Sentiment Analysis of Mental Health Statement

1st Yaxuan Huang
Department of Biostatistics
University of Michigan
Ann Arbor, United States
yaxuanh@umich.edu

***Abstract*—Mental health is crucial for individuals to function effectively in daily life. With the advent of mobile health technologies, digital tools are being developed to monitor mental health in real-time, deliver personalized notifications, and encourage healthier behavioral patterns. This study aims to evaluate machine learning methods that can assist mobile health applications in the early detection of psychological distress. Specifically, Support Vector Machines (SVM), Long Short-Term Memory (LSTM) networks, and Bidirectional Encoder Representations from Transformers (BERT) are applied to sentiment analysis on a publicly available dataset from Hugging Face. The SVM and LSTM models yield accuracies of approximately 77%, while BERT outperforms them with an accuracy of 83%. In conclusion, transformer-based architectures, such as BERT, offer promising potential for integration into mobile health tools to mitigate the severity of mental health deterioration and reduce suicide rates.**

***Index Terms*—Sentiment Analysis, Mental Health, LSTM Network, Bidirectional Encoder Representations from Transformers.**

I. Introduction

Mental health encompasses an individual's psychological, emotional, and social well-being, shaping how they think, feel, behave, and interact with others [1]. Mental health disorders can significantly alter cognition, mood, and behavior, thereby affecting daily functioning at home, work, school, and social environments [2]. Recent advances in natural language processing (NLP) and deep learning enable automated sentiment analysis of mental health-related statements, offering the potential for early detection of psychological distress.

With the rise of mobile health (mHealth) technologies, digital tools are being developed to monitor mental health in real time, deliver personalized notifications, and promote healthier behavioral patterns [3]. By identifying linguistic patterns associated with specific mental health conditions, such technologies can enhance early intervention strategies and support mHealth applications to provide timely assistance. These approaches may contribute to reducing the severity of mental health deterioration and lowering suicide rates.

Several prior studies have applied Support Vector Machines (SVM) for sentiment classification, with some enhancing performance through auxiliary techniques such as co-reference resolution [4]. Other classical machine learning models, including Naive Bayes and Random Forest classifiers, have also been utilized in this context [5]. Neural network approaches, particularly Convolutional Neural Networks (CNNs), are frequently adopted when the classification task involves complex linguistic patterns [6]. Ensemble methods have also been explored to leverage the strengths of multiple classifiers and improve overall predictive accuracy [7].

Reported precision values in these studies vary widely, typically ranging from 70–75% for simpler models to approximately 88–89% for more complex, deep learning-based architectures. Among these, neural network-based methods generally achieve the highest accuracy, while SVM remains a strong traditional method, especially when combined with feature engineering or additional data preprocessing steps [4] [5] [6] [7].

In this project, we aim to implement and compare the performance of three representative models for the emotion classification in statements related to mental health. Specifically, we investigate a traditional machine learning algorithm, SVM, a deep learning method using Long Short-Term Memory (LSTM) networks, and a transformer-based model, Bidirectional Encoder Representations from Transformers (BERT).

II. Method

A. Data Description

We utilize a publicly available dataset on Hugging Face (Link). The dataset contains 52,681 mental health statements labeled across seven categories: *Normal, Anxiety, Depression, Suicidal, Stress, Bipolar,* and *Personality disorder*. Our objective is to evaluate the performance of various models in accurately classifying these mental health sentiments and to assess the time required for model training. The dataset exhibits a slight class imbalance in the distribution of sentiment categories. Among the 52,681 annotated statements, the most frequently occurring labels are Normal (16,343 instances) Depression (15,404 instances), and Suicidal (10,652 instances), while categories such as Anxiety (3,841 instances), Bipolar (2,777 instances), Stress (2,587 instances), and Personality disorder (1,077 instances) are underrepresented. This imbalance poses challenges for classification.

B. Data Preprocessing

All statements were converted to lowercase, with punctuation, URLs, and stopwords removed. For the SVM model, the cleaned text was vectorized via the Term Frequency–Inverse Document Frequency (TF-IDF) transformation. In the LSTM network, the text was tokenized into integer sequences using Keras's Tokenizer and padded to a fixed length of 5,000 tokens. For BERT, the text was tokenized using the BERT

WordPiece tokenizer to generate input IDs and attention masks. The labels, originally in words, were mapped to integers in SVM and BERT. In the LSTM model, the labels were further converted into one-hot vectors. The original dataset was randomly divided into training and test sets using an 80%/20% stratified split to preserve the class distribution across all seven sentiment categories.

C. Model Architecture

1) SVM: A linear SVM was employed for multiclass classification. The feature vectors derived from the TF-IDF transformation were fed into the linear SVM. A one-versus-rest strategy was used to address the seven output classes.

2) LSTM: The LSTM networks consisted of an embedding layer, two LSTM layers, a fully connected dense layer, and a softmax output layer. The padded sequences were input into this network architecture.

3) BERT: The statements and corresponding labels were encapsulated into datasets and fed into the BertForSequenceClassification model with seven output labels. Fine-tuning of the model was performed on GPU to accelerate convergence.

All model hyperparameters—such as regularization strength for the SVM, sequence length and embedding size for the LSTM networks, and learning rate and batch size for BERT—were optimized on the training set to maximize performance on the held-out validation data.

III. Results

A. Accuracy

The results show that SVM and LSTM achieve a testing accuracy of approximately 77%, while BERT achieves a higher accuracy of 83%. The classification reports for SVM, LSTM and BERT are presented in Figure 1, 2, and 3. The SVM model shows a more balanced performance across all classes but tends to be slightly weaker in terms of precision, recall, and F1 score, particularly for Personality disorder, Stress, and Suicidal. It is generally less sensitive to the minority classes compared to BERT and LSTM, as indicated by lower recall values in those categories. LSTM performs well across most categories, with high precision and recall, but slightly lags behind BERT in the Suicidal and Personality disorder. BERT has higher precision, recall, and F1 scores across most of the classes, especially for Anxiety, Bipolar, and Normal. The recall for Normal is noticeably high. However, it shows slightly lower performance in Stress and Suicidal compared to the other two models.

B. Time

Although BERT achieves the highest accuracy among the three methods, it requires significantly more time for model training compared to SVM and LSTM. Specifically, SVM completed training in 5 minutes, LSTM took 50 minutes, and BERT required 80 minutes. Therefore, SVM is the fastest method of the three.

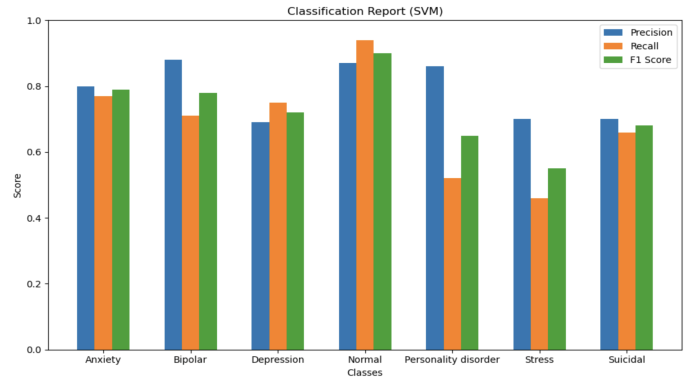

Fig. 1. Classification Report of SVM.

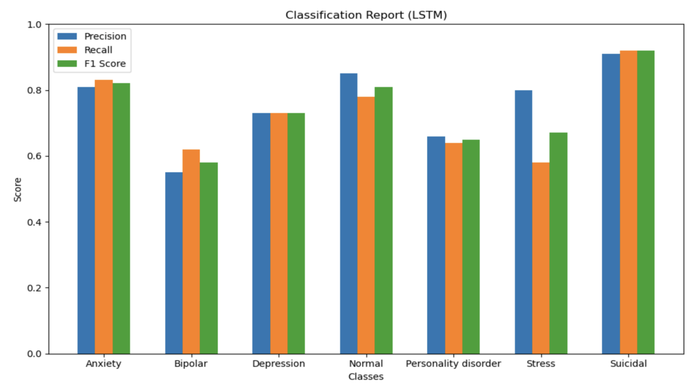

Fig. 2. Classification Report of LSTM Networks.

IV. Conclusion

SVM achieves comparable accuracy while requiring less time than LSTM. BERT, however, outperforms the other two methods in terms of accuracy. In conclusion, due to its high accuracy, the transformer-based architecture can be effectively integrated into mobile health tools to enhance the early detection of mental disorders and enable timely interventions. Future work may explore the integration of additional modalities, such as facial recognition and voice detection, to further enhance these mHealth tools.

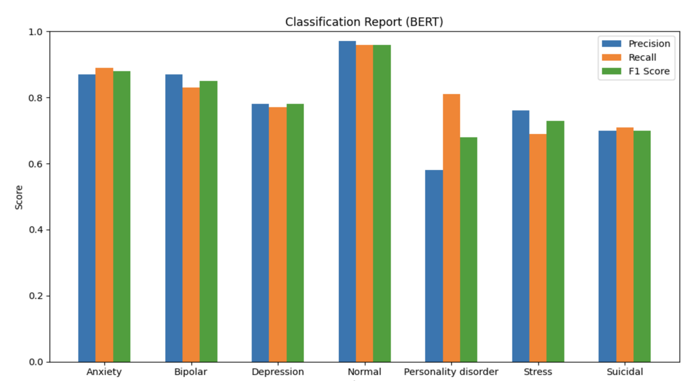

Fig. 3. Classification Report of BERT.

Acknowledgment

The author thanks Dr. Zhang for her help in this project and her exceptional ability to explain the material in STATS 507. The Github repository containing the code can be accessed here (Link).

References

[1] Centers for Disease Control and Prevention, “About mental health,” 2023. [Online]. Available: https://www.cdc.gov/mental-health/about/index.html.

[2] Cleveland Clinic, “Mental Health Disorders: Types, Diagnosis & Treatment Options,” 2022. [Online]. Available: https://my.clevelandclinic.org/health/diseases/22295-mental-health-disorders.

[3] Chou, Y. H., Lin, C., Lee, S. H., Chang Chien, Y. W., and Cheng, L. C. (2023). Potential mobile health applications for improving the mental health of the elderly: A systematic review. Clinical Interventions in Aging, 1523-1534.

[4] Devi, D. N., Kumar, C. K., and Prasad, S. (2016, February). A feature based approach for sentiment analysis by using support vector machine. In 2016 IEEE 6th international conference on advanced computing (IACC) (pp. 3-8). IEEE.

[5] Vanaja, S., and Belwal, M. (2018, July). Aspect-level sentiment analysis on e-commerce data. In 2018 International conference on inventive research in computing applications (ICIRCA) (pp. 1275-1279). IEEE.

[6] Duncan, B., and Zhang, Y. (2015, July). Neural networks for sentiment analysis on Twitter. In 2015 IEEE 14th International Conference on Cognitive Informatics and Cognitive Computing (ICCI* CC) (pp. 275-278). IEEE.

[7] Zirpe, S., and Joglekar, B. (2017, January). Polarity shift detection approaches in sentiment analysis: A survey. In 2017 International conference on inventive systems and control (ICISC) (pp. 1-5). IEEE.