Project Report: Mental Health Sentiment Analysis Using NLP and Power BI

---

1. Introduction

Mental health is one of the most critical yet overlooked aspects of overall well-being. With the rise of social media and digital communication platforms like Reddit, individuals are increasingly expressing their emotions and mental health struggles online. This project leverages Natural Language Processing (NLP) and sentiment analysis to understand the emotional tone of mental health-related posts on Reddit. The project combines the power of Python (for sentiment analysis) and Power BI (for data visualization) to provide insights into online expressions of depression and overall sentiment distribution.

2. Objective

The primary goals of this project are:

- To analyze Reddit comments related to mental health.

- To perform sentiment analysis and classify emotions as Positive, Neutral, or Negative.
- To identify patterns in depression-related posts.
- To visualize the sentiment and depression trends using Power BI.

3. Tools and Technologies Used

- **Programming Language**: Python
- **Libraries**: Pandas, Numpy, TextBlob, Matplotlib, Seaborn
- **Data Visualization**: Power BI
- **IDE/Notebook**: Jupyter Notebook
- **Dataset Source**: Reddit mental health discussions (depression_dataset_reddit_cleaned.csv)

4. Methodology

Step 1: Data Loading and Initial Exploration

- The dataset was loaded into a Jupyter Notebook.

- Key columns included: clean_text, is_depression (1 = depressed, 0 = not depressed).
- Checked for null values, data types, and general structure.

Step 2: Sentiment Analysis

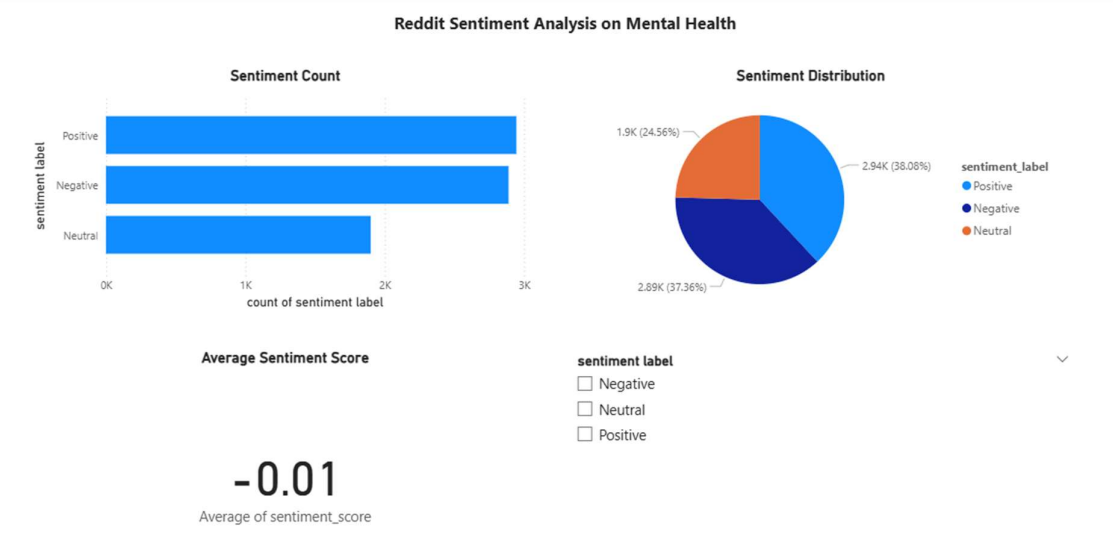
- Used the TextBlob library to compute polarity scores (sentiment_score).
- Based on score ranges, sentiment was categorized into:
  - Positive (score > 0)
  - Neutral (score = 0)
  - Negative (score < 0)
- Added a new column sentiment_label to reflect the classification.

Step 3: Data Cleaning and Exporting

- Ensured all relevant columns were cleaned and formatted.
- Exported the final dataset as sentiment_data.csv for visualization.

Step 4: Power BI Dashboard Creation

- Imported sentiment_data.csv into Power BI.
- Created the following visualizations:



-

5. Key Insights

- A large portion of posts analyzed exhibited negative sentiment, reflecting the emotional weight of the topic.
- Depression-tagged posts had significantly lower sentiment scores on average.
- Positive sentiments were fewer but present, indicating supportive or recovery-based discussions.

- Visual patterns revealed sentiment trends across depressed and non-depressed groups.

6. Challenges and Solutions

- **Challenge**: No labels for sentiment in the original dataset.
    - **Solution**: Applied TextBlob to compute and label sentiment automatically.
- **Challenge**: Ensuring compatibility between Python output and Power BI input.
    - **Solution**: Cleaned and saved the final dataset in .csv format with correct delimiters.

7. Conclusion

This project demonstrates the potential of combining Natural Language Processing and data visualization tools to gain insights into sensitive topics like mental health. By analyzing online conversations, we can better understand public emotional states and the prevalence of depression-related discussions. This type of analysis can be extended to larger datasets and integrated with real-time monitoring tools to

support mental health awareness and intervention efforts.

8. Future Scope

- Integrate more advanced NLP models like VADER or transformers (BERT).
- Create a Streamlit or web-based dashboard for real-time visualization.
- Collaborate with mental health organizations for real-world applications.
