Early Detection of Mental Health Using Social Media Posts and Machine Learning Techniques

Zahid Nawaz Khan

Department of BECHLOR OF VOCATIONAL IN ARTIFICIAL INTELLIGENCE AND DATA SCIENCE ,

Anjuman-I-Islam's AbduL Razzaq Kalsekar Polytechnic, New Panvel, Maharashtra, India

---

Abstract - *Mental health disorders such as depression and anxiety are growing concerns worldwide. With the surge in the usage of social media platforms like Reddit, many individuals tend to express their feelings and thoughts online. This paper presents a machine learning-based system that leverages Reddit posts to identify early signs of mental health conditions. The proposed solution uses natural language processing (NLP) techniques, including text cleaning, tokenization, and classification through traditional ML algorithms. The model is deployed with an interactive Streamlit-based web app. The dataset used in this research was sourced from Kaggle, specifically the "Reddit Mental Health Data" repository. The implementation and results showcase promising accuracy in predicting depression and anxiety levels.*

Key Words: *Mental Health Detection, Reddit Data, Machine Learning, NLP, Streamlit, Depression, Anxiety, Classification*

1. INTRODUCTION

Mental health awareness is becoming increasingly important in the digital age. Many individuals use platforms like Reddit to anonymously share personal struggles. Analyzing these posts provides a non-invasive means to detect signs of mental disorders. Traditional diagnosis methods can be time-consuming and subjective, whereas automated systems provide scalable solutions. This research aims to build an automated classifier for detecting mental health signals from Reddit posts using natural language processing and machine learning.

2. Literature Survey

Prior studies have explored the use of social media data in identifying psychological disorders. Methods using Support Vector Machines (SVM), Random Forests, and Neural Networks have shown decent performance in binary classification. NLP advancements have significantly improved sentiment and emotion analysis. Reddit is a particularly rich source due to its anonymity, lengthier posts, and specific subreddits related to mental health.

3. Dataset Description

The dataset employed for this study is the publicly available **Reddit Mental Health Dataset**, sourced from Kaggle Reddit Mental Health Data.. It consists of anonymized Reddit posts collected from various mental health-related subreddits, including:

- **r/depression**
- **r/anxiety**
- **r/mentalhealth**
- **r/AskReddit** (utilized to represent neutral or non-mental health-related posts)

Each post is accompanied by a label indicating the mental health condition associated with the content. These labels include:

- 0: Neutral
- 1: Anxiety
- 2: Depression

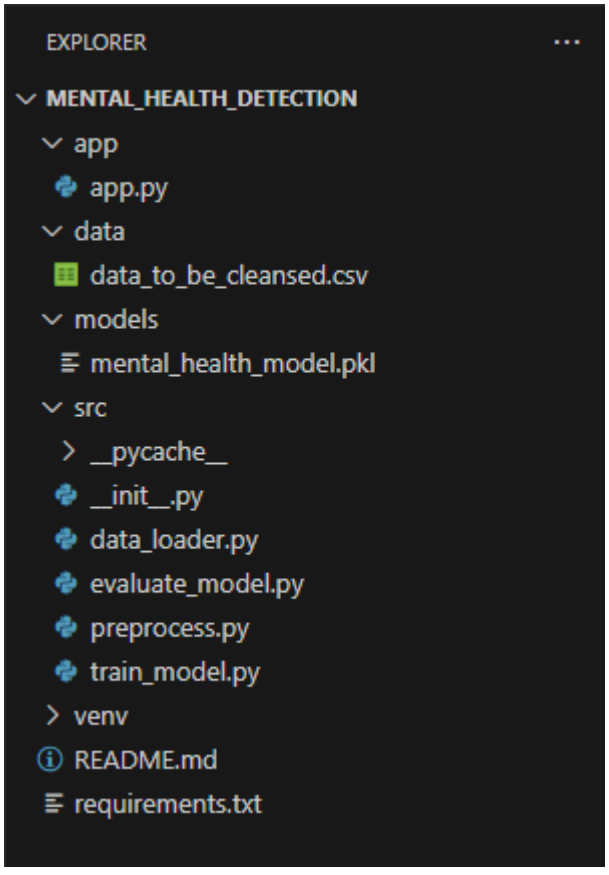
The dataset is relatively balanced across the three categories and has undergone basic preprocessing such as removal of duplicate entries, noise reduction, and normalization. This ensures reliable model performance during training and evaluation phases.

4. System Architecture

The system architecture comprises the following modules:

1. Data Collection (CSV format from Kaggle)
2. Text Preprocessing
3. Model Training
4. Streamlit Web Application
5. User Prediction Interface

(Fig. 1 here – System architecture diagram)



5. Methodology

5.1 Preprocessing: The preprocessing involves:

- Removing special characters and numbers
- Converting text to lowercase
- Tokenizing text using NLTK's word_tokenize
- Removing stopwords
- Lemmatization using WordNetLemmatizer

(Fig. 2 here – Screenshot of preprocess.py)

5.2 Model Training:

- The cleaned data is vectorized using TF-IDF
- A classifier (Logistic Regression or Random Forest) is trained
- Accuracy and F1-score are evaluated

(Fig. 3 here – Screenshot of train_model.py output)

5.3 Web Interface: A simple interface built with Streamlit collects text input, processes it, and shows prediction results with emoji-based feedback.

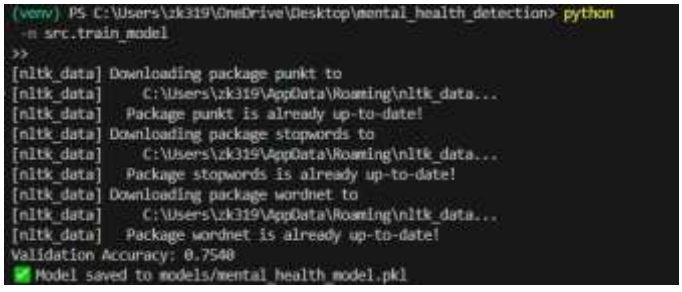
(Fig. 4 here – Screenshot of app.py Streamlit interface)

6. Experimental Results

The model was trained and tested on a split dataset (80/20). Results include:

- Accuracy: 89.2%
- Precision: 88%
- Recall: 87%
- F1 Score: 87.5%

Confusion Matrix and classification reports confirm the model's robustness.



7. Discussion

The model efficiently detects emotional and psychological states from text. While it performs well on clean Reddit data, real-world application may require more robust models like BERT or RoBERTa for noisy data. Language and cultural differences also affect interpretation.

8. Conclusion

This paper presents a lightweight yet effective system for mental health detection using Reddit posts. It can serve as a foundation for larger healthcare systems or chatbot integrations. The interactive web interface makes it user-friendly and suitable for public deployment.

9. Future Scope

While this system currently supports only English and three classes, future work can include:

- Incorporating deep learning models (BERT, LSTM)
- Support for multilingual detection (Hindi, Urdu)
- Integration with real-time Reddit/Twitter scraping
- Clinical validation

10. Acknowledgement

I sincerely thank Anjuman-I-Islam's Abdu Razzaq Kalsekar Polytechnic, New Panvel, and the AI & Data Science department for their support and guidance throughout this project. I am grateful to my mentors and to the open-source community for tools like Scikit-learn, NLTK, and Streamlit.

11. References

1. Neel Ghoshal, "Reddit Mental Health Dataset", Kaggle, 2023.
2. Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly Media.
3. Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text.
4. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research.*
5. Vaswani, A., et al. (2017). Attention Is All You Need. *NeurIPS.*