

Machine learning for depression analysis

DSF - PT07 GROUP 11
21 November 2024



Outline

1. Group Introduction
2. Problem Statement
3. Approach
4. Data Understanding
5. Sentiment Analysis
6. Data Preparation
7. Modeling
8. Model Evaluation and Interpretability
9. Demonstration
10. Recommendations
11. Challenges





Meet the Team



Nduku
Kiteng'e



John Isaac
Mwangi



Phoebe
Wawire



Geoffrey
Shikanda



Joseph
Matheri



Cecily
Wahome



Keziah
Gicheha

Can we use data science to differentiate between depressive and non-depressive content based on text patterns on Reddit?

Problem Statement



Objectives

1. To predict whether a Reddit post indicates depression or non-depression based on its content.
2. Develop a predictive model for classifying depressive vs. non-depressive posts.
3. Deploy a user-friendly tool to analyze reddit posts.



Target Audience

We identified **three** primary audiences.



Social Media Platforms

Have a vested interest in improving content moderation and detecting high-risk posts



AI and Social Impact Investors

Would provide the necessary capital to scale and maintain the project, seeking a return on investment or social impact metrics as the tool gains adoption.



Academic and Research Institutions

Would develop studies related to mental health trends, language analysis, and social media behavior.





Hypothesis

An **increased** number of negative posts could mean that the person is likely to end up depressed.





Ethical Considerations

SAFETY FIRST

Given the sensitive nature of the data, ethical and responsible use is at the forefront of our approach. These are the key ethical principles guiding our model



NON ALARMIST MESSAGING

The model will be used as a **supportive** tool to flag potential depressive posts, not as a definitive diagnostic tool.



PRIVACY PROTECTION

All analysis will respect user anonymity, with no personal identification attached to the posts.



NON STIGMATIZING

The model will **avoid attaching stigmatizing terms** to users



Our Approach

Data Understanding

Session to load and review the data to understand the information it can provide. Scrape reddit for new data

Tools used:

- Bar graphs
- Line graphs
- Word Cloud
- Histograms
- Box plots

Data Preparation and Preprocessing

Includes data cleaning for missing values, data types and null values and feature engineering to enhance potential analysis methods and variables. Initial sentiment analysis was carried out using VADER.

Modelling and interpretation

Using the cleaned and prepared data, we review the potential models based on the objectives. This will help us to identify the best one to use.

Models included:

- Logistic regression
- Random Forest
- BERT

Recommendations

Develop a list of recommendations for the clients especially around which models would work best for future sentiment analysis projects.

Data understanding



Data source: Kaggle.

Data scraping: (same features but no labels)

Tool used - PRAW: The Python Reddit API Wrapper

Posts collected from subreddits (*Depression, SuicideWatch, Teenagers, happy & Deep thoughts*).

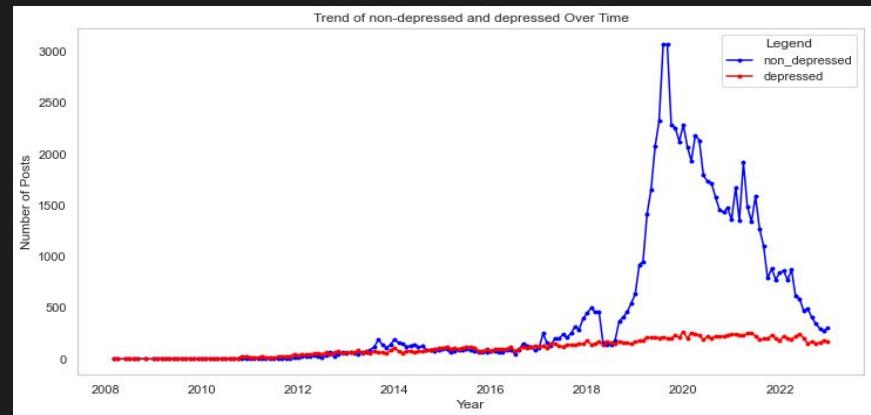
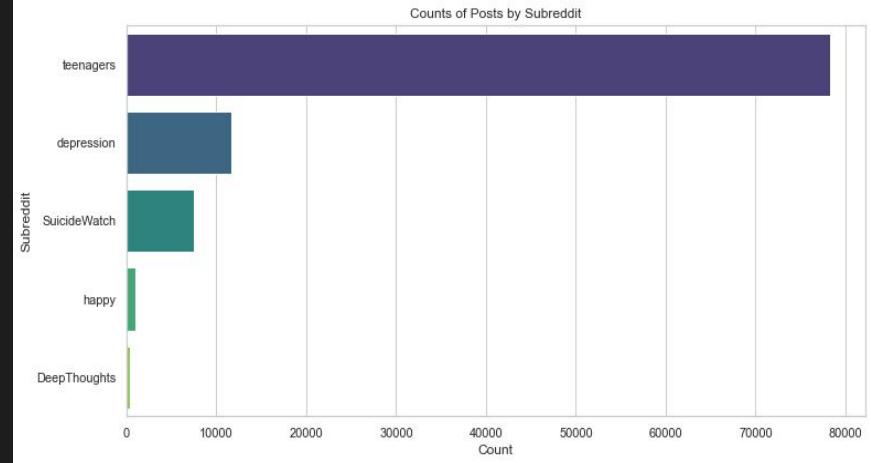
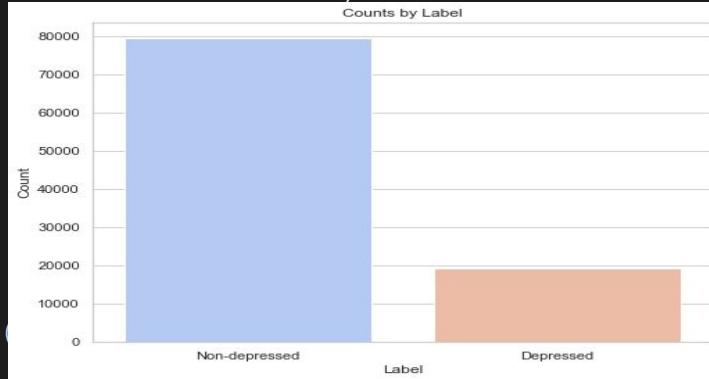
Size: 98, 826 posts, 7 features, 2 classes (Depressed = 1, Non-depressed = 0).

Features:

Text: Title and body of posts (primary inputs).

Numerical: date, num_comments, upvotes.

Labels: Derived from subreddits (0/1)(binary classification).



Sentiment Analysis



What We Did:

- Analyzed the emotional tone in Reddit posts to identify patterns of depression and well-being.
- Used a tool called **VADER** to measure:
 - Positive(+1)**, **Negative(-1)**, and **Neutral (0)** emotions.
 - Combined these into an overall **sentiment score**.

How It Works:

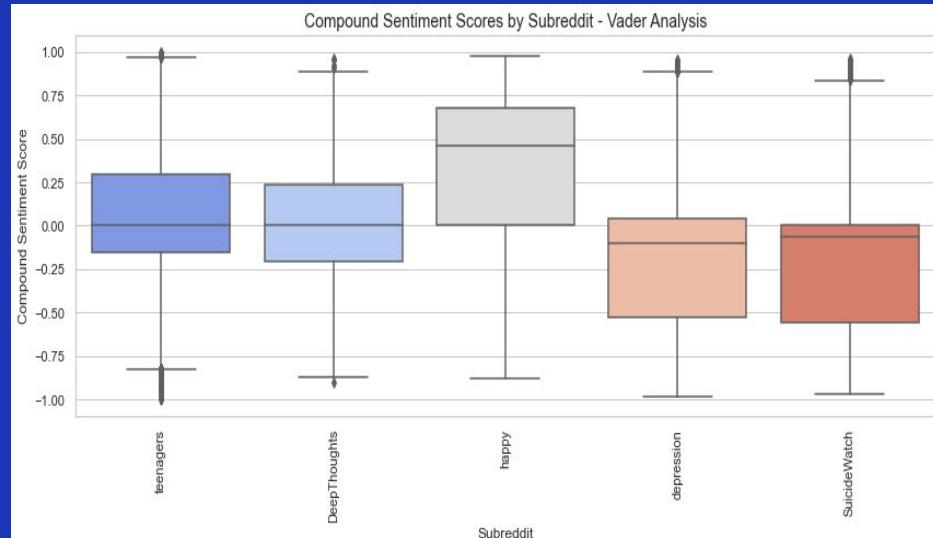
- Scores range from negative (sad) to positive (happy).
- Example:
 - "I feel hopeless." → Negative.(-1 compound score)
 - "Today was a great day!" → Positive.(+1 Compound score)

Why It Matters:

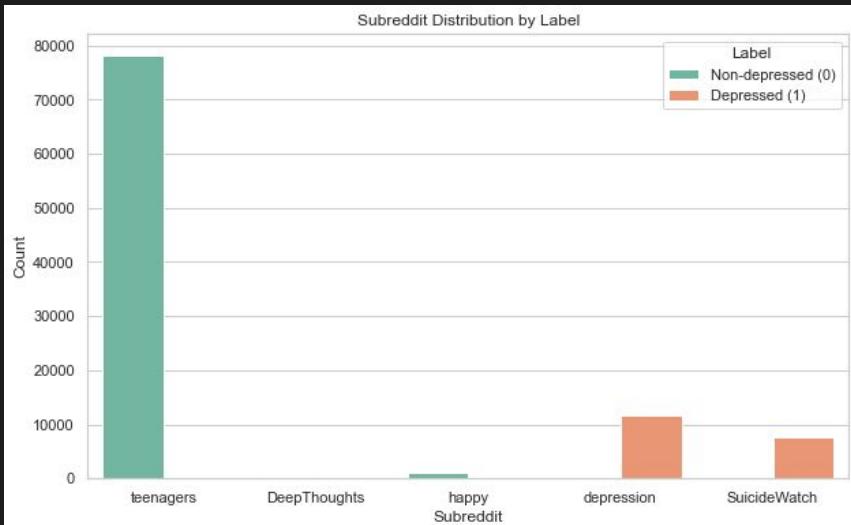
- Helps us understand the dataset in more depth.
- Helps us understand the emotional tone of conversations.
- Can support early detection of mental health challenges.

Key Insights:

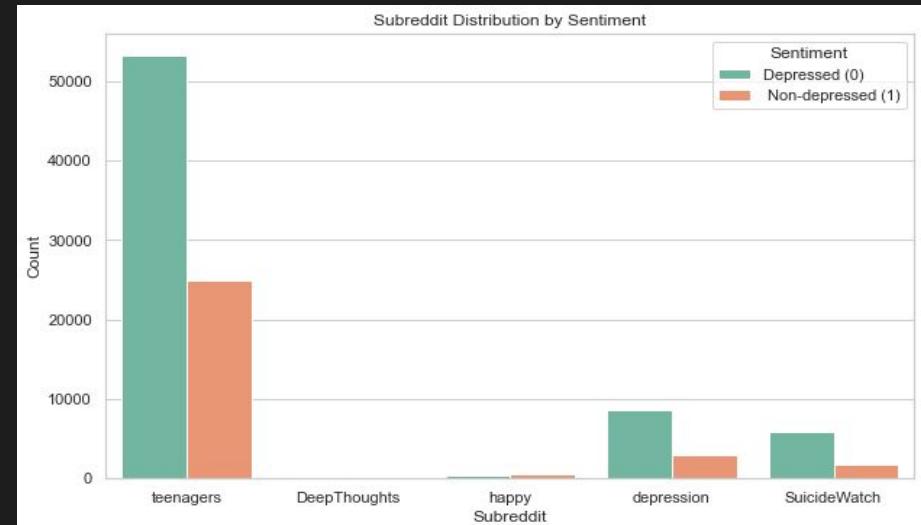
- Posts from depression-related forums have more negative emotions.
- Positive sentiment is more common in forums like *Happy* and *Teenagers*



Sentiment Analysis - (EDA VS VADER)



EDA



VADER

Data Preparation and Modelling



What is Data Preparation?

The process of cleaning, organizing, and transforming data to make it ready for analysis.

Why is it Important?

- Ensures accuracy and reliability of results.
 - Removes irrelevant or incorrect information.
 - Makes data consistent and easy to analyze.

How We Prepared the Data

1. **Cleaning Text:** Removed noise (e.g., links, special characters).
 2. **Filtered out non important words:** Common words (like "and," "the," and "is") that don't add much meaning were removed
 3. **Standardizing Word Forms** - Words were converted to their base form



Outcome:

A clean and structured dataset ready for analysis, enabling us to uncover patterns and insights about mental health trends.



Model Evaluation

Key Metrics Used:

Recall: Measures how well the model identifies depressed posts (important for mental health contexts).

Precision: Ensures the model minimizes false positives.

Accuracy: Overall correctness of predictions.

Logistic Regression with tuned parameters outperformed Random Forest due to its higher recall, ensuring more depressed posts are correctly identified.

Random Forest			
Training accuracy score		0.99	
Testing Accuracy Score	0.91		
Labels	Precision score	Recall Score	F1 score
0 (Non Depressive)	0.94	0.95	0.95
1(Depressive)	0.79	0.73	0.76

Logistic Regression			
Training accuracy score		0.89	
Testing Accuracy Score		0.88	
Hyper Parameter Tuning	0.92		
Labels	Precision score	Recall Score	F1 score
0 (Non Depressive)	0.97	0.88	0.92
1(Depressive)	0.64	0.9	0.75

Model Interpretability



Why Interpretability Matters:

- Transparency in decision-making is essential, especially in sensitive domains like mental health.
- Helps ensure trust and validate that the model identifies relevant patterns.

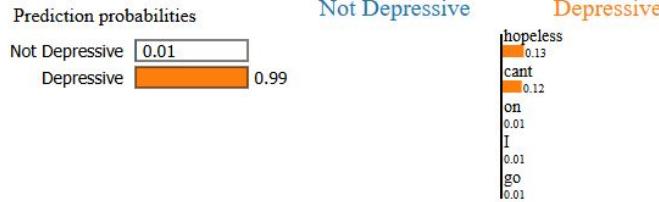
What is LIME?

- Local Interpretable Model-Agnostic Explanations (LIME).
- A technique to explain individual predictions by highlighting features that influenced the outcome.

How LIME Helped:

- For posts classified as "depressed," LIME identified significant keywords or phrases (e.g., "*hopeless*," "*can't go on*").
- For non-depressed posts, it highlighted positive sentiment words (e.g., "*happy*," "*excited*").

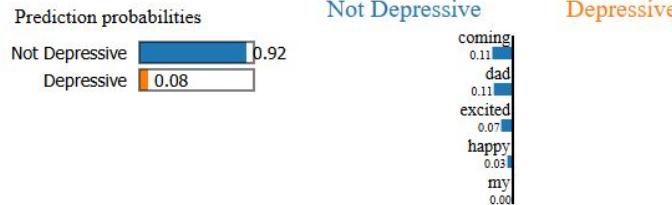
Prediction Insights



Text with highlighted words

It is **hopeless** I **can't** go on

Prediction Insights



Text with highlighted words

I am happy my dad is **coming** tomorrow, **excited**!

Real-Time Sentiment Analysis

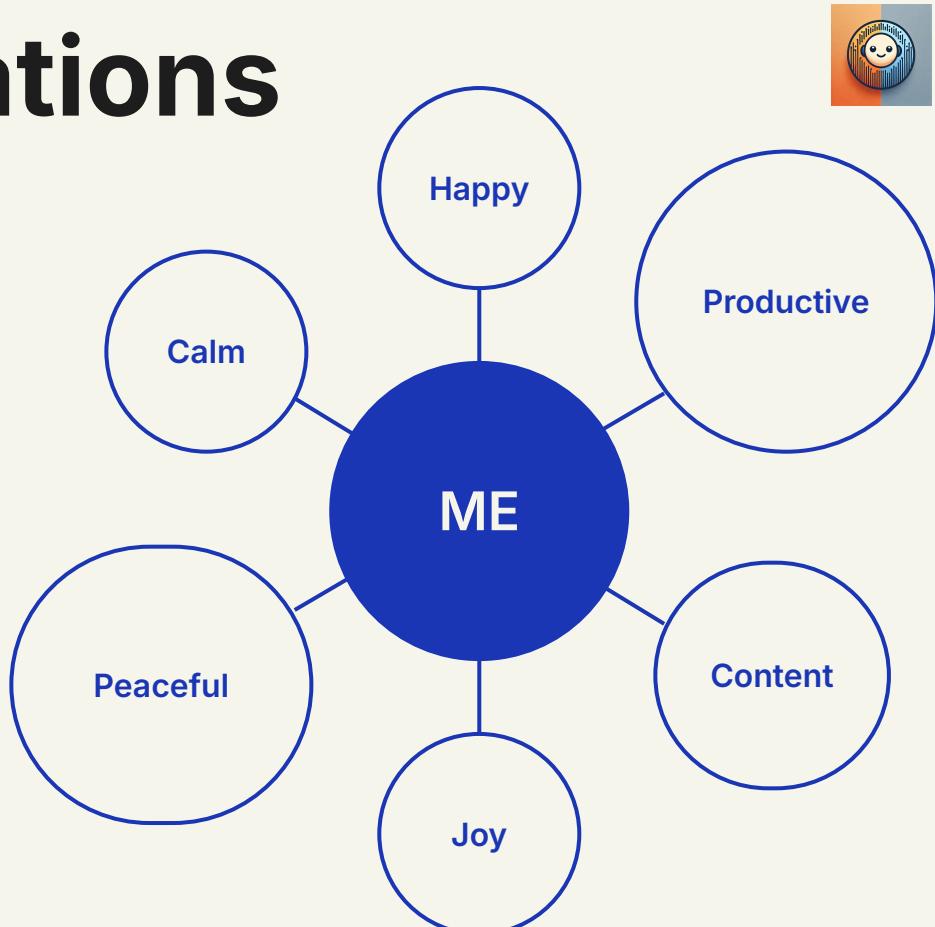
<https://moodlens.streamlit.app>

**Together we can bridge the gap
between technology and
well-being**



Recommendations

- We propose working with medical professionals and researchers to develop a more robust criteria for analysing reditts.
- We propose to improve our model to allow for batch processing which would be super useful for larger platforms.
- We recommend that in future, the app should keep a record/database /reports for future reference to allow for analysis of continuous patterns for specific users and share resources.



Challenges

- While labels were predefined based on subreddits, sentiment scores from VADER didn't always align, making interpretation and alignment of the results challenging
- Imbalanced Dataset - The majority of posts came from non-depressed subreddits, creating a skewed label distribution. This made it challenging to train models effectively on the minority class
- Computational Constraints - Limited computing power for advanced model. For better precision and analysis, the team would require high computing power to process sentiments





Thank you! Questions?



