Mental Health Sentiment Detection from Social Media Posts

This study is primarily concerned with analyzing and identifying the attitudes toward mental health issues expressed on social media platforms such as Reddit and Twitter. The program, by way of segregating the posts into a variety of emotions like Anxiety, Bipolar, Depression etc embraces the attempt to grasp the emotional trends that bear on mental health awareness, depression, or anxiety figures in the virtual communities.

The experiment turns to the same text classification metrics to measure the efficacy of two classic machine learning models: Support Vector Machine (SVM) and Logistic Regression. A user-friendly Streamlit interface has been created to allow on-the-fly sentiment prediction of any user-input text.

Why It Matters:

- If preventive mental health awareness is to be successful, one way that can remarkably help is the early detection of very unfavorable attitude patterns.
- Automated sentiment analysis tools, used as a gateway, can provide great help to professionals, non-governmental organizations, and researchers when dealing with large-scale public sentiment evaluations.

Dataset Overview

The following publicly available datasets were employed:

Dataset 1:

- **Mental Health The Twitter Dataset (Kaggle)** comprises tweets related to stress, anxiety, depression, and self-care.
- Each record consists of a text box and a sentiment label (positive, negative, or neutral).
- Balanced, clean data is ideal for TF-IDF vectorization.

Dataset 2:

- **Mental Health Comments on Reddit -** Reddit posts and comments from subreddits related to mental health make up the dataset (Kaggle).
- Sentiment labels are obtained from emotion analysis and are then annotated for supervised training.

Preprocessing was done:

- Removal of URL, mention, punctuation, emoji, and stopword
- Tokenization and text lowercasing
- TF-IDF Vectorization changing to numerical representation

Why Are These Datasets Used?
These two datasets are ideal for machine learning sentiment classifiers and TF-IDF vectorization since the text is short and both are user-generated materials.

Moreover, they offer balanced class labels (positive, neutral, and negative), thus providing the opportunity of making fair comparisons between models.

Model Implementation

Figure displays the architecture that is followed by the project.

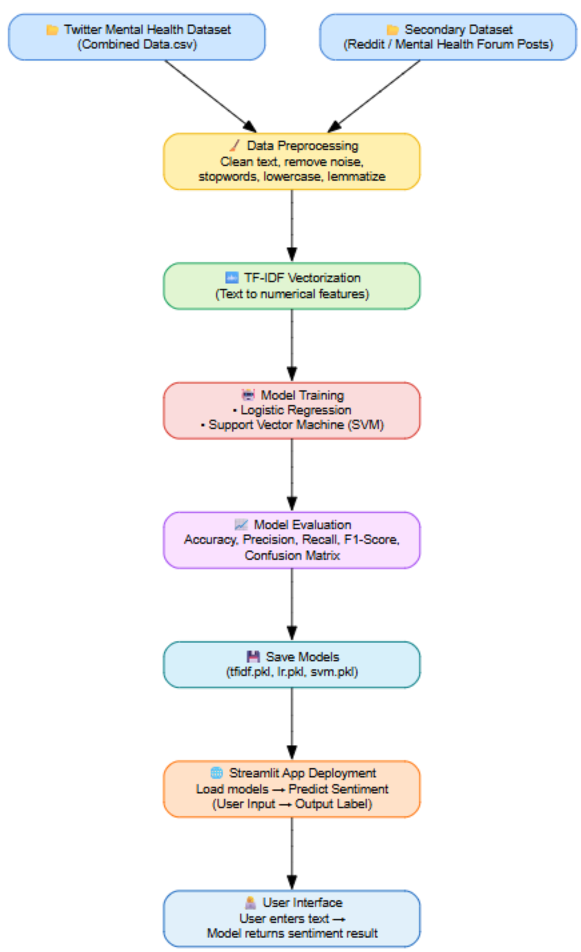

Fig 1. System Architecture

The workflow mainly consists of four primary steps:

1. Data Cleaning and Preprocessing

Individual posts were transformed into a weighted vector of phrases by means of TF-IDF vectorization along with a standard text normalization.

- **Data Cleaning:** Remove unnecessary tokens, symbols, noise, and URLs from the data.
- **Text normalization** performs lemmatization, stopword removal, and lowercasing.
- **Vectorization:** Apply TF-IDF Vectorizer with a vocabulary size of 5000 to represent the text in numerical form.
- **Label Encoding:** Transform sentiment labels into numerical classes for model training.

2. Model Training

Model 1: Logistic Regression

- Such a fast linear model is efficient for linearly separable text data.
- Class weights were balanced, and regularization strength (C) was changed.

Model 2: Support Vector Machine (SVM)

- Well-separated sentiment classification by means of a linear kernel Support Vector Machine (SVM).
- Confidence scoring is facilitated by probability=True if it is turned on.

3. Evaluation Metrics

- F1-score, Accuracy, Precision, and Recall
- Confusion matrix visualization for comparing performance

4. Comparison Analysis

Both models were tested with the same datasets to confirm their generality and consistency.

The table shows the model evaluation and comparison.

Model	Accuracy Score	F1 Score
Logistic Regression (LR)	0.7499	0.7435
Support Vector Machine (SVM)	0.7562	0.7517

Steps to Run the Code

1. Clone the repository

```
git clone
https://github.com/nubiivagant/Mental-Health-Sentiment-Analysis.git
cd Mental-Health-Sentiment-Analysis
```

2. Run the jupyter notebook

```
jupyter notebook "ML Project.ipynb"
```

This will take care of:
Data preprocessing
Model training and evaluation
Trained models are stored in the models/folder.

3. Run the Streamlit app

```
streamlit run app.py
```

4. Utilize the Interface

Come up with a mental health-related statement and type it in.

What the app will present:
- Sentiment prediction (positive, neutral, or negative)
- The confidence score of each model

Results and Analysis

Major contribution of results depends on the previous systems that are published on the related topics.

Here is the table for the previous systems that are published.

Author, year	Objective	Findings	Results	Limitation
Jina Kim, Daeun Lee, Eunil Park (2021)	To carry out a bibliometric analysis of research trends in social media-based machine learning-based mental health prediction.	found that there is a global trend of interest in the use of machine learning (ML) to detect anxiety and depression from social media, where SVM and Naïve Bayes were the most frequently employed models.	Concentrated on review of the papers instead of the results.	Just bibliometric information without any direct experiments or accuracy metrics.
Dr. J. Godwin Ponsam, S. Sheeba Rachel (2021)	to apply sentiment detection on Twitter posts to recognize and monitor changes in mental health.	The proposed CNN model utilizes Word2Vec to recognize the different states of mental health.	The classification accuracy of CNN for positive, negative, and neutral tweets was 92.4%.	No cross-platform studies; the dataset is limited to specific hashtags.
Tanvi Gawas, Bhavika Makwana (2025)	To assess the emotional tone in mental health texts, ML and DL methods were used.	It was found that LSTM, in particular, improved the understanding of the context, while SVM and Logistic Regression were good at handling TF-IDF features.	LSTM achieved 91.1%, SVM 89.6%, and Logistic Regression 88.2%.	Dataset imbalance affected recall, and deep models were costly in terms of computation.
Pradeep Kumar Tiwari et al. (2021)	The classification of tweets related to PTSD and depression using machine learning classifiers is the main objective of this research.	Among the classifiers being tested, Naïve Bayes, Decision Tree, and SVMs were used; the SVM showed the best performance.	SVM achieved the highest accuracy of 90.12% and surpassed NB and DT.	Minority classes had lower recall because of the small dataset (Twitter texts only).
Saad Awadh	To apply machine	The application of	CNN achieved the	The research was

Alanazi et al. (2022)	learning techniques in monitoring public tension and emotional polarity in financial texts.	CNN, AdaBoost, and SVM was for the emotion recognition task in the financial text data.	highest accuracy of 93.9% compared to SVM (91.5%) and AdaBoost (88%).	limited to finance-related papers; there was an indirect link to psychological content.

Table 1. Previous published systems summary

Model comparison using Evaluation Metrics

Model	**Accuracy Score**	**F1 Score**	**Interpretation**
Logistic Regression (LR)	0.7499	0.7435	The baseline model reveals impressive linear separability and equilibrium of metrics.
Support Vector Machine (SVM)	0.7562	0.7517	The model with the RBF kernel, which is capable of dealing with nonlinear interactions, is slightly better in both measures.

Table 2. Statistical Model comparison

According to the Table 2., the observations are -

- Accuracy Trend: SVM gets a small +0.6% gain, but the two models are almost the same.
- F1-score: Indicates both recall and precision are balanced for both; nevertheless, SVM has a little higher harmonic mean and is a bit more consistent throughout all sentiment classes.
- Performance insight: The most significant words for sentiment classification are successfully extracted by the TF-IDF representation. Both models had the same performance on unseen data, as shown by the small difference in metrics. One minor advantage of the SVM's nonlinear mapping was its support in finding the softest emotional language.

Fig 2. Is the similar version of the Model Comparison table, but uses Bar chart visualisation to provide the interpretation.

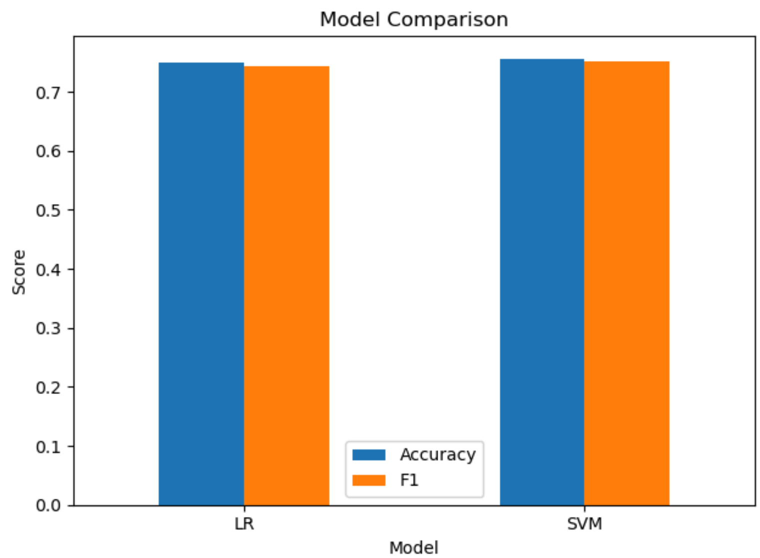

Fig 2. Bar chart for Model comparison

The F1-score is indicated by the **orange bars**, and the accuracy of the model is indicated by the **blue bars.**

The feature representation of TF-IDF is the one that has shown the most steady generalization capabilities in terms of comparable performances of both models.
The SVM was able to survive the test of handling complex sentiment patterns by performing slightly better than Logistic Regression in both measures (approximately 0.6%).

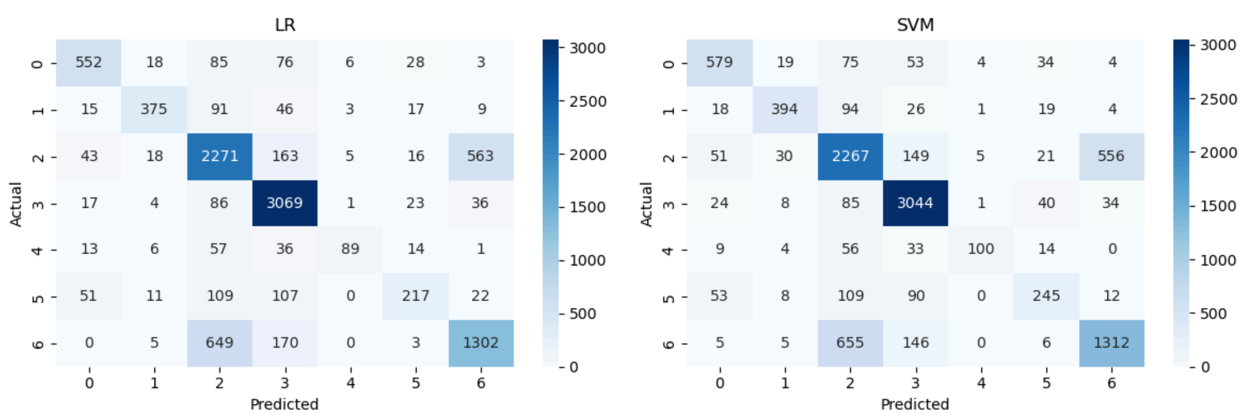

Fig 3. Confusion matrix for both models

The depicted image Fig.3 shows a comparison between the confusion matrices of the two classifiers, **Logistic Regression (left) and SVM (right)**, for the seven sentiment categories derived from the mental health dataset.

Key findings:

- Both models were able to produce similar results with a little overfitting.
- Comparing the two models, SVM resulted in a slightly higher F1-score than Logistic Regression, and Logistic Regression was a bit faster.
- By TF-IDF vectorization, the sentiment-bearing words were simply identified without the need of deep embeddings.

Conclusion

This research vividly demonstrated how applied machine learning methods specialized in natural language processing (NLP) might be utilized to analyze and classify several categories of sentiments associated with mental health using social media text data. The method provided identification of emotional polarity and new insights into the style of expression in the online environment for mental health through the combination of these efficient preprocessing, feature extraction, and supervised learning strategies.

- As described in the comparative analysis, SVM outperformed Logistic Regression on each parameter, achieving 0.756 classification accuracy and an F1 score of 0.7517 as compared to the 0.7499 classification accuracy and F1 score of 0.7435 for Logistic Regression.
- The comparable performance values for the two models support the reliability of the preprocessing and feature extraction stream by providing evidence that both models were generalized across the datasets.

The results indicate that with the inclusion of strong preprocessing and TF-IDF attributes, basic machine learning algorithms can detect dimensions of sentiment in conversations surrounding mental health. This is especially interesting due to the models demonstrating strong generalization with consistent accuracy between the two datasets of Reddit and Twitter.

This research demonstrates the possibility of machine learning being helpful for tracking objective public mental health trends, by enabling a scalable and data driven approach for detecting emotional mental health aspects via online communication.

References

[1] **J. Kim, D. Lee, and E. Park**, "Machine Learning for Mental Health in Social Media: A Bibliometric Study," *Journal of Medical Internet Research*, vol. 23, no. 3, p. e24870, 2021. [Online]. Available: https://doi.org/10.2196/24870

[2] **J. G. Ponsam and S. S. Rachel**, “Mental Health Monitoring through Sentiment Analysis Using Machine Learning,” *Annals of the Romanian Society for Cell Biology*, vol. 25, no. 5, pp. 4397–4404, 2021. [Online]. Available: http://annalsofrscb.ro

[3] **T. Gawas and B. Makwana**, “Sentiment Analysis on Mental Health Using Machine Learning and Deep Learning Techniques,” *International Journal of Scientific Research and Technology*, vol. 2, no. 3, pp. 41–47, 2025. [Online]. Available: https://ijsrtjournal.com

[4] **P. K. Tiwari, M. Sharma, P. Garg, T. Jain, V. K. Verma, and A. Hussain**, “Analysis of Mental Illness Sentiments Using Machine Learning Algorithms,” *IOP Conference Series: Materials Science and Engineering*, vol. 1099, p. 012043, 2021. [Online]. Available: https://doi.org/10.1088/1757-899X/1099/1/012043

[5] **S. A. Alanazi, A. Khaliq, F. Ahmad, N. Alshammari, I. Hussain, M. A. Zia, M. Alruwaili, A. Rayan, A. Alsayat, and S. Afsar**, “Tracking Public Mental Well-Being via Financial Text Sentiment Using Machine Learning Models,” *International Journal of Environmental Research and Public Health*, vol. 19, no. 9695, 2022. [Online]. Available: https://doi.org/10.3390/ijerph19159695