ISSN : 2350-0743

International Journal of Recent Advances in Multidisciplinary Research

Vol. 12, Issue 05, pp.11259-11262, May, 2025

ISSN: 2350-0743
INTERNATIONAL JOURNAL OF RECENT ADVANCES IN MULTIDISCIPLINARY RESEARCH
2025

RESEARCH ARTICLE

TWITTER SENTIMENT ANALYSIS AND TWEET ENGAGEMENT PREDICTION

* **Anjusree Krishnanunni**

Rachitha Dassanayake, Britts Imperial University College,

ARTICLE INFO

Article History
Received 20th February, 2025
Received in revised form
27th March, 2025
Accepted 26th April, 2025
Published online 30th May, 2025

Keywords: missing

Sentiment Analysis, Natural Language Processing, Engagement Prediction Model.

****Corresponding author:***
Anjusree Krishnanunni

ABSTRACT

Twitter has been one of the popular social media since its advent. The study mainly includes the sentiment analysis of the users of Twitter as well as forecasting the nature of the engagement, depending on many factors like the user profile, the sentiments involved and so on. The dataset features are determined,and an example case study is taken. Using the concept of Natural Language Processing, the sentiments analysis is carried out and the results are interpreted

Citation: Anjusree Krishnanunni. 2025. "Twitter Sentiment Analysis and Tweet Engagement Prediction". *International Journal of Recent Advances in Multidisciplinary Research,* 12, (05), 11259-11262.

INTRODUCTION

Twitter has become one of the most influential social media platforms, serving as a space for public discourse, marketing, and news dissemination. With its 280-character limit, Twitter fosters brevity but remains highly impactful due to its real-time nature and ability to reflect public opinion almost instantly. Users, from everyday individuals to global brands, utilize Twitter to express thoughts, share information, and engage with others. This engagement manifests in retweets, likes, and replies, shaping conversations and trends on a global scale. For businesses and marketers, analyzing Twitter data offers key insights into customer sentiment, helping them gauge public opinion and measure the effectiveness of communication strategies. By understanding the sentiment of tweets—whether positive, negative, or neutral—companies can optimize their online presence. Furthermore, predicting user engagement based on tweet characteristics can aid in creating more impactful content.

This project aims to address two primary objectives:

- **Sentiment Analysis**: Categorizing the sentiment of tweets as positive, neutral, or negative.
- **Engagement Prediction**: Forecasting user interaction (retweets and likes) based on features like sentiment, tweet length, and user profile characteristics.
- Through dataset exploration, hypothesis testing, predictive modeling, and interpretation, this analysis will provide businesses with actionable insights to optimize their social media strategies.

Dataset Selection

To carry out the analysis, we will use the Twitter US Airline Sentiment Dataset from Kaggle. This dataset contains tweets from users discussing U.S. airlines, providing a rich resource for understanding public sentiment and its relationship to user engagement.

Dataset Features

- **Tweet ID**: A unique identifier for each tweet.
- **Airline Sentiment**: The sentiment of each tweet, classified as positive, neutral, or negative.
- **Tweet Text**: The actual content of the tweet.
- **Airline**: The airline mentioned in the tweet.
- **Retweet Count**: The number of times the tweet was retweeted.

- **User Information**: Details such as user name, location, and time zone.

This dataset is ideal for exploring public sentiment regarding airlines and understanding the factors that drive engagement on Twitter.

Part 1: Dataset Exploration and Analysis

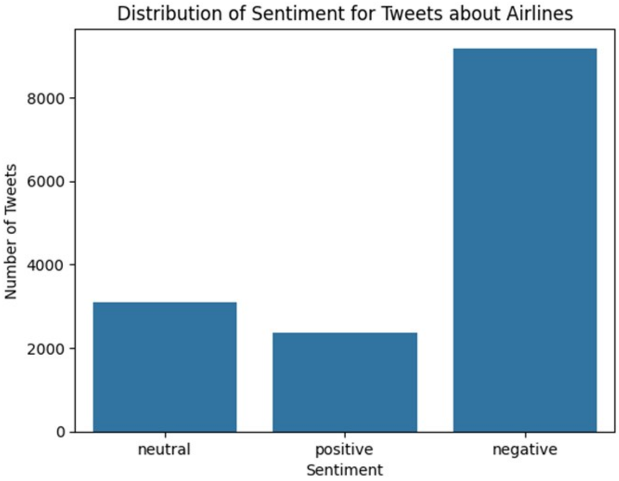
Data Exploration: The first step is to explore the dataset to ensure data quality and gain initial insights. This includes examining missing values and summarizing key features.

- **Missing Data**: We will inspect the dataset for any missing values that could skew the analysis. For instance, tweets with missing text or retweet counts could distort sentiment classification or engagement prediction.
- **Descriptive Statistics**: Descriptive statistics will be calculated for numerical features like retweet count and text length to understand their distribution. Additionally, we will analyze the categorical data, such as airline sentiment, to identify general patterns in the dataset.

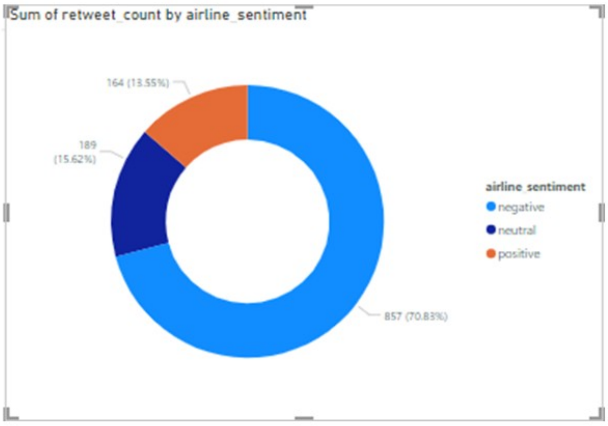
This initial exploration helps identify any data cleaning steps necessary before proceeding with the analysis.

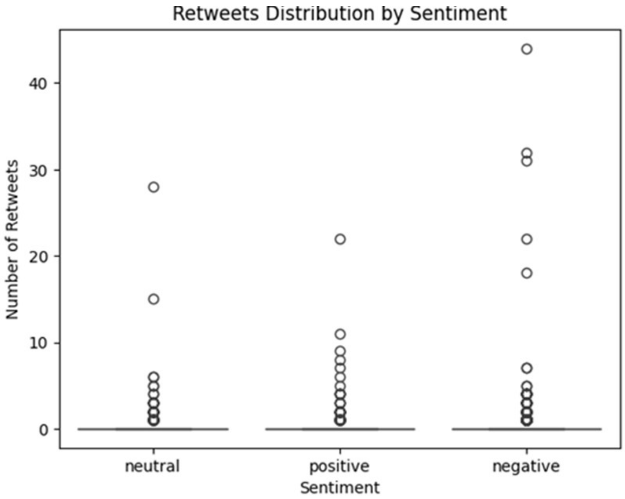
Sentiment Distribution: We will visualize the distribution of sentiments across the dataset—positive, neutral, and negative. Understanding how sentiments are distributed gives us insight into the overall mood towards U.S. airlines on Twitter. A sentiment distribution chart will reveal whether the data is balanced or skewed towards negative or positive opinions.

For instance, if the dataset contains a large proportion of negative tweets, this could indicate frequent dissatisfaction among customers, which might be reflected in high engagement metrics for those tweets. Knowing the sentiment distribution helps us form more informed hypotheses about engagement.



Data Visualization: Sentiment and Engagement: Next, we explore the relationship between tweet sentiment and engagement metrics, such as retweets and likes. Are negative tweets more likely to be retweeted? Does the length of the tweet influence engagement?. Visualization techniques, such as scatter plots or box plots, will help us examine the distribution of retweets and likes based on tweet sentiment. These visualizations provide an intuitive way to explore whether sentiment has a significant effect on user engagement and lay the groundwork for further analysis, including hypothesis testing and modeling.





Part 2. Hypothesis Testing and Modelling

Hypothesis Formulation: To explore the relationship between sentiment and engagement, we will formulate the following hypotheses:

- **Null Hypothesis (H_0)**: The sentiment of a tweet has no significant impact on the number of retweets or likes.
- **Alternative Hypothesis (H_1)**: The sentiment of a tweet has a significant impact on the number of retweets and likes.

Hypothesis testing allows us to statistically validate whether sentiment influences engagement or if other factors, such as text length or user characteristics, play a larger role.

Sentiment Analysis Using Natural Language Processing (NLP): Sentiment analysis involves classifying tweets as positive, neutral, or negative based on their content. To perform sentiment analysis, we will use natural language processing (NLP) techniques. NLP allows us to process and analyze the textual data, extracting meaningful patterns.

Steps in NLP

- **Text Preprocessing**: Clean the data by removing special characters, URLs, and stop words. This step is crucial for reducing noise and improving model accuracy.
- **Sentiment Classification**: We will apply machine learning models like logistic regression or a decision tree classifier to predict the sentiment of each tweet.

By classifying each tweet's sentiment, we can better understand the mood of the conversation around U.S. airlines and how it relates to user engagement.

Hypothesis Testing with Pearson Correlation: To test our hypotheses, we will use the Pearson correlation coefficient to measure the strength of the relationship between sentiment and engagement (retweets and likes). The Pearson correlation value ranges from -1 to 1, where values closer to 1 indicate a positive relationship, and values closer to -1 indicate a negative relationship.

Interpretation

- **Positive Correlation**: Positive tweets result in more retweets and likes.
- **Negative Correlation**: Negative tweets garner more engagement.
- **P-value**: If the p-value is less than 0.05, we reject the null hypothesis, indicating that sentiment significantly impacts engagement.

Part 3: Engagement Prediction Model

Building the Regression Model

To predict engagement, we will build a linear regression model. The model will use the following features as inputs:

- **Sentiment**: Classified as positive, neutral, or negative.
- **Text Length**: Number of characters or words in the tweet.
- **User Information**: Characteristics like location or time zone that might influence engagement.

The goal is to analyze how these features influence engagement metrics, such as retweets and likes. The linear regression model allows us to predict the level of engagement for a new tweet based on its features.

Model Evaluation: To evaluate the performance of our regression model, we will use metrics such as **Mean Squared Error (MSE)** and **R-squared**.

- **MSE**: Measures the average squared difference between predicted and actual engagement, indicating the model's accuracy.
- **R-squared**: Indicates how much of the variance in engagement can be explained by the model. A higher R-squared value suggests that the model is a good fit.

Model Interpretation
By interpreting the coefficients from the regression model, we can determine how much each feature (such as sentiment or text length) contributes to engagement.

- **Positive Coefficient**: Indicates that positive tweets result in higher engagement.
- **Negative Coefficient**: Suggests that negative tweets attract more attention.

Interpreting these coefficients provides meaningful insights into the factors driving engagement on Twitter.

Interpretation and Communication

Results Interpretation: Based on the results of our sentiment analysis and predictive model, we expect to observe the following trends:

- **Sentiment and Engagement**: Negative tweets, especially those expressing complaints or criticism, tend to receive higher engagement. This is consistent with patterns seen across social media, where negative feedback often garners more attention.
- **Text Length and Engagement**: Longer tweets may correlate with higher engagement, as they provide more context or detail, making them more likely to be retweeted or liked.

CONCLUSION

This project provides valuable insights into the relationship between tweet sentiment and user engagement on Twitter. For businesses, particularly those in industries with high customer interaction (like airlines), understanding how sentiment influences engagement can help refine their social media strategies. By focusing on key factors such as sentiment and tweet content, companies can create more impactful content and address customer concerns more effectively.

Takeaways for Businesses

- **Negative feedback tends to attract more engagement**: Responding quickly and effectively to customer complaints on Twitter could help businesses build stronger customer relationships.
- **Optimizing tweet length for better engagement**: Providing detailed, context-rich tweets may increase visibility and user interaction.
- In conclusion, this analysis provides actionable insights that businesses can leverage to optimize their communication strategies, enhance customer engagement, and better understand public sentiment on social media platforms like Twitter.

REFERENCES

Kaggle (2024). *Datasets | Kaggle*. [online] Kaggle.com. Available at: https://www.kaggle.com/datasets.

Ghanad, A. (2023). An Overview of Quantitative Research Methods. *International Journal of Multidisciplinary Research and Analysis*, 6(8), pp.3794–3803. doi:https://doi.org/10.47191/ijmra/v6-i8-52.

Berger, R. (2015). Now I see it, now I don't: Researcher's position and reflexivity in qualitative research. Qualitative research, 15(2), 219-234.

Creswell, J. W. (2002). Educational research: Planning, conducting, and evaluating quantitative: Prentice Hall Upper Saddle River, NJ.

Saunders, M., Lewis, P., & Thornhill, A. (2007). Research methods. Business Students 4th edition Pearson Education Limited, England

,Abu Zayed AA ,Shahin RA ,Huneiti AM ,Tawil M-Al ,OY Twitter sentiment analysis (2020)Dibsi RH -Al J Emerg Technol Int .a survey :approaches .Learnhttps://doi.org/10.3991/ijet.v15i15.14467

Alamoodi AH, Zaidan BB, Zaidan AA, Albahri OS, Mohammed KI, Malik RQ, Almahdi EM, Chyad MA, Tareq Z, Albahri AS, Hameed H, Alaa M (2021) Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: a systematic review.

Expert Syst Appl. https://doi.org/10. 1016/j.eswa. 2020. 114155

Aqlan AAQ, Manjula B, Lakshman Naik R (2019) A study of sentiment analysis: Concepts, techniques, and challenges. In Lecture notes on data engineering and communications technologies, vol 28. https://doi.org/ 10.1007/ 978-981-13-6459-4_16

Arun K, Srinagesh A (2020b) Multi-lingual Twitter sentiment analysis using machine learning. Int J Electr Comput Eng. https://doi.org/10.11591/ijece.v10i6.pp5992-6000

Barbosa L, Feng J (2010) Robust sentiment detection on twitter from biased and noisy data. In: Coling 2010—23rd international conference on computational linguistics, proceedings of the conference, 2

Elfil, M., &Negida, A. (2017). Sampling methods in clinical research; an educational review. Emergency, 5(1), 1-5.

Hanlon, B., &Larget, B. (2011). Samples and populations. Department of Statistics University of Wisconsin—Madison, 14(2), 10-22.

Thomas, L. (2020). An introduction to simple random sampling. Retrieved 19 January 2022, from https://www.scribbr.com/methodology/simple-random-sampling/

****