Mental Illness Detection Using Natural Language Processing and Machine Learning

Ruhit Ahmed Rizon

Applied Research Project submitted in partial fulfillment of the requirement for the degree of

M. Sc. In Artificial Intelligence

At Dublin Business School

Supervisor: Satya Prakash

December 2023

Declaration

"I hereby declare that, except for instances where it is obviously acknowledged by references, this Applied Research Project, which I have submitted to Dublin Business School for the award of M.Sc. in Artificial Intelligence, is the result of my own investigations. Moreover, no other degree has received this material as a submission."

Signed: Ruhit Ahmed Rizon

Student Number: 10640049

Date: 08-01-2023

Acknowledgement

I want to sincerely thank Satya Prakash for his guidance and unwavering watchfulness as I completed my master's thesis. Additionally, I am grateful to Professor Assem Abdelhak for helping me choose a research topic in the first place. Additionally, I would like to thank my loved ones for their encouragement and helpful feedback on the research.

Abstract

Mental illness nowadays has a huge impact on public health that encourages new methods of early identification and intervention. This dissertation will explore the bridge between Natural Language Processing and Machine Learning methods. The dataset is taken from Zenodo that has 104 files and one million rows and reflects different mental health conditions. Altogether total rows were one million and one hundred thousand is taken randomly from them where each file contributes ten percent of the data. Preprocessing techniques were applied to improve the quality of the train set such as stopword removal, lemmatization, punctuation removal, special character and number removal, and data merging. For model training different feature engineering techniques were used such as TF-IDF, Min-Max scaling, Standard scaling and Log scaling. On the other hand, four different classifiers were used to evaluate the effectiveness of predicting mental diseases from text. They are Multinomial Naive Bayes, Logistic Regression, Random Forest and Gradient Boosting. Grid search and Random search were also be used to investigate the difference between the results of Logistic Regression and Multinomial Naive Bayes. To evaluate the performance of each model different techniques were used like accuracy, weighted precision, F1 scores, and confusion matrices. Logistic Regression was the best model which was min-max scaled.

Table of Contents

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of Figures

Chapter 1

Introduction

Mental health is the fundamental basis for overall well-being, affecting all elements of the human experience. As societies advance and our understanding of mental health evolves, we increasingly recognize its substantial impact on individuals, families, and communities. Mental health disorders have become widespread globally, transcending national, cultural, and economic boundaries. The World Health Organization (WHO)[10] emphasizes the need to tackle mental health issues worldwide, since they estimate that approximately 25% of individuals may experience mental or neurological disorders at several stages in their lives. The evolution of therapeutic methodologies and the cultivation of mindsets represent the defining characteristics of the intricate narrative that encompasses the historical progression of mental healthcare. In the 18th century psychiatric asylums were introduced. Before that it used to treat with supernatural beliefs that were beyond any scientific logic. But the practice of threating mental illness through scientific ways started at mid-20th century when different effective therapy and diagnosis were introduced, and they proved to be effective curing mental illnesses. Still there were some lacings left to make it fully accurate. Social stigma around mental illnesses still made it harder to seek help and care timely. Traditional diagnostic procedures often suffer from different limitations as they mostly rely on subjective assessments and self-reported symptoms.

The modern era has brought opportunities as well as challenges for mental health. Nowadays social media gives people an opportunity to share mental health experiences and also emotional wellbeing. However, things like cyberbullying and social media-induced anxiety have added a few extra problems to peoples' mental health. For this Machine Learning (ML), Natural Language Processing (NLP) comes with a lot of opportunities. NLP gives machines the leverage to understand human language whereas ML allows machines to predict something from those languages. For this now machines are capable of understanding large number of text data and ML identifies the patterns of them. However, we must consider the social context and ethical use of these tools. This research explores the different opportunities of tech innovations, social issues and history that can medication and identification of mental illnesses through ML and NLP. The prior aim is to facilitate medical professionals and patients or normal people’s early identification and fast cure.

The main difficulty is that existing diagnostic methods are not capable enough to handle the change of spectrum of different conditions of mental health. Self-reporting and time-consuming diagnosis are common limitations and mass people struggle for this. However, making prompt and accurate diagnosis is more difficult as there are a lot of different mental health issues and also lack of communication makes them harder. These models can decrease these problems and opens up a new window that allows better care.

There is a huge potential of this research that can give a new perspective which can be significant. The project intends to provide effective, impartial and timely insights through NLP and ML. Furthermore, proactive and data-driven interventions have the power to lower the stigma that refers to mental health issues which is very important to improve diagnostic accuracy. The consequences of the findings would help a huge number of patients across the world. It can also decrease time, save money and give better treatment by which lot will be benefited and also will change the global view of mental health.

This research uses cutting-edge NLP and ML approaches that create predictive models to decrease the gaps in mental health diagnosis. The main interest of these models are, they can identify different mental health conditions after they are trained with different textual datasets. The ultimate goal is to deliver fast, reliable and accurate results for mental health treatment. The project aims to give a new dimension that will decrease mental health advocacy and also increase mental well-beingness.

As a part of the research, a randomly chunked data was selected from 104 files that has altogether one million rows of text data. Each file provides 10% of the data to make the dataset more balanced. File merging, text preprocessing and for feature extraction TF-IDF were used to preprocess the data that makes the data more noise free, easily processable by machines. For preprocessing methods like lemmatization, stopword removal, number removal and lower case is applied. Different scaling methods also applied such as Min-Max, Log, and Standard scaling. For classification four algorithms were used: Multinomial Naive Bayes, Logistic Regression, Random Forest, and Gradient Boosting. On the other hand, hyperparameter tuning techniques were used to observe the results.

As our dataset is not balanced that is found from data exploration, we focused on F1 scores. Logistic regression has the highest F1 score around 71%. However Gradient boost algorithm and Random Forest have almost similar F1 score around 65%. But the least performing model is Naive Bayes. Confusion matrices, F1 scores, and weighted precision were among the evaluation measures used to give a detailed picture of the models' performance. This study lays the groundwork for further research in the field of mental health detection by highlighting how NLP and ML have the power to revolutionize diagnostic procedures and enhance patient outcomes.

Research Question

Can algorithms for machine learning and natural language processing accurately forecast mental health illnesses from textual data, supporting early detection and intervention strategies?

Research Objective

In order to improve the efficiency and accuracy of mental health disorder detection through the analysis of textual data, this research aims to develop and evaluate predictive models using Natural Language Processing and Machine Learning techniques. Ultimately, this research will contribute to early intervention strategies.

Chapter 2

Literature Review

Low et al[1] collected the data from reddit, a very popular website that has mental support group subfora(subreddit) which captures instant experiences regarding pandemic. Text processing and machine learning is being used on the dataset to analyze the COVID-19 effects on people mental health and their needs. Quarantine showed the increase of different mental disorder in people mind such as Post-Traumatic Stress Disorder (PTSD), Major Depressive Disorder, and Generalized Anxiety Disorder [2]. Early data of China also be analyzed that shows depressive symptoms (50.4%), anxious symptoms (44.6%), insomnia (34%) and general distress (71.5%) [3] comparatively high in healthcare workers. Adults that are younger with pre-existing mental and physical health conditions point out worsening symptoms that are depression, anxiety [4]. In terms of using Natural Language Processing for mental health is a booming field nowadays and posts in subreddit has unique topic. Unlike standard clinical datasets, Reddit's anonymous, free-form posts offer an ecological documentation of sensitive first-hand experiences, and its publicly available, historical data allows for comparisons. Machine learning models have been used to classify and characterize Reddit posts from mental health subreddits, achieving 98% accuracy using N-gram language modeling and LIWC [5]. Although demographic data for individual subreddits is not accessible, Reddit users as a whole are primarily young (22% 18-29yo; 14% 30-49yo), male (67%), and American (49.9%).

The study extracted data from 15 mental health communities, two broad mental health subreddits, and 11 non-mental health subreddits using the pushshift API, focusing on specific issues and preprocessing details. Posts were analyzed for features like LIWC, sentiment analysis, basic word counts, punctuation, readability metrics, TF-IDF ngrams, and manually built lexicons related to suicide, economic stress, isolation, substance use, domestic stress, and guns. Using an 80-20 train-test split, they used binary classification on 15 mental health subreddits in comparison to a control group. In order to assess possible dataset, shift and performance using weighted F1, two more test sets, comprising mid-pandemic posts and r/COVID19_support posts, were constructed to test the pre-pandemic model. Three linear models and two complicated tree ensemble classifiers were tested in the study with the goal of identifying feature importance using the least complex model. Data from verified instances and COVID-19-related tokens was collected between January 2020 and April 2020. For every 90 characteristics and 28 subreddits, a linear regression was fitted, and the slope × R 2 represented the rate of change. A Mann-Whitney U test was used to assess the absolute difference between years. A comparison analysis of 15 mental health subreddits during the COVID-19 outbreak was done. The feature set was condensed to 30 PCA components and the post-pandemic posts were downsampled to a balanced representation of 1500 posts per subreddit. Twenty clusters were found using the scikit-learn SpectralClustering function, and cluster-characteristic characteristics were found using Wilcoxon rank-sum tests. LDA model estimation was carried out using the gensim library, and several models were developed to evaluate topic stability. In order to compare the similarities between the subreddits, the study also calculated the asymmetric Hausdorff distance across time. We extracted the median distance value from fifty bootstrapping samples. The purpose of the study was to ascertain how the pandemic affected subreddits related to mental health. In order to identify psychologically relevant COVID-19 posts on Reddit, a study used binary classification algorithms using data collected in the middle of the pandemic. As compared to the motion-related lexicon, the results indicated a considerable drop in the use of tokens associated with isolation, economic stress, home, and anxiety. There is a link of ρ = 0.83 between the global COVID-19 instances and the mean proportion of posts connected to the virus. Post-traumatic stress disorder, borderline personality disorder, EDAnonymous, and r/alcoholism were the mental health subreddits with the greatest negative semantic change. The language use across subreddits is a continuum, but it also includes significant axes of variation, as indicated by the proximity of clusters for closely related conversation themes, according to a clustering analysis of pre-pandemic posts. Using both supervised and unsupervised clustering, the study looks at the post distribution in mid-pandemic mental health subreddits. Topics including alcoholism and addiction, health anxiety, eating disorders and alcoholism, schizophrenia, ADHD, and autism were mostly represented in the pre-pandemic LDA model. The tokens for autism and ADHD are divided into the mid-pandemic LDA model, which also includes a family theme and a PTSD token. The "Health Anxiety" and "Life" subjects are on the rise, while the "Alcoholism and Addiction" topic is on the decline, according to Wilcoxon signed-rank tests on distributions of pre- and mid-pandemic postings across pre-pandemic model themes. During the pre-pandemic and mid-pandemic periods, the prevalence of the topics of health anxiety and life greatly increased, whereas the popularity of the issue of alcohol and addiction significantly declined. Common themes and pain spots are indicated by the topic distribution, which may assist determine the format and content of mental health services. According to the study, which focuses on mental health issues among Reddit users, linguistic alterations during the epidemic are typical. But the study doesn't connect adjustments to occurrences. The study focuses on the increased clinical needs to offer reddit users useful tools. The research analyzes posts from other subreddits like r/COVID19_support that could benefit the users along with r/Suicide watch. The results show that Natural Language Processing (NLP) could aid the users by moderating answers. Moreover, the study refers to more studies in future that allows to understand the ability of some health groups to decrease the effects of pandemic.

Le Glaz et al. (2021) [6] did thorough research using ML and NLP which was published between August 2018 and February 2020. The search was PRISMA-guided, that summed up 58 relevant papers from 327 original entries. Analysis employing EHRs, social media, and clinical notes showed that studies on depression and suicide risk were conducted often. Getting rid of symptoms, grading the severity of the illness, evaluating treatment, and identifying psychopathological hints were common goals. Standard NLP was combined with medical ontology mapping during preprocessing. High-performing classifiers such as SVM and regression were used over transparent techniques in the majority of investigations. The most popular ones were R and Python. The authors draw the conclusion that while ML and NLP offer novel research paradigms, they currently mostly validate clinical intuitions. They stress taking ethical considerations and cultural biases into account.

The exponential expansion of user-generated content on social media platforms has sparked interest in using these data sources for stress detection and mental health monitoring. For this purpose, Nijhawan et al. [7] offer a thorough overview of sentiment analysis and emotion detection methods used with Twitter data. Using deep learning models like BERT and machine learning algorithms like random forest, their methodology includes gathering, cleaning, extracting features from the data, and modeling it. An exploratory analysis reveals that the sample tweets are mostly about student life and neutral/sad emotions. Underlying themes about sadness and college are revealed through topic modeling with LDA. Experiments on binary sentiment classification reveal that random forest outperforms decision trees and logistic regression, achieving 97.78% accuracy. In terms of multi-class emotion classification—joy, sadness, fear, anger, and neutrality—the refined BERT model achieves 94% accuracy. The real-time text input capabilities of the framework are demonstrated through a web portal. Though ethical considerations must be made, the authors highlight how social media analytics using natural language processing and machine learning offers novel opportunities to detect signals of psychological distress for proactive mental healthcare. In a related study, Zhang et al. thoroughly [9] examine 399 papers from 2012 to 2021 that use NLP methods to identify mental illness. They reveal an upward trend in this application's research, favoring deep learning over more conventional machine learning techniques that rely on human feature engineering. Publication analysis reveals that social media data—particularly Twitter—is widely used to cover a range of mental health issues. Models like CNN, RNN, and Transformer architectures are frequently used. The model interpretability, dataset bias, and ethical guidelines for using personal data are some of the major issues that are brought up. Explainable AI, multi-modal approaches, semi-supervised learning, and stringent consent protocols are the main points of advice. When combined, these papers offer thorough summaries of recent developments using natural language processing (NLP) and machine learning to leverage online textual data for mental health monitoring and psychiatric risk assessment. They also draw attention to significant shortcomings in terms of privacy, transparency, and robustness, which call for continued study.

A thorough analysis of how natural language processing (NLP) methods have been used to various user-generated texts in order to draw conclusions about mental health and wellbeing can be found in the paper written by Calvo et al. [8] The writers list the various data sources that were consulted, including blogs and social media sites like Facebook, Twitter, and Reddit. Numerous studies have concentrated on identifying particular moods, emotions, depression, risk factors for suicide, and other mental health conditions from the language used in these texts. The methods comprised taking different language, behavioral, social, and demographic characteristics out of the texts and applying machine learning techniques for classification and prediction, such as support vector machines, naive Bayes, conditional random fields, and deep learning. Identifying distressed individuals, classifying texts pertaining to mental health issues, and combining texts to describe trends at the population level have been important tasks. Self-reports, participation in mental health forums, physician diagnoses, and manual annotation have all provided ground truth data. The majority of work uses lexical databases, such as WordNet, ANEW, and LIWC, along with topic models and metadata, such as time and location, to detect emotions. Beyond diagnosis, some systems have dabbled in creating personalized interventions through natural language generation, such as individualized information delivery, support request prioritization, or human-robot interaction. However, there are still ethical issues with identifying at-risk people without their consent, as well as technical difficulties with processing informal text. The review, taken as a whole, covers a wide range of multidisciplinary literature on the use of NLP in mental health applications. It offers a useful framework and common language for integrating viewpoints from computational linguistics, human-computer interaction, and mental health research. In order to provide a solid foundation for future research on applying NLP techniques to understand mental health from everyday texts, the review highlights major trends and gaps.

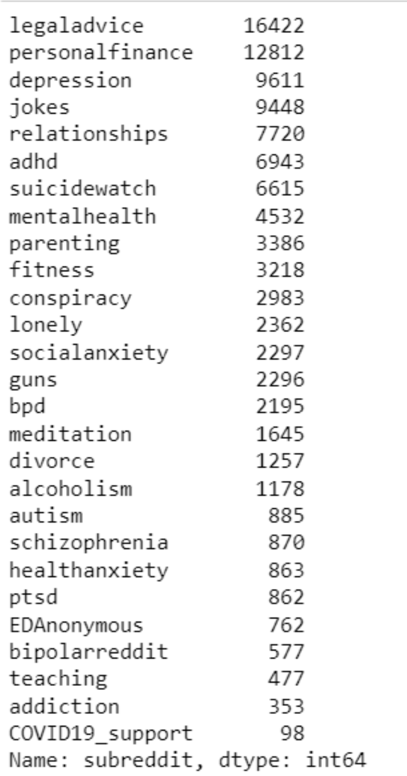
Chapter 3

Methodology

These days, mental sickness is a big problem that affects a lot of people. We can now find them earlier and faster than ever before thanks to progress in technology. Traditional methods were used by doctors in the past to find illnesses, and they took a long time. Some people said it got worse. So, if we can come up with a way to find people who might have mental illnesses, we will save a lot of time and give doctors more time to think about how to fix them. We started our study because we were interested in whether or not algorithms for natural language processing and machine learning could correctly talk about mental diseases from written data to help with early diagnosis.

Data Collection

The data[11] was obtained from Zenodo and consists of 104 files containing Reddit postings from 27 distinct subreddits. The data originated from a compilation of 15 mental health support organizations, documenting the years 2018-2020. There are a total of 104 distinct files. Initially, we employed the pandas library in Python to merge the files into a single entity. It would facilitate the testing process if we randomly selected around one hundred thousand data points from the larger dataset. Each of the files contains 10% of the data in that one hundred thousand data, ensuring a rather fair distribution across the collection.



```
legaladvice          16422
personalfinance      12812
depression            9611
jokes                 9448
relationships         7720
adhd                  6943
suicidewatch          6615
mentalhealth          4532
parenting             3386
fitness               3218
conspiracy            2983
lonely                2362
socialanxiety         2297
guns                  2296
bpd                   2195
meditation            1645
divorce               1257
alcoholism            1178
autism                 885
schizophrenia          870
healthanxiety          863
ptsd                   862
EDAnonymous            762
bipolarreddit          577
teaching               477
addiction              353
COVID19_support         98
Name: subreddit, dtype: int64
```

Figure 1 data distribution of 102667 data

Feature Extraction

For feature extraction and dataset exploration we applied Principal Component Analysis (PCA). For dimensionality reduction to numerical features PCA is applied. After that in the numerical features Term Frequency-Inverse Document Frequency (TF-IDF) is applied that reflects the significance of the words in a document relative to their occurrence across all documents. Then t-Distributed Stochastic Neighbor Embedding(t-SNE) is applied to visualize all the 27 different classes in different colors in three dimensions.

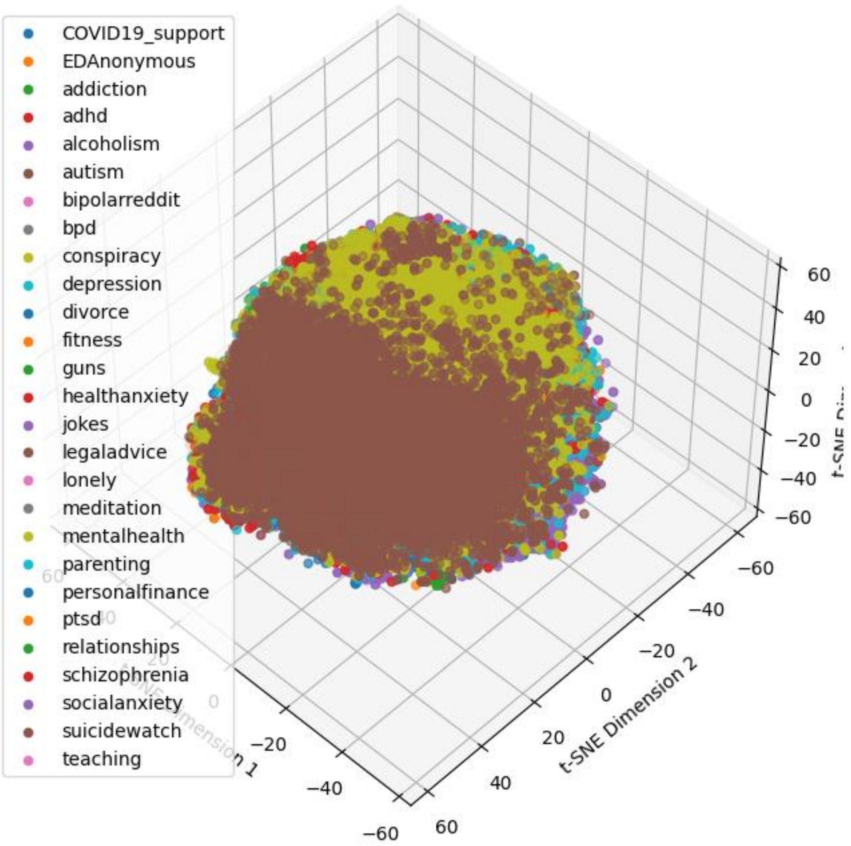


Figure 2 t-SNE representation of all the classes in different colors

Uniform Manifold Approximation and Projection (UMAP) is also applied for the reduction of additional dimensionality that proves lower dimensional representation of the numerical features. Here the sample size is exactly one hundred thousand. As two components for UMAP were only chosen, the dimensionality would be UMAP1 vs UMAP2.

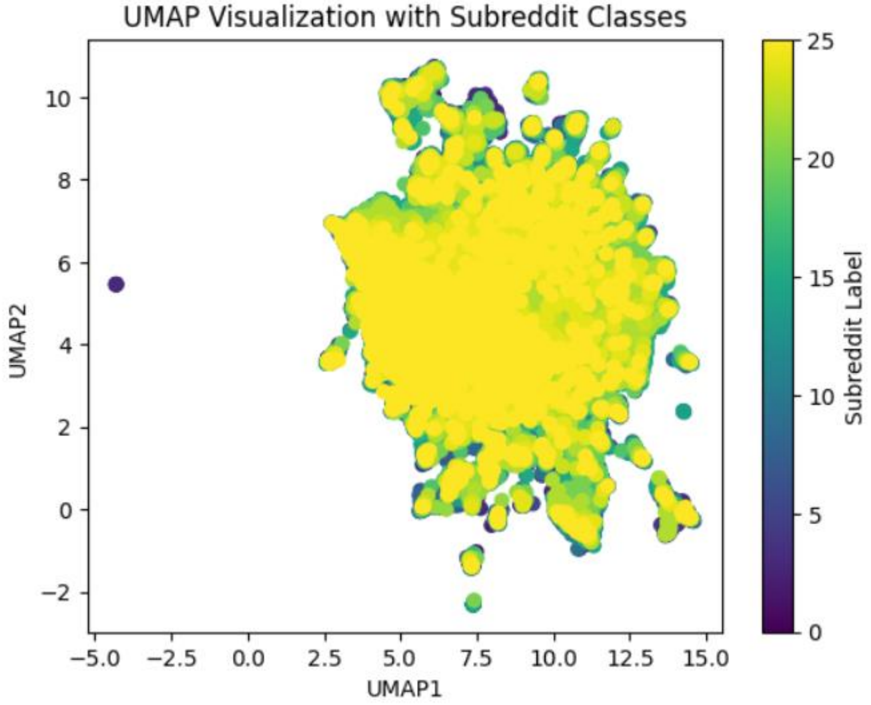


Figure 3 visualization of UMAP data into two dimensions.

Word cloud is also applied just to show the top words used in the modified dataset. The size of the sample is one hundred thousand.

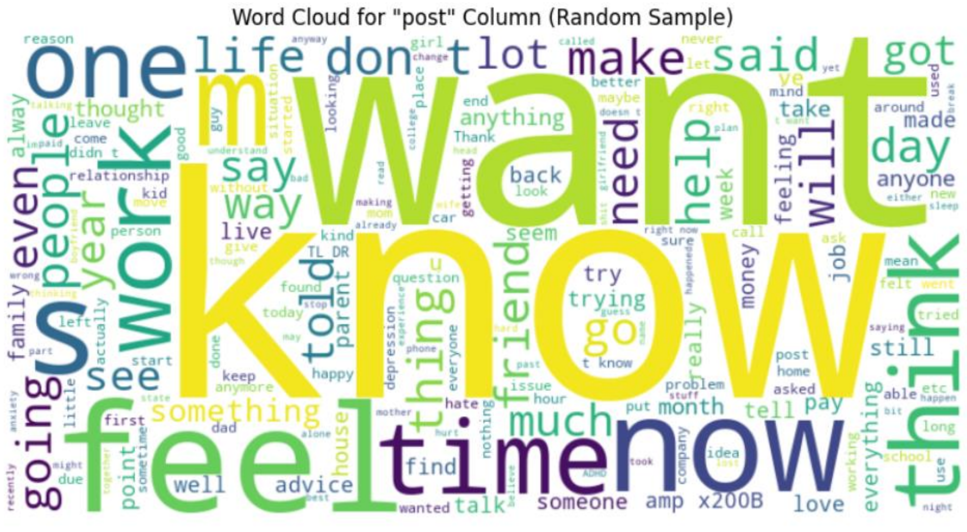


Figure 4 visualization of top words in words cloud.

TF-IDF is also visualized in a bar chart where top words are shown in terms of the TF-IDF score. Only the top twenty are shown in the bar chart.

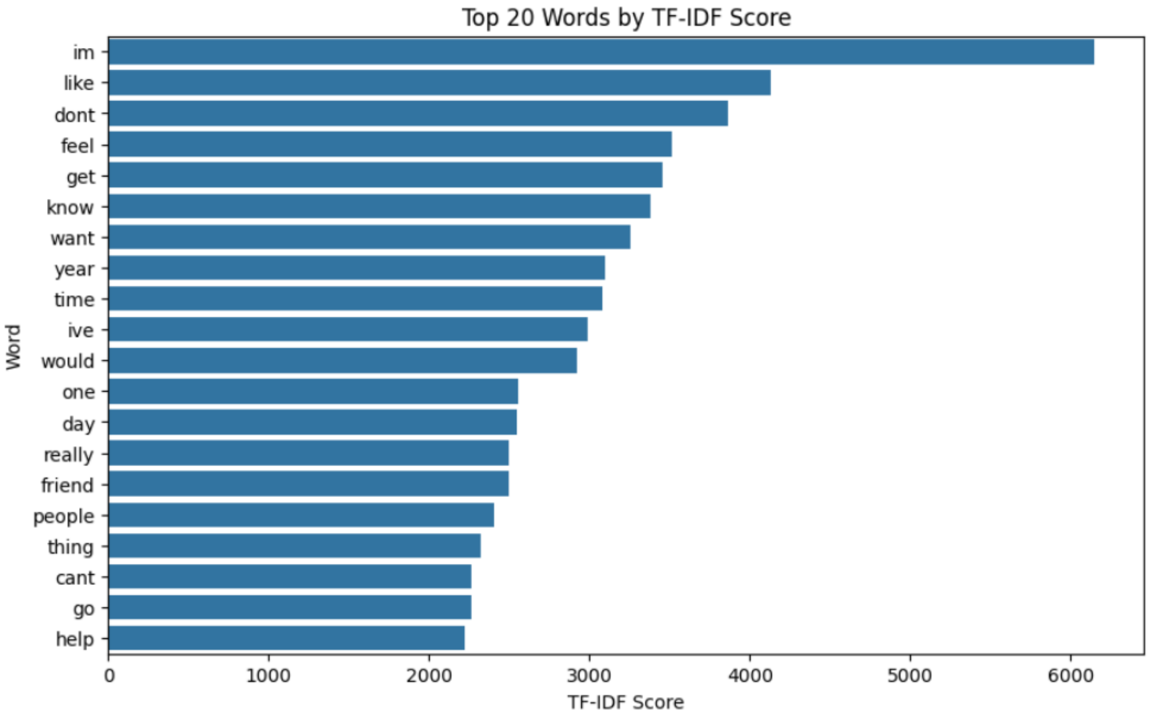


Figure 5 representation of the top twenty words by TF-IDF Score

Data Preprocessing

Textual data underwent multiple preprocessing processes before analysis. All content was changed to lowercase to equalize capitalization. Next, numeric digits were deleted because they didn't include analysis-relevant data. Punctuation and other non-alphanumeric symbols were removed to minimize noise and normalize punctuation-only texts. Avoiding stopwords like "a", "and", and "the" let the analysis focus on more important phrases. Lemmatization mapped terms to their base forms, treating inflections of the same root word equally.

Final TF-IDF weights were determined for each period. TF-IDF shows how essential a word is to a document across all documents. More significant phrases had more effect in the analysis with TF-IDF weights. The textual data was cleaned and normalized after lowercasing, digit removal, symbol stripping, stopword removal, lemmatization, and TF-IDF weighting to decrease noise and variability for better analysis.

Feature Scaling

Based on the data, we utilized several feature scaling strategies to normalize the range of independent variables in order to facilitate efficient modeling. A logarithmic scale was applied to the features to decrease skewness and limit the influence of outliers. This was accomplished using log scaling. With the use of standard scaling, the characteristics were standardized such that they had a mean of zero and a standard deviation of one. This was done to bring them into similar ranges. The data was then rescaled to this constrained range after the min-max scaling technique was applied, which turned the characteristics into a defined range of 0 to 1. The characteristics of the data were standardized by applying different scaling techniques that include log scaling, standard scaling, min-max scaling. Outliers did not have enough influence when log scaler was applied. Min-max scaling confined all the characteristics to the range of 0 to 1. On the other hand, all the variables were standardized by standard scaling. The preprocessing was done to make the machine learning model more accurate and scaled data helped to increase the accuracy of the models.

Chapter 4

Data Splitting

During the research, four different machine learning algorithms were used to analyze the results. The dataset was split into 80:20 ratio where the training dataset was 80 and testing data set is 20. The training dataset was used to train the algorithms and the test dataset was there to evaluate the performance of the models. The models that were used in the research are: Multinomial Naive Bayes, Logistic regression (LR), Random Forest(RF) an Gradient Boosting Algorithm(GB). Jupyter notebook was being used to perform the training.

Training and Testing

Multinomial Naive Bayes (NB) Model Training

Multinomial Naive Bayes were used to train the preprocessed dataset. The method is based on Bayes' theorem that has a huge capability to assume characteristics that are independent from the other ones. During the training process, model was fitted into labeled data where numerical and linguistic representation of the posts on reddit were corresponded. Data scaling methods were also applied to check the effectiveness of the model. Lastly hyperparameter tuning methods were tested to Multinomial Naive Bayes to enhance the prediction of the model.

Logistic Regression (LR) Model Training

Logistic Regression model was trained on the training dataset that was preprocessed. The Logistic Regression technique is a linear classification method that provides an estimate of the likelihood of belonging to a certain category. In the process of training, the features were extracted from both numerical and textual representations of the Reddit posts. In the training session, the best possible weights were finalized. The model is being adjusted at tie time of training process to get the best possible performance.

Random Forest (RF) Model Training

Random Forest model was trained with the training dataset. Moreover, this ensemble learning system uses a lot of decision trees to improve its accuracy and predictions. In the training session a huge number of decision trees were generated by the model. These trees were derived from various subsets of the training data. Optimization was performed on the hyperparameters, which included the number of trees and the tree depth, reaching the highest feasible level of performance.

Gradient Boosting (GB) Model Training

On the training dataset, the Gradient Boosting model was trained to perform its functions. Gradient Boosting is a technique that is used in ensembles to construct decision trees in a sequential manner, with each tree fixing the faults that were made by the tree that came before it. During its training, the model demonstrated an iterative improvement in its predictive powers by reducing the residuals. Several hyperparameters, such as the learning rate and tree depth, were adjusted to perfection in order to achieve the best possible overall performance from the Gradient Boosting model.

Model Prediction and Evaluation

Prediction

For prediction on the test data, we have used trained Naïve Bayes, Logistic Regression, Random Forrest and Gradient Boosting classifier. We also applied these algorithms to min-max and standard scaled data just to compare which one performs better.

Evaluation Metrics

We calculated accuracy, precision, recall, F1 score and confusion matrix to evaluate the models. Overall accuracy is computed by comparing the actual labeled data with the trained data. Then precision, recall and F1 score is computed binary classification. On the other hand, detailed confusion matrix and confusion matrix heatmap is implemented to break down true positives, true negatives, false positives, and false negatives.

Scaling Techniques Impact

Effect of Min-max scaling and standard scaling on the model performance is also analyzed. This shows whether scaling boosts the performance of the classifiers or not.

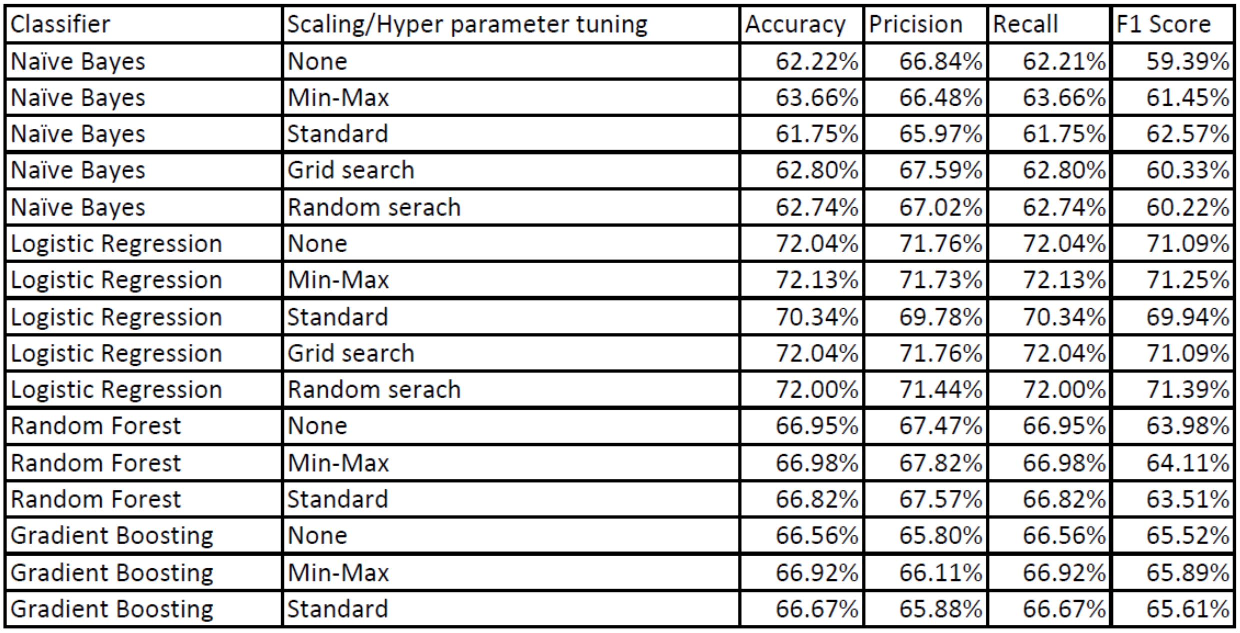
Hyperparameter Tuning Impact

Grid search and Random search are also applied in the models. This allows us to compare the performance of optimized hyperparameters versus the default configuration.

Chapter 5

Results

The study investigates several methods and employs diverse scaling and Hyperparameter tweaking strategies. This will enable us to determine whether we can address the research topic through this study. This work examines the intricate relationship between the performance of classifiers and the features of datasets, with a special focus on the F1 score as a reliable indicator when dealing with unbalanced data. We will be examining the findings of the research.



Classifier	Scaling/Hyper parameter tuning	Accuracy	Pricision	Recall	F1 Score
Naïve Bayes	None	62.22%	66.84%	62.21%	59.39%
Naïve Bayes	Min-Max	63.66%	66.48%	63.66%	61.45%
Naïve Bayes	Standard	61.75%	65.97%	61.75%	62.57%
Naïve Bayes	Grid search	62.80%	67.59%	62.80%	60.33%
Naïve Bayes	Random serach	62.74%	67.02%	62.74%	60.22%
Logistic Regression	None	72.04%	71.76%	72.04%	71.09%
Logistic Regression	Min-Max	72.13%	71.73%	72.13%	71.25%
Logistic Regression	Standard	70.34%	69.78%	70.34%	69.94%
Logistic Regression	Grid search	72.04%	71.76%	72.04%	71.09%
Logistic Regression	Random serach	72.00%	71.44%	72.00%	71.39%
Random Forest	None	66.95%	67.47%	66.95%	63.98%
Random Forest	Min-Max	66.98%	67.82%	66.98%	64.11%
Random Forest	Standard	66.82%	67.57%	66.82%	63.51%
Gradient Boosting	None	66.56%	65.80%	66.56%	65.52%
Gradient Boosting	Min-Max	66.92%	66.11%	66.92%	65.89%
Gradient Boosting	Standard	66.67%	65.88%	66.67%	65.61%

Figure 6 all the results of different classifiers with scaled and hyperparameter tuned.

Figure 6 illustrates the outcomes of several classifiers. The graphic displays the outcome of the base classifier, as well as the scaled classifiers and hyperparameter tuned classifiers. Accuracy is the predominant outcome in other investigations. However, because of the imbalanced nature of our dataset, our study will prioritize the F1 score. This metric takes into account both recall and accuracy, resulting in a fairer evaluation. According to the figure, logistic regression (LR) with hyperparameter tuned achieved the highest accuracy of 71.39%. Conversely, the Naïve Bayes (NB) basic model has a poor performance rate of 59.39%. Other Naïve Bayes models that have been scaled and hyperparameter tuned likewise have relatively poor performance compared to other classifiers. The second model utilized is the Gradient Boosting approach, which has been scaled using the min-max scaling technique and achieved an accuracy of 65.89%. The base Gradient Boosting method and the standard scaled Gradient Boosting approach yield comparable results, with accuracies of 65.52% and 65.61% respectively. Lastly, the Random Forest algorithm has moderate performance, with an accuracy rate of around 63% when compared to other algorithms.

Challenges

A major obstacle faced throughout this research was related to the availability of computer resources and the pace at which processing could be done. Examining a dataset consisting of various mental health subjects gathered from Reddit required significant computational capacity, and the constraints of the existing hardware became evident. As the dataset was huge, several tasks took a long time to complete. Especially training the datasets in different models, extracting the features. The reason behind this is natural language processing and machine learning algorithms are very complicated and needs a lot of computational power. That is why it slowed down the research a lot for the long processing time.

Future Work

In future study, the main focus would be decreasing the imbalance nature of the dataset with various techniques such as over sampling, under sampling, random over sampling and random under sampling. The main goal is to contribute fairer distribution of data of mental illness for training and testing of the model. On the other hand, several classifiers will be tested to check their effectiveness for mental illness detection such as Support Vector Machines (SVM) and Bayesian Additive Regression Trees (BART). Min-max scaling and standard scaling will also be applied in Random Forest and Gradient Boosting models to compare with others. In addition, deep learning methods as well as neural network methods will also be explored and will conduct different experiments and evaluate their effectiveness. If better results found in those experiments, a web application will be developed that will take data from the users and predict mental illness and also advice some of the medication from a ethical point of view.

Conclusion

As early detection is crucial for mental illnesses, machine learning and natural language processing algorithms were used to predict mental illness from text data. The main goal of the study was to determine if these methods can accurately predict mental illnesses just using the text data given by an individual. Though we have unbalanced data Logistic Regression showed good performance that has an F1 score of 71.39%. This proves that NLP methods can effectively detect sickness from text data. The outcomes could be better if the data was balanced, and it could give very high scores. The device used to perform the experiment was not strong enough. So, with better device it would be faster and more efficient. Moreover, The study is a prior example that Machine Learning and Natural Language Processing are capable of predicting mental health from text alone that may help future endeavors. The ethics of exploiting people's linguistic data for predictive modeling of sensitive health issues also need to be carefully considered. In order to improve models while guaranteeing proper usage, future work will concentrate on balancing the dataset, adding more classifiers like SVM and BART, investigating scaling strategies, and experimenting with neural networks. The objective is to prioritize model fairness, accountability, and openness while improving prediction performance and accessibility through an intuitive application.

References

1. Low, D.M., Rumker, L., Talkar, T., Torous, J., Cecchi, G. and Ghosh, S.S., 2020. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during covid-19: Observational study. Journal of medical Internet research, 22(10), p.e22635.
2. Brooks, S.K., Webster, R.K., Smith, L.E., Woodland, L., Wessely, S., Greenberg, N. and Rubin, G.J., 2020. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The lancet, 395(10227), pp.912-920.
3. Duan, L. and Zhu, G., 2020. Psychological interventions for people affected by the COVID-19 epidemic. The lancet psychiatry, 7(4), pp.300-302.
4. Alonzi, S., La Torre, A. and Silverstein, M.W., 2020. The psychological impact of preexisting mental and physical health conditions during the COVID-19 pandemic. Psychological trauma: theory, research, practice, and policy, 12(S1), p.S236.
5. Shen, J.H. and Rudzicz, F., 2017, August. Detecting anxiety through reddit. In Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality (pp. 58-65).
6. Le Glaz, A., Haralambous, Y., Kim-Dufor, D.H., Lenca, P., Billot, R., Ryan, T.C., Marsh, J., Devylder, J., Walter, M., Berrouiguet, S. and Lemey, C., 2021. Machine learning and natural language processing in mental health: systematic review. Journal of Medical Internet Research, 23(5), p.e15708.
7. Nijhawan, T., Attigeri, G. and Ananthakrishna, T., 2022. Stress detection using natural language processing and machine learning over social interactions. Journal of Big Data, 9(1), pp.1-24.
8. Calvo, R.A., Milne, D.N., Hussain, M.S. and Christensen, H., 2017. Natural language processing in mental health applications using non-clinical texts. Natural Language Engineering, 23(5), pp.649-685.
9. Zhang, T., Schoene, A.M., Ji, S. and Ananiadou, S., 2022. Natural language processing applied to mental illness detection: a narrative review. NPJ digital medicine, 5(1), p.46.
10. World Health Organization (2001) 'The World Health Report 2001 - Mental disorders affect one in four people', WHO, Available at: https://www.who.int/news-room/detail/28-09-2001-the-world-health-report-2001-mental-disorders-affect-one-in-four-people (Accessed: 15 December 2023)
11. Zenodo. (2020). Reddit Mental Health Dataset. Zenodo. https://doi.org/10.5281/zenodo.3941387