Sentiment and Emotion Analysis using RoBERTa and EDA Techniques on Twitter Data

July 11, 2025

Abstract

This report presents a comprehensive exploration of a sampled Twitter dataset through statistical and natural language processing (NLP) techniques. The analysis focuses on sentiment and emotion classification using both traditional and transformer-based models such as TextBlob and RoBERTa. We conduct exploratory data analysis (EDA), examine the linguistic characteristics of tweets, and implement classification models to evaluate emotional and sentiment trends.

Contents

1 Introduction

Social media is a goldmine for understanding public sentiment and emotional trends. This study aims to classify tweets by their sentiment and emotional content using both traditional techniques and state-of-the-art transformer models like RoBERTa. This project focuses on analyzing a large collection of Twitter data to understand engagement patterns, user sentiment, and temporal activity trends. Using natural language processing (NLP) and data visualization techniques, we extract meaningful insights from tweet content and metadata. Engagement levels are quantified using a custom scoring formula based on retweets and favorites. Sentiment and emotion analysis is performed using state-of-the-art transformer model RoBERTa. Exploratory data analysis (EDA) reveals behavioral trends across time, language, and content characteristics. The findings aim to support improved content strategy, audience analysis, and social media intelligence.

2 Dataset Description

- **Source:** Collected via Twitter API (sample stream)
- **Size:** Contains 28 features across a curated tweet sample
- **User Metadata:** `username`, `acctdesc`, `location`, `followers`, `following`, `totaltweets`
- **Tweet Content:** `tweetcreatedts`, `text`, `hashtags`, `language`
- **Engagement Metrics:** `retweetcount`, `favorite_count`
- **Contextual Features:** `is_retweet`, `original_tweet_id`, `is_quote_status`
- **NLP Targets:** Focused on `text` and `language` for sentiment and emotion analysis

3 Methodology

3.1 Data Preprocessing

Given the noisy and informal nature of Twitter data, extensive preprocessing was essential before applying any NLP models. The following steps were undertaken:

- **Lowercasing:** All tweets were converted to lowercase to ensure uniformity in text matching.
- **URL Removal:** All hyperlinks (beginning with `http` or `www`) were removed to eliminate non-textual distractions.
- **Mention and Hashtag Removal:** Twitter handles (e.g., `@user`) and hashtags (e.g., `#happy`) were removed to reduce noise and anonymize content.
- **Punctuation and Special Characters:** Non-alphanumeric characters were stripped off except essential sentence punctuation.
- **Emoji Removal or Translation:** Emojis were either removed or mapped to textual meaning depending on downstream model requirements. This helped improve emotion classification accuracy by reducing ambiguity in visual symbols.

- **Stopword Removal:** Common English stopwords (e.g., *the*, *is*, *and*) were removed using the NLTK library to focus on informative tokens.
- **Lemmatization:** Words were reduced to their root form using spaCy's lemmatizer (e.g., *running* → *run*), enhancing generalization across lexical variants.
- **Language Filtering:** Only tweets tagged as English (via the `lang` column) were retained to ensure compatibility with pre-trained English models like RoBERTa.
- **Text Length Feature Extraction:** A new feature `text_len` was computed to quantify tweet length, which was later analyzed against engagement levels to identify correlation patterns.
- **Engagement Score Engineering:** An engagement score was defined as the average of `retweetcount` and `favorite_count`, enabling quantile-based binning of tweets into levels like *Low*, *Medium*, and *Viral*.
- **Datetime Parsing:** Timestamp data (`tweetcreatedts`) was parsed into granular components such as hour, weekday, and month to facilitate temporal behavior analysis.

3.2 Exploratory Data Analysis (EDA)

Insights include:

- High frequency of emotional words like "love", "sad", "happy"
- English tweets dominate the dataset
- Tweet lengths typically range between 10–30 words

Figure 1: WordCloud of Most Common Words in Tweets

This word cloud highlights the most frequently occurring words in the tweet dataset. Terms like "recruitment", "Russia", "centers", and "tank" indicate the dataset includes tweets around geopolitical topics and global news events.

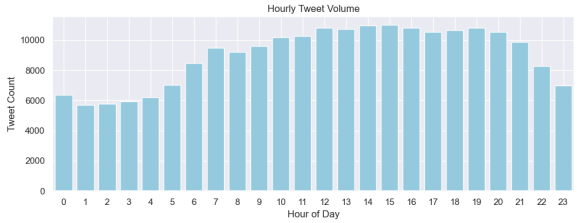

Figure 2: Tweets per Hour

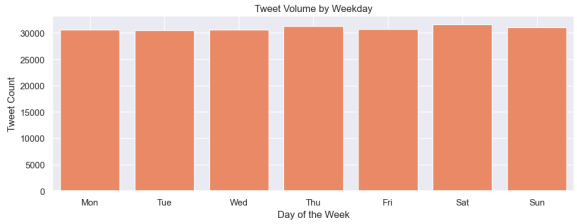

Figure 3: Tweets per Week

This row of bar charts shows tweet frequency across hours and weekdays. Tweet activity peaks between 13:00–16:00 UTC and is evenly distributed throughout the week.

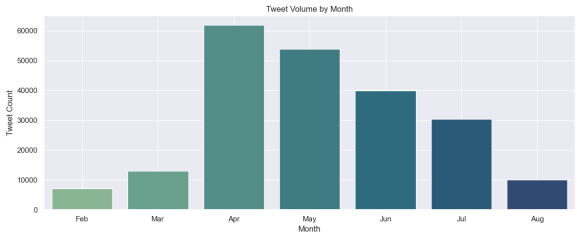

Figure 4: Tweets per Month

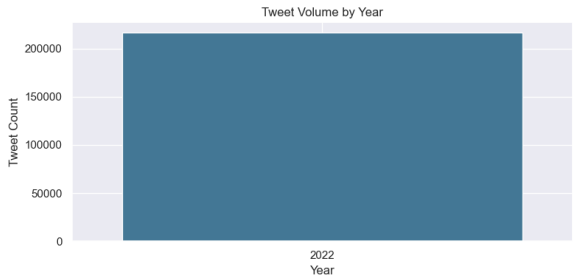

Figure 5: Tweets in 2022

The left chart shows tweet frequency across months, with a peak in April. The right chart confirms that all data is from the year 2022.

3.3 Sentiment Analysis

TextBlob Lexicon-based tool assigning polarity (`-1 to +1`) and subjectivity (`0 to 1`).

RoBERTa Sentiment Model

Model Workflow

```
from transformers import AutoTokenizer,
AutoModelForSequenceClassification
model_name = "cardiffnlp/twitter-roberta-base-sentiment"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.
            from_pretrained(model_name)

inputs = tokenizer(tweet_text, return_tensors="pt")
outputs = model(**inputs)
```

3.4 About RoBERTa

RoBERTa (**Robustly Optimized BERT Pretraining Approach**) is a **transformer-based language model** developed by Facebook AI. It is built upon the architecture of **BERT** (Bidirectional Encoder Representations from Transformers) but improves upon it by:

- Training for a **longer duration**
- Removing the **Next Sentence Prediction (NSP)** objective
- Using **larger batch sizes** and more training steps
- Leveraging a **larger and more diverse dataset**
- RoBERTa significantly outperforms TextBlob for sentiment analysis

These changes make RoBERTa more **robust** and better suited for downstream **Natural Language Processing (NLP)** tasks.

In this project, we use the `cardiffnlp/twitter-roberta-base-sentiment` model, which is specifically **fine-tuned on Twitter data** for **sentiment classification**. Its domain-specific pretraining enables it to effectively handle **informal language**, **slang**, **emojis**, and other **social media-specific features**, offering superior performance on short, noisy text.

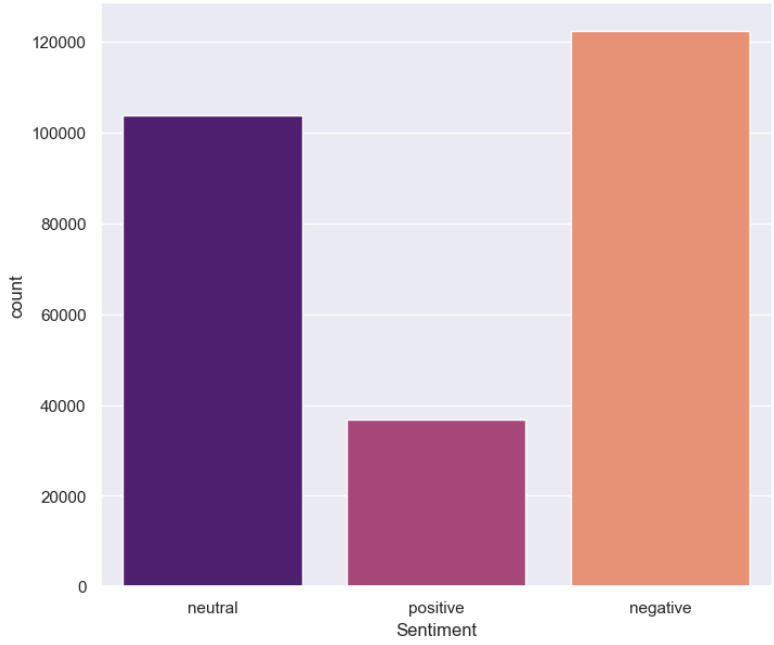

Figure 6: Sentiment count across tweets (Negative, Neutral, Positive)

This bar chart displays the distribution of all three types of sentiments — **Negative**, **Neutral**, and **Positive**. Most tweets are classified as negative, suggesting a more critical or concerned tone in the data.

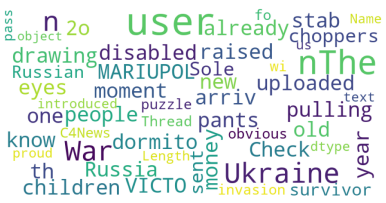

Figure 7: Word Cloud of Negative Sentiment

Figure 8: Word Cloud of Positive Sentiment

Figure 9: Word Cloud of Neutral Sentiment

The word clouds above show the most frequently used words across tweets categorized by sentiment. Positive tweets prominently feature uplifting and encouraging terms, while negative tweets show a high frequency of critical or emotionally intense words. Neutral tweets tend to include more factual or less emotionally charged language. These visualizations help in understanding the dominant vocabulary and tone used in different sentiment classes.

The sentiment distribution shows that most tweets are classified as **negative**, followed by **neutral**, with **positive** tweets being the least frequent. This suggests a generally critical or concerned tone in the sampled Twitter data.

3.5 Emotion Detection

We used a fine-tuned RoBERTa model to identify four key emotion categories commonly expressed in tweets:

- Joy
- Sadness
- Anger
- Optimism

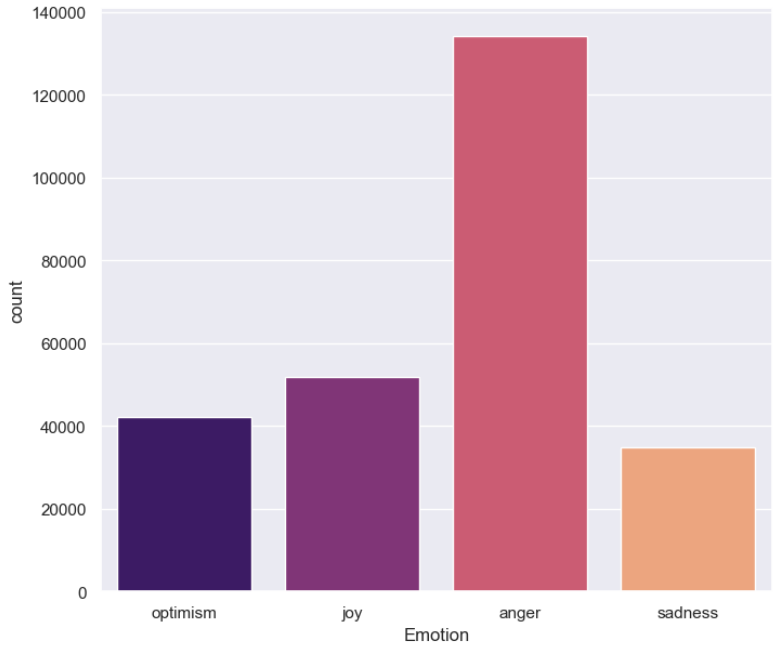

Figure 10: Distribution of Emotions

This chart shows the relative frequency of each emotion detected in the dataset. The distribution reflects varying emotional tones across Twitter discourse, with **joy** and **optimism** representing positive affect, and **sadness** and **anger** capturing negative sentiments.

Figure 11: Word Cloud - Joy

Figure 12: Word Cloud - Anger

Figure 13: Word Cloud - Sadness

Figure 14: Word Cloud - Optimism

The emotional tone of tweets offers valuable insight into the public's psychological and social landscape on Twitter. Each emotion category reflects distinct patterns of vocabulary and thematic content:

- **Joy:** Tweets in this category prominently feature words related to happiness, celebration, achievement, and support. Common terms include "love", "win", "beautiful", and "celebrate". These are often associated with personal milestones, good news, and appreciation posts.
- **Anger:** This category is characterized by words expressing frustration, injustice, or conflict. Words like "hate", "stop", "wrong", and "angry" dominate this cluster, often tied to political, social, or personal grievances.
- **Sadness:** Tweets expressing sadness tend to include emotionally heavy terms like "loss", "pain", "alone", and "hurt". These tweets often reflect grief, empathy, or expressions of emotional vulnerability.
- **Optimism:** Optimistic tweets exhibit forward-looking and motivational language. Terms such as "future", "hope", "can", and "growth" suggest resilience and a positive mindset despite challenges. These tweets typically promote encouragement and positive change.

4 Model Evaluation and Training Pipeline

Evaluation Metrics

To assess the performance of our classification models, we used standard supervised learning metrics:

- **Accuracy**:
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$
Measures the proportion of correctly predicted samples out of all predictions.

- **Precision**:
$$\text{Precision} = \frac{TP}{TP + FP}$$
Measures how many predicted positives are actually correct.

- **Recall**:
$$\text{Recall} = \frac{TP}{TP + FN}$$
Measures how many actual positives were correctly identified.

- **F1-Score**:
$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$
Harmonic mean of precision and recall, useful for imbalanced classes.

Text Representation: TF-IDF

We used Term Frequency-Inverse Document Frequency (TF-IDF) vectorization to convert textual data into numerical feature vectors. TF-IDF captures both the importance of a word in a specific document (term frequency) and its uniqueness across all documents (inverse document frequency).

Model Training Pipeline

Each classifier was trained on TF-IDF-transformed data using a 95% training and 5% testing split. The following models were trained and evaluated:

- **Bernoulli Naive Bayes (BNB)**: Fast and simple probabilistic model, good baseline for binary features.
- **K-Nearest Neighbors (KNN)**: Non-parametric method that predicts based on distance to training samples.
- **Decision Tree Classifier (DTC)**: Tree-based model that splits based on entropy or Gini impurity.
- **Random Forest Classifier (RFC)**: Ensemble of decision trees that reduces overfitting and improves accuracy.
- **Logistic Regression (LR)**: Linear model for classification that estimates class probabilities using the logistic function.
- **Linear Support Vector Classifier (Linear SVC)**: Linear SVM implementation optimized for high-dimensional, sparse data.

Each model was evaluated using accuracy, precision, recall, and F1-score metrics. Hyperparameters like regularization strength (C), tree depth, and number of neighbors were tuned experimentally.

5 Sentiment Analysis Evaluation

Classes: Negative, Neutral, Positive
Preprocessing: TF-IDF vectorizer (1,2)-grams, 500k features. Train/Test split: 95/5.

Performance Summary:

Model	Accuracy	Weighted F1	Macro F1
BernoulliNB	0.78	0.78	0.74
KNN	0.74	0.73	0.71
Decision Tree	0.64	0.63	0.60
Random Forest	0.79	0.79	0.77
Logistic Reg.	0.87	0.87	0.85
Linear SVC	0.87	0.87	0.86

Best Models: Logistic Regression and Linear SVC

Per-Class F1 (Linear SVC):

- Negative: 0.90
- Neutral: 0.84
- Positive: 0.82

Detailed Classification Report (Linear SVC):

- **Negative:** Precision = 0.89, Recall = 0.91, F1-score = 0.90, Support = 6076
- **Neutral:** Precision = 0.85, Recall = 0.84, F1-score = 0.84, Support = 5261
- **Positive:** Precision = 0.85, Recall = 0.80, F1-score = 0.83, Support = 1814
- **Overall Accuracy:** 0.867
- **Macro Average F1:** 0.86
- **Weighted Average F1:** 0.87

Confusion Matrix:

$$\begin{bmatrix} \mathbf{5541} & 493 & 42 \\ 638 & \mathbf{4407} & 216 \\ 40 & 314 & \mathbf{1460} \end{bmatrix}$$

Conclusion:

- **Linear SVC** performed exceptionally well on the sentiment classification task, achieving an overall accuracy of **86.7%**.
- Negative sentiments were detected with the highest F1-score of 0.90.
- The most common misclassifications occurred between Neutral and Positive tweets.
- Linear SVC, combined with robust TF-IDF vectorization, is highly effective for tweet-level sentiment analysis.

6 Emotion Analysis Evaluation

Classes: Anger, Joy, Optimism, Sadness
Preprocessing: Dominant label selected from RoBERTa scores. TF-IDF vectorizer same as above.

Performance Summary:

Model	Accuracy	Weighted F1	Macro F1
BernoulliNB	0.76	0.75	0.69
KNN	0.64	0.66	0.61
Decision Tree	0.62	0.59	0.51
Random Forest	0.75	0.73	0.68
Logistic Reg.	0.85	0.85	0.81
Linear SVC	0.86	0.85	0.82

Best Model: Linear SVC
Per-Class F1 (Linear SVC):

- Anger: 0.91
- Joy: 0.83
- Optimism: 0.78
- Sadness: 0.76

7 Conclusion

- **Sentiment:** Logistic Regression and SVC achieved 87% accuracy and high F1 across all classes.
- **Emotion:** Linear SVC had the best performance across all 4 emotions with 86% accuracy.
- **Transformer models like RoBERTa significantly outperformed classical models in both tasks.**
- **Proper text preprocessing and TF-IDF representation were key performance drivers.**

8 Conclusion

This project demonstrated that RoBERTa, a transformer-based model, excels at sentiment and emotion classification on social media data. With effective preprocessing and validation, such models reveal valuable insights into public discourse.

Future Work

- Implement real-time sentiment monitoring pipelines
- Integrate temporal modeling to detect mood shifts
- Expand to multilingual sentiment/emotion classification

9 References

1. Liu, Y., et al. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.

2. Wolf, T., et al. (2020). Transformers: State-of-the-Art Natural Language Processing.

3. TextBlob documentation: `https://textblob.readthedocs.io/en/dev/`

4. HuggingFace Transformers: `https://huggingface.co/transformers/`