Proceedings of the 2nd International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/132/2024.20541

A Machine Learning-Enhanced Chat Application for the Identification of Mental Disorders

Qian Cao[1,5,*], Zhihao Yan[2,6], Zihang Gong[3,7], Jin Huang[4,8]

[1]School of Computer Science, Wuhan University, Wuhan, 430072, China
[2]Bishop Guertin High School, Nashua, NH 03060, USA
[3]Magee Secondary School, Vancouver, BC, V6M 4M2, Canada
[4]Westridge School for Girls, Pasadena, CA 91105, USA

[5]qianc7611@gmail.com
[6]yzh20070227@gmail.com
[7]leogong778@gmail.com,
[8]jessicahuang9122@gmail.com
*corresponding author

Abstract. The prevalence of mental disorders is increasing, but they continue to be underdiagnosed and under addressed. Social media platforms offer novel opportunities for detecting potential mental health issues through the analysis of user-generated content. This paper presents a chat-based program developed using machine learning models trained on a dataset of comments from Reddit users. The program is capable of predicting the type of mental illness based on user input. This study provides a detailed comparison of various classification algorithms, including Naïve Bayes, Logistic Regression (LR), Support Vector Machines (SVM), and Random Forests (RF). Additionally, the paper discusses relevant machine learning techniques from previous studies. The results indicate that LR model, particularly with a uni-gram feature representation, outperforms other models with an accuracy of 0.81 and demonstrates the fastest processing speed. Future research directions include the integration of Large Language Models and the development of a multilingual chat interface.

Keywords: Social media, mental disorder, machine learning, classification algorithm

1. Introduction

Mental illness, also referred to as a mental disorder, pertains to a condition that affects an individual’s cognitive functions, emotional regulation, behavior, and value judgments [1]. The manifestations of various mental disorders range from abrupt shifts in personality to enduring feelings of melancholy. Individuals may exhibit a variety of symptoms such as increased heart rate, nausea, panic attacks, sleep disturbances, and suicidal ideation [2]. Traditional methods of identifying mental patients rely on professional expertise. Due to the absence of an objective criterion or consistent physical symptoms for mental disorders [3], they can often be mistaken for unstable emotions, making the diagnosis of mental illnesses heavily dependent on the judgment of psychiatrists. Becoming a psychiatrist requires years of training and practical experience for many healthcare professionals. Consequently, self-diagnosis by patients or detection by their friends or family members is challenging. These circumstances underscore the pressing necessity for advanced technologies to identify potential mental health patients.

The widespread adoption of social media has led to an increasing number of individuals choosing to utilize social media platforms as their primary means of expressing emotions. Research suggests a correlation between users' self-assessment and their social media posts [4], and self-assessment plays a significant role in the potential causes of mental health disorders. By collecting a substantial dataset comprising comments from Reddit users, it is possible to conduct a preliminary assessment by creating a model based on the analysis of these comments, and potentially offer more detailed treatment options. The development of these models involves the application of supervised machine learning techniques. By evaluating the performance of various models, we identified the most effective one. Ultimately, the model is integrated into a chat program capable of providing basic assessments based on user input.

2. Dataset

The dataset was obtained from Hugging Face and comprises roughly 42,100 records of Reddit users' self-comments labeled with different mental sicknesses, including depression and Post-Traumatic Stress Disorder (PTSD).

The dataset comprises 42,113 columns and 2 columns. The columns are labeled'body' and 'label', where the 'label' column contains eight particular numerical values compared to eight diverse mental clutter categories. The objective of the show is to classify the substance within the 'body' column, coming about in a prediction that matches one of the eight numerical labels.

Hugging Face could be a stage that specializes in Normal Dialect Handling (NLP) and Counterfeit Insights (AI), advertising a wide cluster of instruments and assets advantageous for designers and analysts. The essential advertising is the Transformers library, which contains pre-trained models outlined for errands such as text classification, interpretation, and summarization. Moreover, the Datasets library gives get to pre-existing datasets for machine learning ventures, encouraging the procurement of expansive datasets for preparing and testing purposes. Hugging Face too highlights an Application Programming Interface (API) that streamlines show sending and prediction making through an API, as well as Spaces, a stage for creating and sharing intelligently machine learning applications. These comprehensive instruments and assets have built up Hugging Face as a key player in AI and NLP communities, fostering innovation and the selection of progressed arrangements.

Reddit may be a social stage where clients lock in talks on different themes such as news, legislative issues, courses, and maladies. The stage permits posts of up to 40,000 characters, empowering clients to supply nitty gritty portrayals of their encounters and feelings. Within the proposed dataset, the normal content length is 168.6 words. With an endless client base of hundreds of millions of month-to-month dynamic clients, Reddit cultivates an interesting community culture that advances the open sharing of individual stories, counting encounters with sicknesses. This openness is profitable for research purposes. Research conducted by De Choudhury demonstrates that Reddit clients share their encounters, feelings, and the impacts of mental illness on their day-by-day lives [5]. This characteristic renders Reddit a practical stage for distinguishing mental clutters. Past research has included gathering individual accounts from Reddit to set up a dataset for detecting uneasiness [6], whereas other considerations have utilized comments and posts from Reddit to compile a dataset for recognizing self-destructive eagerly [7].

3. Methodology

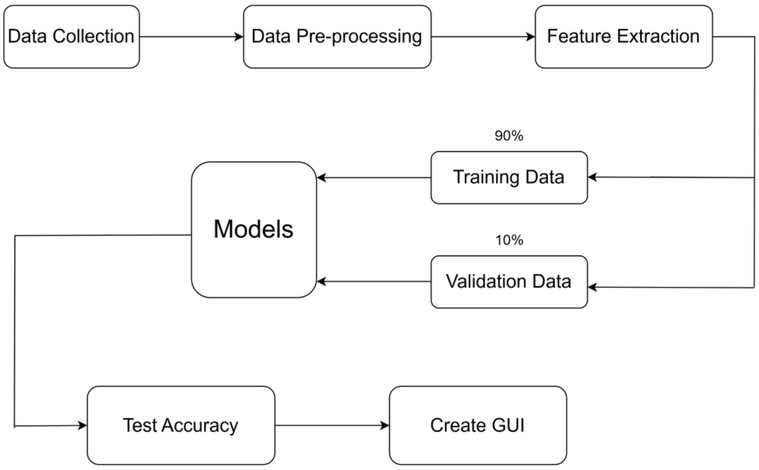

Figure 1. Methodology for developing machine learning-based mental illness detection program
Note: The figure above illustrates the comprehensive procedure employed in the development of the program.

The first stage involves data collection. The data was sourced from Hugging Face, an online open-source data platform. A dataset containing comments on mental illness was selected due to its comprehensiveness, simplicity, and the two-column structure that helps in reducing the workload. The dataset's characteristics include: Number of Rows = 42113, Number of Columns = 2, Number of Classes = 8.

Table 1. Distribution of text data across mental illness categories

Name	Scale	Amount
None	16.7%	7044
ADHD	14.9%	6280
BPD	13.5%	5667
Anxiety	13.3%	5581
PTSD	11.9%	5028
Depression	11.6%	4902
OCD	9.7%	4068
Bipolar	8.4%	3543

The second step involves Data Pre-processing. Initially, the text is converted to lowercase. Subsequently, functions from the Regular Expression Library in Python are utilized to eliminate all punctuation and irrelevant characters. Stopwords are then removed. Additionally, WordNetLemmatizer is employed for lemmatization, aiming to enhance text consistency and model performance.

The third step involves Feature Extraction, where Term Frequency-Inverse Document Frequency (TF-IDF) is utilized to assess the significance of individual words in a document within a collection or corpus. TfidfVectorizer, a class in the scikit-learn library, is employed to convert a set of documents into a matrix of TF-IDF features. Pipelines, a feature in scikit-learn, are beneficial for combining multiple steps for cross-validation with varying parameters. Pipelines encompass all estimators that execute fit and transform methods, facilitating the application of a series of data transformations followed by a final estimator. The vectorizer is incorporated into the Pipeline alongside model training algorithms, resulting in a Pipeline object capable of supporting training and prediction tasks.

The fourth step involves Model Training, where various algorithms are utilized to train the model and evaluate their performance.

3.1. Logistic Regression (LR)

LR is a statistical model where a logistic curve is utilized to fit the dataset [8]. Unlike Decision Trees or SVM, LR offers a clear probabilistic interpretation and allows for easy updating with new data [9]. This model makes fewer assumptions compared to others, as it does not assume a specific distribution of independent variables or a linear relationship between predictors and the target variable. It is capable of handling interaction effects, nonlinear effects, and power terms. However, it necessitates a large sample size to produce stable results. In this study, with a sample size of 42,000, LR model performed well and was selected as the preferred model.

Distinguished from the traditional dichotomous LR model, the research focuses on multinomial LR, specifically classifying eight types. The multinomial LR model is a simple extension of the binomial LR model. It is employed when the dependent variable consists of more than two nominal or unordered categories [10].

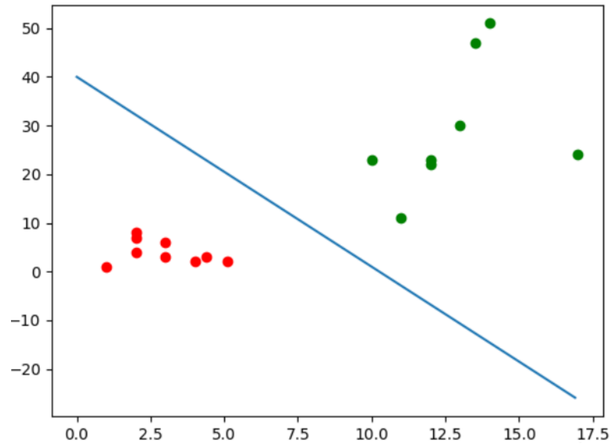

Figure 2. Illustration of dichotomous LR

Note: Multinomial LR can be decomposed into a series of dichotomous LR models.

3.2. Naïve Bayes

Naïve Bayes is a probabilistic classification algorithm that leverages the Bayes' Theorem beneath the suspicion of indicator autonomy [12]. It surveys the values of particular highlights to calculate the likelihood of a course based on Bayes' hypothesis, eventually selecting the course with the most elevated likelihood.

As the hypothetical establishment of Naïve Bayes classification [11], Bayes's hypothesis speaks to a straightforward expansion of conditional likelihood, giving an equation for deciding the likelihood of occasion A taking after the event of occasion B.

$$P(A|B)= \frac{P(A\cap B)}{P(B)} \text{ And } P(B|A)=\frac{P(A\cap B)P(B)}{P(A)}$$

Naïve Bayes has been commonly utilized in text classification assignments [14], as illustrated in this consideration, where the objective is to discover in case a client shows a particular mental illness through their posts on the Reddit stage. This strategy is well-suited for multivariate information and is particularly beneficial for dealing with expansive datasets. Besides, it is comparatively less computationally serious than other calculations such as Back Vector Machine (SVM) and Irregular Timberland (RF). The test comes about demonstrates that Naïve Bayes shows the quickest handling speed among the four calculations considered.

In this study, Naïve Bayes showed marginally second-rate execution in accuracy, precision, recall, and F1-score compared to other models. In any case, its comes about was still important and given a benchmark for comparison. The Naïve Bayes show accomplished an accuracy of 0.66, precision of 0.74, recall of 0.66, and an F1-score of 0.64 on the evaluated dataset. These measurements were hardly lower when differentiated with elective models such as LR, SVM, and RF inside this particular setting.

3.3. SVM

SVM is a broadly recognized and commonly utilized supervised learning strategy [13]. It serves as a strong classification model, particularly compelling in taking care of high-dimensional datasets. The elemental rule behind SVM includes distinguishing the ideal hyperplane that can viably isolate distinctive classes inside a dataset. The key objective is to maximize the edge between classes to upgrade the model's generalization capability. SVM is able to tend to nonlinear information by leveraging part functions to map information into a higher-dimensional space [15]. It is critical to recognize that whereas SVM offers certain hypothetical focal points, preparing the demonstration on a huge dataset may be time-consuming. In this study, the training time of SVM was altogether longer compared to other algorithms, but the results were palatable.

3.4. RF

RF is a machine learning algorithm that points to progress and shows accuracy by developing decision trees in a collective way [16,17]. The learning handle is conducted freely for each tree by utilizing an arbitrarily chosen subset of highlights, which serves to diminish show changeability and the hazard of overfitting. RF too suits a wide extend of input factors and gives an evaluation of the significance of each variable. In this consideration, the RF algorithm was employed to distinguish pointers of mental health concerns in printed posts extricated from Reddit. The tall level of accuracy, flexibility, and vigor of RF contributed to its adequacy when connected to broad printed datasets.

The fifth step involves model evaluation. To evaluate the models, this study compared the algorithms based on the following factors:

$$\text{Accuracy} = \frac{\text{number of patients predicted correctly}}{\text{number of all indivduals}}$$

$$\text{Precision} = \frac{\text{number of patients predicted correctly}}{\text{number of indivduals predicted as patients}}$$

$$\text{Recall} = \frac{\text{number of patients predicted correctly}}{\text{number of all patients}}$$

A trade-off between precision and recall is evident in machine learning models. Typically, as the precision of a model increases, the recall rate decreases. To assess a model based on the specific objectives and to adjust the emphasis between precision and recall, F-score is introduced. F-score, denoted as F_i, represents the weighted harmonic mean of the combined recall and precision metrics. The parameter i enables researchers to find a suitable balance between precision and recall. A value of i less than 1 prioritizes precision, while a value greater than 1 emphasizes recall [20]. F-score provides a comprehensive evaluation of algorithm performance across all classes. This study primarily utilizes F1-score for evaluation purposes.

$$\text{F1} = \frac{2precision \cdot recall}{precision + recall}$$

The sixth step involves Graphical User Interface (GUI) development. GUI is created using the Tkinter library in Python, which offers a simple and effective method for developing GUI applications [18].

The application comprises the following essential components:

3.5. Chatbox

A scrollable text widget has been implemented to exhibit the conversation between the user and the chatbot. This widget utilizes ScrolledText widget from Tkinter and is configured to be read-only, except during the transmission of new messages.

3.6. User Input Entry

An entry widget allows users to input their messages. This widget gathers user input and forwards it to the chatbot for further processing.

3.7. Send Button

A button is implemented to initiate the process of transmitting the user's message to the chatbot. Upon clicking the button, the user's message is extracted from the entry widget, exhibited in the chatbox, and analyzed by the machine learning model to produce a suitable response.

The primary functionalities of GUI application include the following:

3.8. Text Preprocessing

The user input undergoes preprocessing to transform it into a format appropriate for the machine learning model. This process includes converting the text to lowercase, removing punctuation and irrelevant characters, excluding stopwords, and applying lemmatization.

3.9. Model Prediction

The preprocessed text is subsequently inputted into LR model, which has been previously trained to categorize the text into various mental disorder classifications or ascertain the absence of a significant medical condition.

3.10. Response Generation

According to the model's prediction, a corresponding response is generated from a predefined set of messages. These responses are intended to offer support and provide fundamental advice concerning the identified mental disorder.

3.11. Displaying Response

The generated response is presented in the chatbox, mimicking a dialogue with the user.

4. Results

4.1. TF-IDF Scoring

TF-IDF is a numerical strategy utilized to evaluate the centrality of a particular word inside a specific report or a collection of archives. It comprises two key components: TF and IDF. TF speaks to the proportion of the recurrence of a term in an archive to the whole number of terms within the same record. IDF, on the other hand, gauges the significance of a term over the complete archive collection by taking the logarithm of the entire number of reports within the corpus isolated by the number of records containing the term.TF-IDF score is calculated as the product of TF and IDF, assigning higher scores to terms with high TF and low IDF. This technique is commonly employed in document retrieval to rank documents based on their relevance to a query and in comparing documents for similarity. By emphasizing important terms, TF-IDF improves the effectiveness of text analysis.

This study determined TF-IDF values for both Uni-grams and Bi-grams. Subsequently, the top 10 items was identified based on TF-IDF scores for each category to assess the significance of an item within the documents.

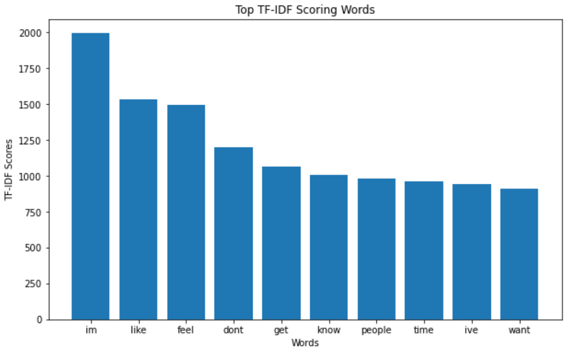

Figure 3. Top 10 Uni-grams based on TF-IDF scores

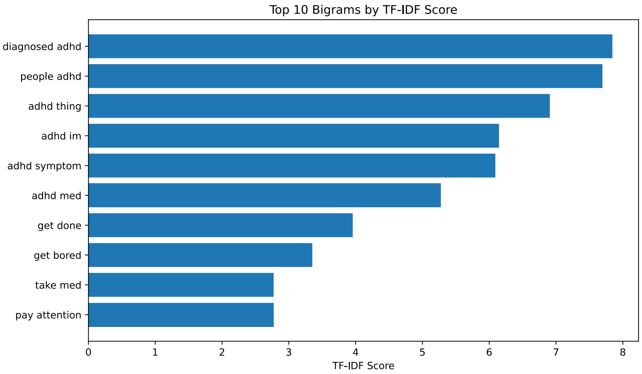

Figure 4. Top 10 Bi-gram2 based on TF-IDF scores

4.2. Model Performance

LR model was chosen for integration into the chat program due to its superior average performance in comparison to other models. The performance metrics for each model are outlined in the table above, revealing that LR exhibited the highest levels of accuracy, precision, recall, and F1-score. Consequently, this model was selected for incorporation into the chat program.

The running results of the four models are outlined in Table 2.

Table 2. Performance metrics of four machine learning models for mental illness detection

Model	Accuracy	Precision	Recall	F1-score
LR	0.81	0.81	0.81	0.81
Naïve Bayes	0.66	0.74	0.66	0.64
SVM	0.80	0.81	0.80	0.80
RF	0.76	0.77	0.76	0.76

Note: The values presented are the weighted averages across eight mental illness categories.

In Table 2, LR model demonstrates superior performance compared to the other three models. While SVM model may closely approximate the performance of LR model, it requires significantly more time to execute due to extensive computational requirements. Taking into account a comprehensive evaluation, this study utilized LR model for the proposed chat program.

Naïve Bayes exhibited slightly inferior performance in accuracy, precision, recall, and F1-score compared to alternative models. However, its results remained valuable and provided a benchmark for the study. The evaluation of Naïve Bayes on the dataset yielded an accuracy of 0.66, precision of 0.74, recall of 0.66, and an F1-score of 0.64. These metrics were marginally lower when contrasted with other models such as LR, SVM, and RF within the specific context of the study.

In LR analysis, a comparison was made between the unigram model and the bi-gram model. The results indicate that the uni-gram model demonstrates higher accuracy.

Table 3. Comparison between Uni-gram and Bi-gram models

Model	Accuracy	Precision	Recall	F1-score
Uni-gram	0.81	0.81	0.81	0.81
Bi-gram	0.48	0.46	0.47	0.45

5. Conclusion

This considers diagrams the improvement of a chat-based program outlined for the discovery of mental disarranges, with a center on basic perspectives such as dataset choice, feature extraction, algorithm execution, and execution assessment. The dataset, sourced from Hugging Face, comprises Reddit comments labeled with eight mental health conditions. TF-IDF was utilized for feature extraction, which empowered consequent machine learning applications. A comparative examination was conducted utilizing a few broadly recognized supervised learning calculations, counting LR, Naïve Bayes, SVM, and RF. The execution of these models was thoroughly assessed utilizing fitting measurements, and LR demonstrated developed as the foremost compelling, accomplishing the most elevated levels of accuracy, precision, recall, and F1-score. Thus, this demonstration was integrated into the chat program's GUI.

Modern mental clutter location frameworks overwhelmingly center on optimizing prediction accuracy, precision, and recall, regularly neglecting the significance of computational efficiency. In this research, the choice of LR over SVM was somewhat driven by contemplations of computational time, underscoring the requirement for timely detection to avoid the acceleration of extreme indications, such as suicidal ideation. Future research ought to investigate the computational complexity of different calculations, with the point of creating a more computationally proficient arrangement that empowers faster location.

Furthermore, most existing models are limited to English-speaking users, despite the fact that social media platforms are global and users communicate in multiple languages. In regions such as China, where the incidence of mental health issues is particularly high, users may express their emotions differently in their native language. Therefore, future studies should focus on developing multilingual mental disorder detection programs to enhance their applicability and effectiveness across diverse linguistic and cultural contexts.

References

[1] Qiao, J. (2020) A Systematic Review of Machine Learning Approaches for Mental Disorder Prediction on Social Media. CDS, Stanford, CA, USA, pp. 433-438,

[2] Mental disorders. (2013) [online] Available: https://www.who.int/mental_health/management/en/.

[3] X. Wang, C. Zhang, Y Ji, L. Sun, L. Wu and Z. Bao. (2013) A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network. LNCS., 201-13.

[4] Vazire, S., Gosling, S. D. (2004). e-Perceptions: Personality Impressions Based on Personal Websites. J Pers Soc Psychol, 87(1), 123–132.

[5] De Choudhury, M., De, S. (2014) Mental health discourse on Reddit: Self-disclosure social support and anonymity. Proceedings of ICWSM, pp. 71-80.

[6] Shen, J. H. and Rudzicz, F. (2017) Detecting anxiety on Reddit. CLPsych, pp. 58-65.

[7] De Choudhury, M., Kiciman, E., Dredze, M., Coppersmith, G., Kumar, M. (2016) Discovering shifts to suicidal ideation from mental health content in social media. CHI, pp. 2098-2110

[8] Lorena, A.C. et al. (2011) Comparing machine learning classifiers in potential distribution modelling. Expert Syst. Appl., 38:5268-5275.

[9] Singh, A., Thakur, N., Sharma, A. (2016) A review of supervised machine learning algorithms. INDIA Com, New Delhi, India., pp. 1310-1315.

[10] Bayaga A. (2010) Multinomial Logistic Regression: Usage and Application in Risk Analysis. JAQM., 5(2).

[11] Reddy E M K, Gurrala A, Hasitha V B, et al. (2022) Introduction to Naive Bayes and a review on its subtypes with applications. Bayesian reasoning and gaussian processes for machine learning applications., 1-14.

[12] Rish I. (2001) An empirical study of the naive Bayes classifier[C]//IJCAI 2001 workshop on empirical methods in artificial intelligence., 3(22): 41-46.

[13] Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. (1998) Support vector machines. IEEE Intell Syst Appl., 13:18–28.

[14] Flach P A, Lachiche N. (2004) Naive Bayesian classification of structured data. ML., 57: 233-269.

[15] Cho G, Yim J, Choi Y, Ko J, Lee SH. (2019) Review of Machine Learning Algorithms for Diagnosing Mental Illness. Psychiatry Investig., 16(4):262-269.

[16] Breiman, L., Friedman, H., Olshen, R.A., Stone, C J. (1984) Classification and regression trees, Wadsworth and Brooks. Monterrey, CA.

[17] Jaime Lynn Speiser, Michael E. Miller, Janet Tooze, Edward Ip. (2019) A comparison of random forest variable selection methods for classification prediction modeling. Expert Syst. Appl., 34:93-101.

[18] Moore A D. (2018) Python GUI Programming with Tkinter: Develop responsive and powerful GUI applications with Tkinter. Packt Publishing Ltd.

[19] Syarif I, Ningtias N, Badriyah T. (2019) Study on mental disorder detection via social media mining[C]//2019 4th International conference on computing, communications and security (ICCCS). IEEE., 1-6.

[20] Sokolova M, Japkowicz N, Szpakowicz S. (2006) Beyond accuracy, F-score and ROC: a family of discriminant measures for performance evaluation[C]//Australasian joint conference on artificial intelligence. Berlin, Heidelberg: Springer Berlin Heidelberg,

[21] Lin T. (1953) A study of the incidence of mental disorder in Chinese and other cultures. Psychiatry. , 16(4): 313-336.