fundamental components of many NLP tasks, including text categorization, sentiment analysis, and machine translation, are tokens.

* + 1. Text-Cleaning - Text cleaning, also referred to as text preparation, is the act of getting unprocessed text data ready for tasks involving natural language processing (NLP). Text must be cleaned and

standardized through a number of procedures in order to be better prepared for analysis. Here are a few typical methods for cleaning text:

Lowercasing: To maintain consistency, all text should be converted to lowercase (for example, "Hello" and "hello" are considered as the same word).

Special characters should be eliminated since they might not be necessary for analysis. These characters include those that are not letters, numerals, or spaces.

Numbers should be removed since they might not be necessary for some text analysis jobs.

Eliminating excess whitespace: To standardize whitespace, eliminate any extra tabs, spaces, or newline characters.

Eliminating stop words: Get rid of frequent terms with little significance, Including "the," "and," and "is."

Lemmatization and stemming: To standardize variants, reduce words to their basic or root

form (e.g., "running" to "run").

Feature Extraction-

Bag of Words(BOW) - A "bag of words" is a straightforward and widely used method for text feature extraction in natural language processing (NLP) and text mining. It depicts a text as a multiset (or "bag") of its words, preserving diversity while ignoring word order and syntax. Every word is regarded as a feature, and each word's frequency serves as a numerical feature vector. This method works well for problems like sentiment analysis, document clustering, and text categorization.

Tf-idf - Term Frequency-Inverse Document Frequency is referred to as TF-IDF. It's a statistical metric used in text mining and information retrieval to assess a word's significance inside a document in relation to a corpus of texts. phrase Frequency (TF) counts how often a word or phrase appears in a document. It is computed by dividing the total number of words in a text by the frequency with which a phrase occurs.

- Inverse Document Frequency (IDF)\*\*: Indicates how significant a phrase is across the board in the corpus. The computation involves dividing the total number of documents in the corpus by the number of documents that include the phrase, resulting in the logarithm. The product of a term's TF and IDF scores yields its TF-IDF score in a document. It shows the degree to which a phrase is pertinent to a certain corpus document. High TF-IDF terms are frequently seen as being more significant or indicative of the document's substance.

Word Embedding - In natural language processing (NLP), word embedding is a method for representing words as dense vectors in a continuous vector space. The closeness of a word's related vectors indicates how similar it is to another word. This is in contrast to conventional techniques such as one-hot encoding, which use sparse vectors to represent words, with a vector containing largely zeros and a single one for each word.

Word2Vec, GloVe (Global Vectors for Word Representation), and FastText are some of the methods often used to learn word embeddings from huge text corpora. These techniques capture contextual and similarity associations between words, which makes them valuable for a variety of natural language processing (NLP) applications, including sentiment analysis, machine translation, and language modeling.

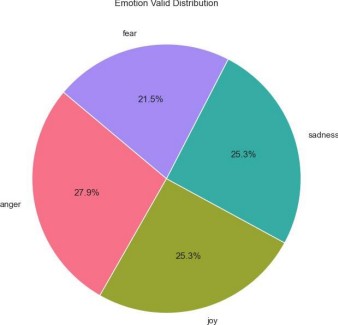
Model Selection –

Emotional Analysis - Training and Testing split of Dataset. Model architecture is defined using a combination of CNN and text-embedding Two branches of the model process the text index input independently, each following consisting of an embedding Layer followed by ID convolutional Layer and pooling.

The outputs of both branches are concatenated and paned through dense Layer for further processing Dropout Layers are used for regularization.

The output Layer with SoftMax activation predicts the probability distribution over the four emotion claves.

Model compilation: The model then compiled with an optimizer, has function and evaluation matrix.



Suicide Level/Stress Prediction – Training and Testing split of Dataset

Binary Classification Problem –

Classifying things into one of two classes or categories is the aim of a binary classification task, a kind of machine learning issue. A common representation of the two classes is 0 and 1, or negative and positive.

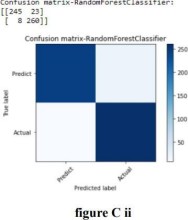
For instance, the objective of spam email detection is to categorize emails as either class 1 (spam) or class 0 (not spam). Similar to this, the aim of a medical diagnostic may be to categorize patients as belonging to a certain class (class 1) or not (class 0).

A machine learning model is trained using a labeled dataset, in which every object is linked to a class label, in order to solve a binary classification issue. In order to forecast fresh, unknown data, the model discovers patterns and correlations in the data. Neural networks, decision trees, logistic regression, and support vector machines (SVM) are popular techniques for binary classification.

Models taken for training

**Regression algorithm:** It is a machine learning technique that uses a classification algorithm to find potential candidates for a categorical dependent variable. Finding the precise model to elucidate the relationship between a set of variables and the separate characteristic of interest (measured variable = reply or final output variable) is the aim of logistic regression. Figure C i mentions the confusion matrix's ultimate output.



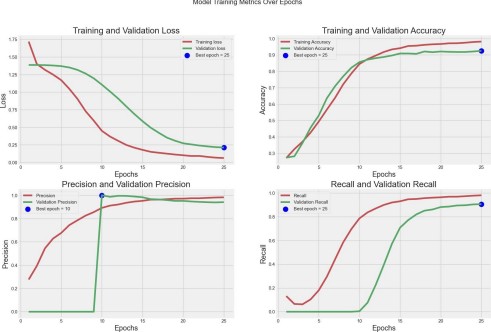


## RESULT & DISCUSSION

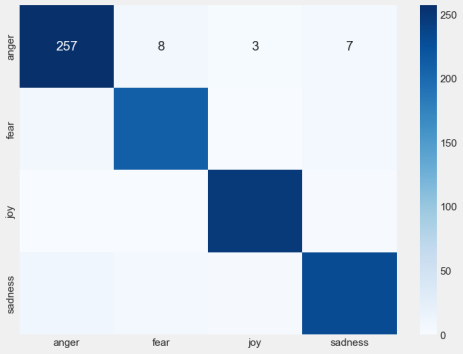
In emotion analysis the model came to be 98% accurate after training

→The model has been trained for 25 epochs.

→Model Training matrix over epochs.



→ Confusion matrix

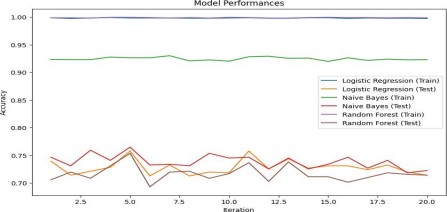


for Stress prediction and suicide prediction there are three models taken for prediction:

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Training Score | Test Score | Difference |
| Logistic Regression | 99.8 | 72.6 | 27.2 |
| Naïve Bayes | 92.7 | 75.0 | 17.7 |
| Random Forest | 99.9 | 72.0 | 27.9 |

Training Score: How well the dataset fits with model during training Test Score: How well the model performing with the test dataset.

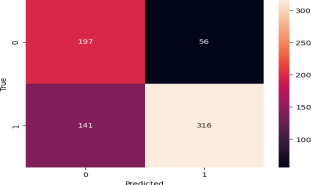
→ Performance of various model with graph



→ Model Accuracy 33.1%

→Confusion matrix

→ Models performance is determined using the method Bagging.



## CONCLUSION & FUTURE SCOPE

Data purification, lost value analysis, probing analysis, and model building and evaluation were the first steps in the analytical process. A greater accuracy score will be found on the public test set, which has the best efficiency. This app can assist in determining the mental health predictions, and further work for this project may

1. Establishing a link between the cloud model and mental health prediction. Streamlining the implementation process in an AI setting.

## REFERENCES

1. Jung Yuchae, Yong Ik. Multimedia Tools andApplications, 76 (9) (2020),

pp.11305-11317 ViewPDF CrossRefView Record in ScopusGoogle Scholar nonparametric analysis of EEG signals. Diss.UniversitiTeknologi MARA, 2020. Google Scholar

1. Norizam, Sulaiman. Determination and classification of human stress index using the nonparametric analysis of EEG signals. Diss. UniversitiTeknologi MARA, 2020. Google Scholar
2. <http://www.mindgarden.com/documents/PerceivedStressScale.pdf> Google Scholar
3. Predicting Anxiety, Depression and Stress in Modern Life using Machine Learning Algorithms (2020). <https://doi.org/10.1016/j.procs.2020.03.442>
4. Stress Detection Using Machine Learning Algorithms(2020) . <https://www.journals.reaim.com/ijresm/article/download/171/154/315>
5. Mental Stress Level Prediction and Classification based on Machine Learning (2020) <https://ieeexplore.ieee.org/document/95888>
6. S. Lu, L. Zhao, L. Lai, C. Shi and W. Jiang, "How do Chinese people view cyberbullying? A text analysis based on social media", *Int. J. Environ. Res. Public Health*, vol. 19, no. 3, pp. 1822, 2022. <https://www.mdpi.com/1660-4601/19/3/1822>
7. Z. Xue, Q. Li and X. Zeng, "Social media user behavior analysis applied to the fashion and apparel industry in the big data era", *J. Retail. Consum. Serv.*, vol. 72, pp. 103299, 2023. <https://www.sciencedirect.com/science/article/abs/pii/S0969698923000462?via%3Dihub>
8. M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh and H. A. Gozali, "Improving text preprocessing for student complaint document classification using sastrawi", *IOP Conference Series: Materials Science and Engineering*, vol. 874, no. 1, pp. 12017, 2020. <https://iopscience.iop.org/article/10.1088/1757-899X/874/1/012017>
9. A. F. Hidayatullah and M. R. Ma'Arif, "Pre-processing tasks in Indonesian Twitter messages",

*Journal of Physics: Conference Series*, vol. 801, no. 1, pp. 12072, 2017. <https://iopscience.iop.org/article/10.1088/1742-6596/801/1/012072>

1. M. Wongkar and A. Angdresey, "Sentiment analysis using Naïve Bayes Algorithm of the data crawler: Twitter", *2019 Fourth International Conference on Informatics and Computing (ICIC)*, pp. 1-5, 2019.

[https://scholar.google.com/scholar?as\_q=TikTok+as+a+Knowledge+Source+for+Programming+Learner](https://scholar.google.com/scholar?as_q=TikTok%2Bas%2Ba%2BKnowledge%2BSource%2Bfor%2BProgramming%2BLearners%3A%2Ba%2BNew%2BForm%2Bof%2BNanolearning%3F&as_occt=title&hl=en&as_sdt=0%2C31) [s%3A+a+New+Form+of+Nanolearning%3F&as\_occt=title&hl=en&as\_sdt=0%2C31](https://scholar.google.com/scholar?as_q=TikTok%2Bas%2Ba%2BKnowledge%2BSource%2Bfor%2BProgramming%2BLearners%3A%2Ba%2BNew%2BForm%2Bof%2BNanolearning%3F&as_occt=title&hl=en&as_sdt=0%2C31)

1. Z. Li, R. Li and G. Jin, "Sentiment analysis of danmaku videos based on naive bayes and sentiment dictionary", *IEEE Access*, vol. 8, pp. 75073-75084, 2020.

[https://scholar.google.com/scholar?as\_q=Sentiment+analysis+of+danmaku+videos+based+on+naive+ba](https://scholar.google.com/scholar?as_q=Sentiment%2Banalysis%2Bof%2Bdanmaku%2Bvideos%2Bbased%2Bon%2Bnaive%2Bbayes%2Band%2Bsentiment%2Bdictionary&as_occt=title&hl=en&as_sdt=0%2C31) [yes+and+sentiment+dictionary&as\_occt=title&hl=en&as\_sdt=0%2C31](https://scholar.google.com/scholar?as_q=Sentiment%2Banalysis%2Bof%2Bdanmaku%2Bvideos%2Bbased%2Bon%2Bnaive%2Bbayes%2Band%2Bsentiment%2Bdictionary&as_occt=title&hl=en&as_sdt=0%2C31)

1. C. Villavicencio, J. J. Macrohon, X. A. Inbaraj, J.-H. Jeng and J.-G. Hsieh, "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naive bayes", *Information*, vol. 12, no. 5, pp. 204, 2021. <https://www.mdpi.com/2078-2489/12/5/204>
2. S. Yadav, M. Timbadia, A. Yadav, R. Vishwakarma and N. Yadav, Crime Pattern Detection Analysis & Prediction, pp. 225-230, 2017.

[https://scholar.google.com/scholar?as\_q=Crime+Pattern+Detection%2C+Analysis+%26+Prediction&as](https://scholar.google.com/scholar?as_q=Crime%2BPattern%2BDetection%2C%2BAnalysis%2B%26%2BPrediction&as_occt=title&hl=en&as_sdt=0%2C31)

[\_occt=title&hl=en&as\_sdt=0%2C31](https://scholar.google.com/scholar?as_q=Crime%2BPattern%2BDetection%2C%2BAnalysis%2B%26%2BPrediction&as_occt=title&hl=en&as_sdt=0%2C31)

1. S. Kurniawan and I. Budi, "Indonesian tweets hate speech target classification using machine learning", *2020 5th Int. Conf. Informatics Comput. ICIC 2020*, pp. 1-5, 2020

[https://scholar.google.com/scholar?as\_q=Indonesian+tweets+hate+speech+target+classification+using+](https://scholar.google.com/scholar?as_q=Indonesian%2Btweets%2Bhate%2Bspeech%2Btarget%2Bclassification%2Busing%2Bmachine%2Blearning&as_occt=title&hl=en&as_sdt=0%2C31) [machine+learning&as\_occt=title&hl=en&as\_sdt=0%2C31](https://scholar.google.com/scholar?as_q=Indonesian%2Btweets%2Bhate%2Bspeech%2Btarget%2Bclassification%2Busing%2Bmachine%2Blearning&as_occt=title&hl=en&as_sdt=0%2C31)

1. A. R. Lubis, M. K. M. Nasution, O. S. Sitompul and E. M. Zamzami, "Spelling Checking with Deep Learning Model in Analysis of Tweet Data for Word Classification Process", *2022 9th International Conference on Electrical Engineering Computer Science and Informatics (EECSI)*, pp. 343-348, 2022.

[https://scholar.google.com/scholar?as\_q=Spelling+Checking+with+Deep+Learning+Model+in+Analysis](https://scholar.google.com/scholar?as_q=Spelling%2BChecking%2Bwith%2BDeep%2BLearning%2BModel%2Bin%2BAnalysis%2Bof%2BTweet%2BData%2Bfor%2BWord%2BClassification%2BProcess&as_occt=title&hl=en&as_sdt=0%2C31)

[+of+Tweet+Data+for+Word+Classification+Process&as\_occt=title&hl=en&as\_sdt=0%2C31](https://scholar.google.com/scholar?as_q=Spelling%2BChecking%2Bwith%2BDeep%2BLearning%2BModel%2Bin%2BAnalysis%2Bof%2BTweet%2BData%2Bfor%2BWord%2BClassification%2BProcess&as_occt=title&hl=en&as_sdt=0%2C31)

1. A. R. Lubis, M. K. M. Nasution, O. S. Sitompul and E. M. Zamzami, "A Framework of Utilizing Big Data of Social Media to Find Out the Habits of Users Using Keyword", *Proceedings of the 8th International Conference on Computer and Communications Management*, pp. 140-144,

Jul. 2020. <https://dl.acm.org/doi/10.1145/3411174.3411195>

1. A. R. Lubis, S. Prayudani, M. Lubis and Al-Khowarizmi, "Analysis of the Markov Chain Approach to Detect Blood Sugar Level", *J. Phys. Conf. Ser.*, vol. 1361, no. 1, 2019. <https://iopscience.iop.org/article/10.1088/1742-6596/1361/1/012052>
2. S. A. El Rahman, F. A. AIOtaibi and W. A. AlShehri, "Sentiment analysis of twitter data", *2019 international conference on computer and information sciences (ICCIS)*, pp. 1-4, 2019.

[https://scholar.google.com/scholar?as\_q=Sentiment+analysis+of+twitter+data&as\_occt=title&hl=en&as](https://scholar.google.com/scholar?as_q=Sentiment%2Banalysis%2Bof%2Btwitter%2Bdata&as_occt=title&hl=en&as_sdt=0%2C31)

[\_sdt=0%2C31](https://scholar.google.com/scholar?as_q=Sentiment%2Banalysis%2Bof%2Btwitter%2Bdata&as_occt=title&hl=en&as_sdt=0%2C31)