NLP Project Report

Suicide Analysis and Prevention Application Using Machine Learning Classifiers

Team Members:

Mayuresh Bhangale

Kartik Gonde

Pavithra Tupakula

---

1. Introduction

Mental health issues such as depression and anxiety are increasingly prevalent, especially among young people aged 15 to 29. According to the World Health Organization, suicide is the second leading cause of death in this age group. While suicidal ideation often goes unspoken in real life, many individuals express distress through online forums like Reddit.

This project aims to build a Natural Language Processing (NLP) pipeline to detect suicidal ideation in Reddit posts. By leveraging machine learning classifiers and real-time data, we strive to create a system that could help identify at-risk individuals early and potentially prevent suicide.

---

2. Objectives

- Detect suicidal ideation using text-based machine learning.
- Use real-time data collection from Reddit for better generalization.
- Provide risk categorization with relevant recommendations (helpline, psychiatrist).
- Evaluate and optimize the model for both accuracy and recall.

---

3. Data Collection and Preprocessing

Sources:

- Subreddits: r/depression and r/SuicideWatch
- Tool: Reddit API (PRAW)

Data Points:

- Features: title, selftext, author, num_comments, url
- Label: is_suicide → 1 for suicidal, 0 for non-suicidal

Preprocessing:

- Tokenization, lowercasing, punctuation removal
- Lemmatization using WordNet
- Removal of stopwords
- Combined title, author, and selftext into megatext_clean

Psychological Stages:

Each post is categorized into one of six stages from “falling short of expectations” to “disinhibition,” allowing for nuanced risk assessment.

---

4. Exploratory Data Analysis (EDA)

a.

Top Words & Word Clouds:

- Extracted using CountVectorizer
- Separate clouds created for each subreddit

b.

Post Length Analysis:

- Avg. r/depression post: 964 words
- Avg. r/SuicideWatch post: 836 words

c.

Double Posters Detection:

- Users who post in both subreddits were identified and analyzed

Initial Scores

model	AUC Score	precision	recall (sensitivity)	best score	train accuracy	test accuracy	baseline accuracy	specificity	f1-score

cvec+ multi_nb	0.72	0.67	0.67	0.65	0.68	0.67	0.52	0.69	0.67
cvec + ss + knn	0.62	0.57	0.57	0.6	0.74	0.57	0.52	0.47	0.57
cvec + ss + logreg	0.73	0.69	0.69	0.65	0.69	0.69	0.52	0.67	0.69

---

5. Model Building

a. Baseline Accuracy:

Assuming all posts are suicidal (majority class) → 71.66%

b. Models Used:

- Count Vectorizer + Multinomial Naive Bayes
- TF-IDF Vectorizer + Logistic Regression
- CNN, LSTM (for future enhancement)

c. Feature Column Chosen:

megatext_clean – combined cleaned text from post, title, and username

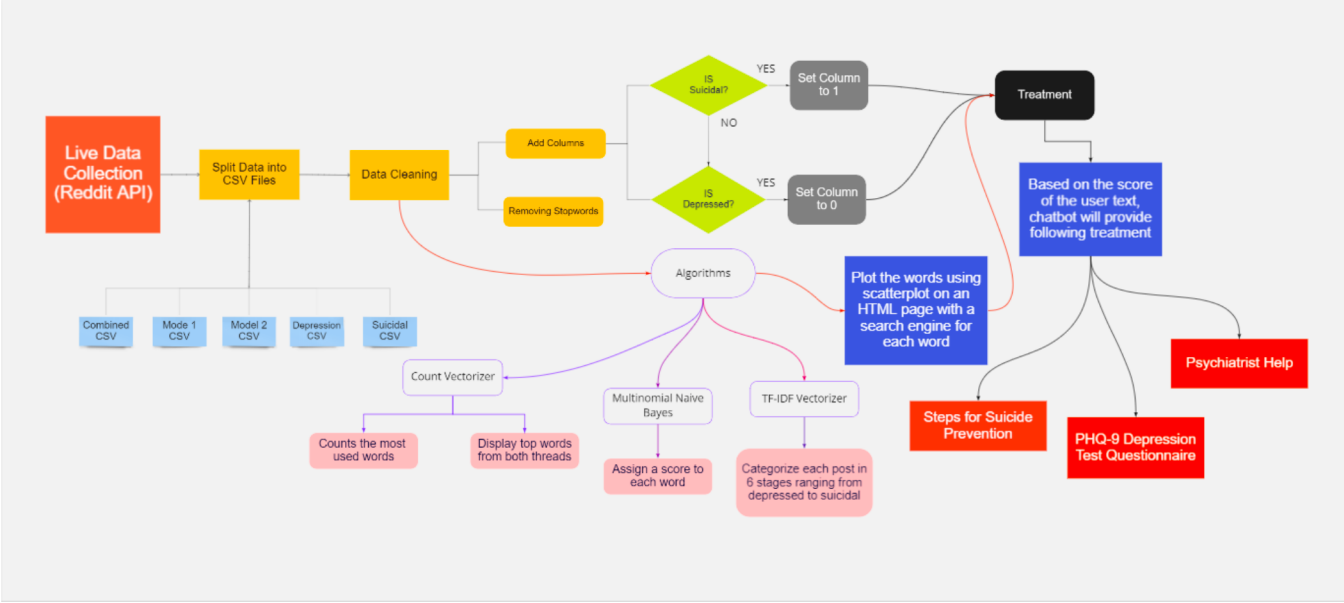
d. Best Model Configuration:

- Model: Count Vectorizer + TF-IDF + Multinomial Naive Bayes
- Parameters: unigram (1,1), stopwords removed, max_features=50

Initial Scores

model	AUC Score	precision	recall (sensitivity)	best score	train accuracy	test accuracy	baseline accuracy	specificity	f1-score
cvec+ multi_nb	0.72	0.67	0.67	0.65	0.68	0.67	0.52	0.69	0.67
cvec + ss + knn	0.62	0.57	0.57	0.6	0.74	0.57	0.52	0.47	0.57
cvec + ss + logreg	0.73	0.69	0.69	0.65	0.69	0.69	0.52	0.67	0.69

Proposed Architecture



6. Results

The final model was evaluated using a test dataset, yielding the following performance:

model	AUC Score	precision	recall (sensitivity)	best score	train accuracy	test accuracy	baseline accuracy	specificity	f1-score
cvec+ multi_nb	0.72	0.67	0.67	0.65	0.68	0.67	0.52	0.69	0.67
cvec + ss + knn	0.62	0.57	0.57	0.6	0.74	0.57	0.52	0.47	0.57
cvec + ss + logreg	0.73	0.69	0.69	0.65	0.69	0.69	0.52	0.67	0.69
tvec + multi_nb	0.73	0.68	0.68	0.65	0.68	0.68	0.52	0.67	0.68
tvec + ss + knn	0.56	0.54	0.55	0.59	0.73	0.55	0.52	0.5	0.54
tvec + ss + logreg	0.73	0.67	0.67	0.65	0.68	0.67	0.52	0.69	0.67
hvec + multi_nb	0.77	0.69	0.68	0.72	0.89	0.68	0.52	0.77	0.68
hvec + ss + knn	0.51	0.75	0.52	0.52	0.52	0.52	0.52	0.0	0.36
hvec + ss + logreg	0.65	0.62	0.62	0.63	1.0	0.62	0.52	0.61	0.62
hvec + multi_nb(tuning)	0.75	0.68	0.68	0.69	0.82	0.68	0.52	0.67	0.68
tvec + multi_nb(tuning)	0.75	0.69	0.69	0.68	0.71	0.69	0.52	0.67	0.69
hvec + multi_nb (tuning_2)	0.76	0.68	0.68	0.69	0.84	0.68	0.52	0.68	0.68
tvec + multi_nb (tuning_2)	0.75	0.7	0.7	0.68	0.71	0.7	0.52	0.69	0.7

As a final step in our EDA, we used Scatter text to produce a user-friendly way of visualising our corpus in HTML.

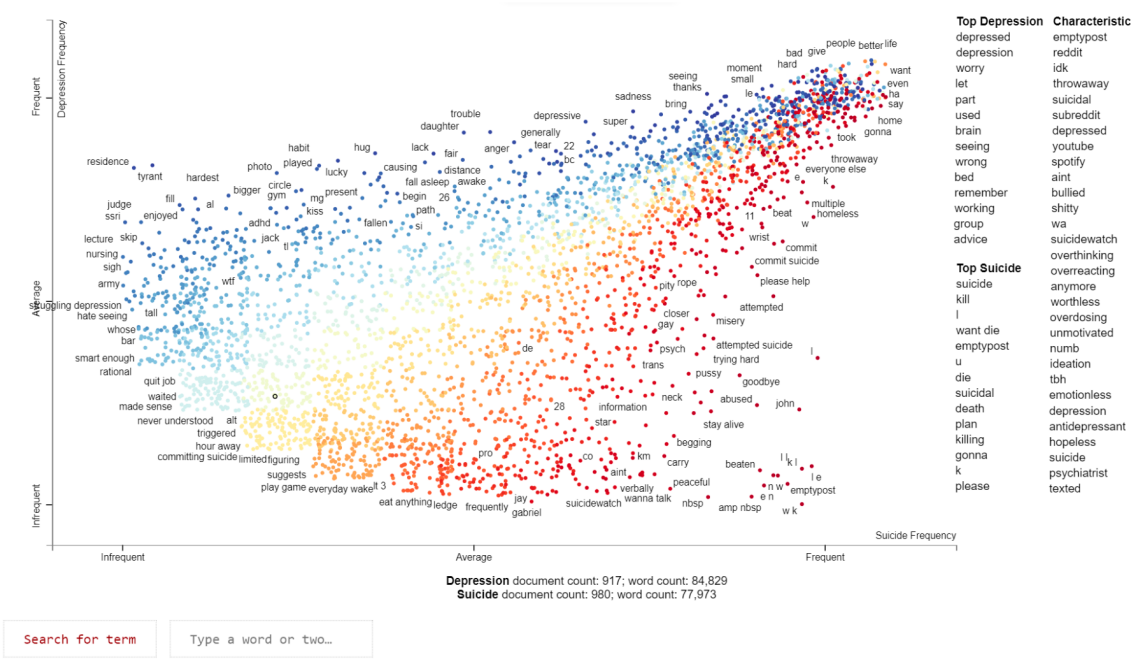
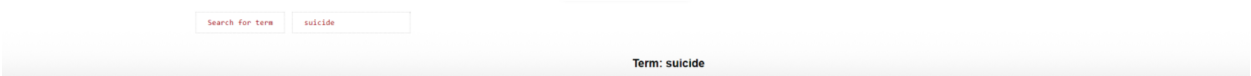


Fig -14: Data Analysis Corpus Scatter Plot



---

7. Model Deployment (Proposed) and future work

A web-based front end is proposed, allowing users to input text which is then analyzed:

- If suicidal risk is detected:

- Display national suicide helpline numbers
- Provide nearby psychiatrist suggestions (via Google Maps API)
- Optionally link to a psychometric assessment

---

8. Conclusion

This project successfully demonstrates the use of NLP for detecting suicidal ideation from social media posts. The trained model achieved **69.89% accuracy** and an **AUC of 75.48%**, suggesting that while improvements are possible, the current system provides a valuable base for real-time suicide prevention applications.

The approach is dynamic, ethical (anonymized data), and scalable for integration into digital mental health tools. More advanced NLP models can improve future performance.

---

10. Future Work

- **Incremental Learning with Updated Data:** Currently, Reddit's free API subscription allows only 1000 posts per subreddit. In the future, we aim to collect posts based on time filters (e.g., by month) to build an incremental dataset. This approach will keep the training data constantly updated and contextually relevant, and we plan to retrain the model periodically on this fresh data to improve robustness and adapt to evolving language patterns.
- **Improve Accuracy & Recall:** Optimize model parameters and features to exceed 80% accuracy.
- **Deploy Deep Learning Models:** Use BERT, RoBERTa, or GPT fine-tuning for better contextual understanding.
- **Temporal Analysis:** Track user post history to detect emotional trajectory over time.
- **Model Interpretability:** Integrate LIME or SHAP to explain predictions.

- **Multilingual Support:** Expand detection capabilities to non-English Reddit communities.

11 . GitHub Link :

https://github.com/mayureshbhangale/NLP

12. Youtube Link

https://youtu.be/7ampm4ZWkFU