Christian Sanchez

Deeksha Patlolla

Salini Pradhan

Dr. Debnath

Social Networking and Informatics

CSCI 4220/5220

12/8/24

Final Project Documentation

Introduction:

This project is focused on the way emotions play a part in interactions within social networks. Through this we are able to make recommendations and predict engagement. This can create healthier interactions and communities on social networks. Natural Language Processing along with artificial intelligence was used in order to predict behavior and for the classification of emotions. This can be called sentiment analysis as emotions are categorized. Many emotions can be dependent on context which can also make it difficult to classify them.

Motivation:

Social media posts convey emotions to the users who view them. We would like to find out more about the unexplored research on the emotions different types of posts can elicit. We wanted to decode user signals to see how emotions would affect user interaction such as engagement and likes. We are motivated to see how examination of emotions on social networks can be navigated to see the relevance of the demographics.

We also aim to enhance the overall health of social networks. We can filter out hate and harmful content to fill social networks with more positivity. This is something platforms like social media platforms have a difficult time with. Real time moderation can be done through artificial intelligence that can process and filter posts and data. Overall we can aim to build a healthier environment for users.

Using this type of sentiment analysis we can also improve recommendation systems. As of now, many recommendation systems seek out user history and what is trending in order to recommend content. With emotions being involved in the recommendation system, users will be able to have deeper connections with the posts they see. For example, if a user sees content that is overall inspiring, future recommended content would also have an inspiring theme.

Description of Dataset:

- The GoEmotions dataset has 27 emotion annotations
- There is an emotional label with each text
- Data is drawn from a diverse amount of content
- Dataset contains emotional categories such as positive and negative emotions.
- The dataset is versatile for NLP.
- The represented emotion from the text is shown in 0 and 1. 0 representing that the emotion is not present in sentiment analysis, and 1 showing that the text does cause that emotion.

Methods of implementation:

1. Data Collection and Preprocessing:

- **Dataset Utilization:** The GoEmotions dataset from Kaggle, containing text data annotated with various emotions, serves as the primary data source. The dataset is loaded into a Pandas DataFrame for manipulation.
- **Preprocessing Steps:**
    - **Text Cleaning:** Initial cleaning steps involve removing formatting issues and standardizing the text.
    - **Feature Engineering:** New features such as text length and sentiment scores are computed. Text length serves as a proxy for user engagement, and sentiment scores are derived using a predefined emotion-to-sentiment mapping.

2. Sentiment Analysis:

- **Emotion Mapping:** Each emotion in the dataset is associated with a sentiment score (positive, negative, neutral), facilitating a quantitative analysis of emotional content.
- **Polarity Calculation:** The sentiment polarity of each text is determined using the TextBlob library, providing an additional layer of sentiment analysis.

3. Network Analysis:

- **Graph Construction:** An interaction graph is built using NetworkX, where nodes represent users and edges represent emotional interactions based on sentiment scores and text attributes.
- **Subgraph Selection:** For efficient computation, a subgraph is created, focusing on a subset of the network, which is used for further analysis.

4. Centrality Measures:

- **Computational Metrics:** Degree and closeness centrality metrics are calculated to identify influential nodes within the subgraph, which are hypothesized to play a crucial role in spreading or dampening emotional content.

5. Predictive Modeling:

- **Feature Selection:** Features for the predictive model include centrality measures, sentiment scores, and other attributes derived from the text and user interactions.
- **Model Training and Testing:** A RandomForestRegressor is employed to predict the engagement level (operationalized as text length), with the dataset split into training and testing subsets for validation.
- **Model Optimization:** Parameters such as the number of trees and tree depth are tuned to balance performance and computational efficiency.

6. Model Evaluation:

- **Performance Metrics:** The model's effectiveness is assessed using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), providing insights into the accuracy of engagement predictions.

7. Visualization:

- **Network Visualization**: The interaction network is visualized to illustrate the connections and flow of emotional content among users, using node sizes to represent degrees of engagement.

8. Tools and Technologies:

- **Programming Languages and Libraries:** Python is used for all computational tasks, leveraging libraries such as Pandas for data handling, NetworkX for network analysis, and Matplotlib for visualization.

- **Development Environment:** The code is developed and tested in a local environment, with considerations for scalability and potential deployment in cloud platforms for larger datasets.

This comprehensive approach not only fulfills the research objectives of understanding and predicting user engagement based on emotional interactions but also sets the stage for future enhancements such as real-time data processing and more complex modeling techniques. This methodical implementation ensures that the findings can significantly impact community moderation and content recommendation systems on social media platforms.

Results and Discussion:

Model Performance

- **Training and Evaluation Speed:** The model training and subsequent evaluation were remarkably swift, with the entire process completed in just about 0.28 seconds, highlighting the efficiency of the implementation.
- **Accuracy of Predictions:**
    - **Mean Squared Error (MSE):** The computed MSE was 898.4513. This indicates the average squared difference between the estimated values predicted by the model and what was actually observed.
    - **Root Mean Squared Error (RMSE):** At 29.97, the RMSE provides a direct measure of the model's prediction error in the same units as the engagement level (text length). This value suggests that the model predictions are generally within 30 characters of the actual text lengths.

Visualization Insights

- **Network Dynamics:** The generated network graph visually depicts the complex interplay of user interactions and emotional exchanges. Larger nodes, representing users with more interactions, suggest varying degrees of influence or activity within the network.
- **Cluster Formation:** The graph also shows how users cluster around specific nodes, which could indicate shared interests or emotional responses. Such patterns are crucial for understanding community structure and dynamics on social media platforms.

Discussion

- **Implications of Findings:**

  The results underscore the potential of using machine learning to predict user engagement based on emotional content. However, the RMSE points to a variability in prediction accuracy that could stem from the inherent complexity and subtlety of emotional expressions in text.

  The efficient processing time suggests that the model is computationally feasible for larger datasets or real-time applications, making it suitable for deployment in dynamic environments like social media platforms.

- **Strategic Insights for Platform Enhancement:**

  By pinpointing influential users and understanding community structures through network analysis, social media platforms can tailor their content delivery and moderation strategies more effectively.

  The insights gained from emotion analysis can guide the development of algorithms that personalize content based on emotional appeal, enhancing user experience and engagement.

- **Limitations and Future Work:**

The current model could be enhanced by incorporating more nuanced sentiment analysis techniques that consider contextual and cultural variations in language use.

Further research could explore the integration of real-time data processing to adaptively predict and respond to changes in user behavior and platform dynamics.

```
○ deeks@deekshas-MacBook-Pro ~ % /usr/local/bin/python3 /Users/deeks/Desktop/SNI/projcodefile.py
  Starting model training...
  Model training completed in 0.28 seconds
  Evaluating the model...
  Mean Squared Error (MSE): 898.4513512010201
  Root Mean Squared Error (RMSE): 29.97417807381914
  Model evaluation completed in 0.02 seconds
```

Figure 1: Terminal output showing rapid training and evaluation of a Python script, detailing model performance metrics including Mean Squared Error (MSE) and Root Mean Squared Error (RMSE).

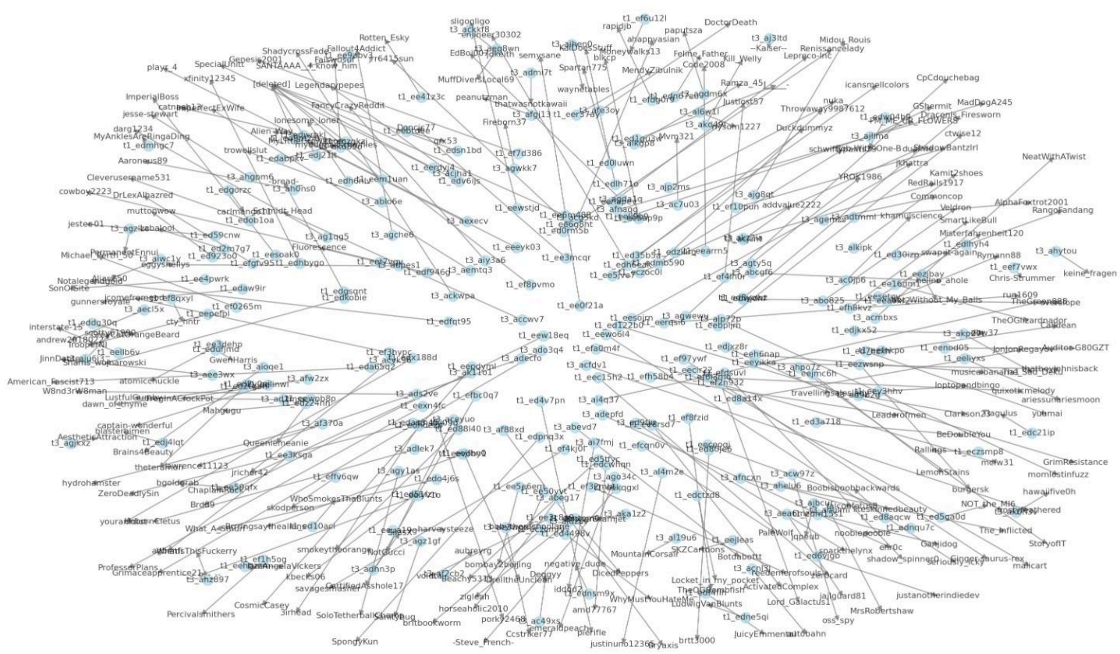

Figure 2: A complex network visualization showing the interconnectivity and relationship dynamics among users based on their emotional interactions within a social media platform.

Conclusion

The study's findings are promising for the application of sentiment analysis in predicting social media engagement. Future iterations of the research could focus on refining the emotional accuracy of the models used and expanding the scalability of the network analysis to encompass larger, more dynamic datasets. This could potentially revolutionize how content is curated and moderated, leading to more resonant and engaging social media environments.