



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 6, June 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Sentiment Analysis of Twitter: A Comprehensive Review and Comparative Study

Vrushank Chaphadkar, Karan Chipade, Dr, Prakash Kene

PG Student, Dept. of Master of Computer Application, P.E.S Modern College of Engineering, Pune, India

Associate Professor, Dept. of Master of Computer Application, P.E.S Modern College of Engineering, Pune, India

ABSTRACT: Sentiment analysis, also known as opinion mining, has gained significant attention in recent years due to the proliferation of social media platforms and the vast amount of user-generated content. Twitter, being a popular microblogging platform, serves as a rich source of real-time information, making it an ideal platform for sentiment analysis research. This paper presents a comprehensive review and comparative study of sentiment analysis techniques applied to tweets, exploring various approaches, challenges, and advancements in the field. The study aims to provide researchers and practitioners with an overview of the existing methodologies, tools, and datasets, enabling them to make informed decisions while designing sentiment analysis models for tweets.

I. INTRODUCTION

Social media platforms have revolutionized the way people communicate and express their opinions. Twitter, with its real-time and concise nature, has become a prominent platform for users to share their thoughts, emotions, and experiences. This vast volume of user-generated content on Twitter presents an opportunity for sentiment analysis, allowing us to gain valuable insights into public opinion, consumer sentiment, and societal trends.

Sentiment analysis, also known as opinion mining, is a subfield of natural language processing (NLP) that aims to determine the sentiment or emotional tone expressed in a piece of text. It involves analyzing the subjective information conveyed by users and classifying it into positive, negative, or neutral sentiments. Sentiment analysis has numerous applications across various domains, including marketing, brand monitoring, customer feedback analysis, political analysis, and financial markets prediction.

II. LITERATURE SURVEY

1. Article: "Twitter mood predicts the stock market" by Johan Bollen, Huina Mao, and Xiao-Jun Zeng (2011)
This influential study investigates the correlation between Twitter mood and stock market movements, using sentiment analysis to predict stock market trends.
2. Article: "Sentiment analysis of Twitter data" by Alec Go, Richa Bhayani, and Lei Huang (2009), This research paper presents a study on sentiment analysis of Twitter data, introducing a sentiment scoring technique and comparing its performance with other approaches.
3. Article: "Twitter sentiment classification using distant supervision" by Alec Go, Lei Huang, and Richa Bhayani (2009), This paper proposes a distant supervision approach for sentiment classification on Twitter by leveraging emoticons as noisy labels and compares it with supervised learning methods.
4. Article: "Sentiment analysis and subjectivity" by Bo Pang and Lillian Lee (2008)
This influential work explores the subjectivity and sentiment classification problem, including the analysis of sentiment-bearing words and the effectiveness of machine learning algorithms.
5. Article: "Twitter sentiment analysis: The good the bad and the omg!" by Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore (2011)
This study investigates sentiment analysis on Twitter by comparing different classification techniques, lexicon-based methods, and examining the challenges and limitations of sentiment analysis in short texts.

III. RESEARCH METHODOLOGY & ANALYSIS

1.1 Data collection and preprocessing:-

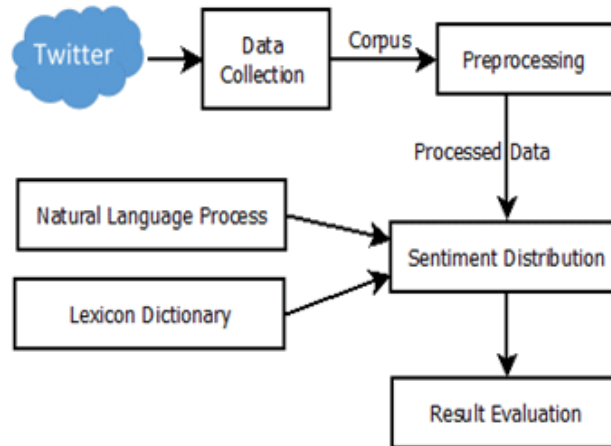


Fig 1: Steps of process of sentiment analysis

The process of sentiment analysis on Twitter involves the following steps:

- 1.Data Collection:** Twitter provides APIs that allow developers to access and collect tweets in real-time or historical data. Researchers can also use publicly available datasets or build custom datasets through manual annotation. The choice of data collection method depends on the research objectives and available resources.
- 2.Preprocessing:** Twitter data requires preprocessing to clean and prepare it for analysis. Common preprocessing steps include tokenization, lowercasing, removing punctuation, stop-word removal, handling URLs, hashtags, mentions, and normalizing informal language, such as slang and abbreviations.
- 3.Sentiment Classification:** The main task in sentiment analysis is to classify tweets into positive, negative, or neutral sentiment categories. Various approaches are used for sentiment classification, including traditional machine learning algorithms (e.g., Naive Bayes, Support Vector Machines), lexicon-based methods, and deep learning models (e.g., Convolutional Neural Networks, Recurrent Neural Networks, Transformer-based models).
- 4.Feature Extraction:** In order to classify sentiment, relevant features need to be extracted from the tweet text. These features can include n-grams, part-of-speech tags, syntactic dependencies, sentiment lexicons, and contextual embeddings. Feature selection techniques may also be applied to identify the most informative features for sentiment classification.
- 5.Model Training and Evaluation:** Sentiment analysis models are trained using labeled datasets, where tweets are manually annotated with sentiment labels. The performance of the models is evaluated using metrics such as accuracy, precision, recall, F1-score, and Area Under the ROC Curve (AUC). Cross-validation and holdout testing are commonly used evaluation techniques.
- 6.Sentiment Analysis Applications:** Sentiment analysis of Twitter data finds applications across various domains. It is widely used in brand monitoring and reputation management to understand customer sentiment towards products and services.

3.2 Traditional Approaches in Sentiment Analysis of Twitter

Traditional approaches in sentiment analysis of Twitter data involve the use of rule-based methods, lexicon-based techniques, and machine learning algorithms. These approaches were widely used before the emergence of deep learning models and still provide valuable insights and baselines for sentiment analysis research. Here are some traditional approaches commonly employed in sentiment analysis of Twitter:

1.Rule-Based Methods: Rule-based approaches rely on handcrafted rules and linguistic patterns to identify sentiment in tweets. These rules can be based on the presence of specific keywords, patterns of negation, intensifiers, and emoticons. Rule-based methods provide interpretability and can capture domain-specific sentiment patterns. However, they may be limited in handling complex linguistic phenomena and require manual rule engineering.

2.Lexicon-Based Methods: Lexicon-based approaches utilize sentiment lexicons or dictionaries that contain words or phrases with associated sentiment scores. Each word's sentiment score is aggregated to determine the overall sentiment of a tweet. Lexicons can be general-purpose or domain-specific. Sentiment scores may also be adjusted based on linguistic modifiers or contextual information. Lexicon-based methods are computationally efficient and can handle

informal language and slang. However, they may struggle with contextual ambiguity and the inclusion of new or domain-specific terms not present in the lexicon.

3. Machine Learning Algorithms: Machine learning algorithms are used to classify tweets into sentiment categories based on manually labeled training data. Commonly used algorithms include Naive Bayes, Support Vector Machines (SVM), Decision Trees, and Random Forests. These algorithms rely on feature engineering, where various textual features such as n-grams, part-of-speech tags, and syntactic dependencies are used to train the classifiers. Machine learning approaches offer flexibility and can capture complex patterns in the data. However, they require extensive feature engineering and may not generalize well to new or domain-specific data.

4. Hybrid Approaches: Hybrid approaches combine multiple traditional methods to improve sentiment analysis performance. For example, a lexicon-based approach can be combined with a rule-based method to handle contextual ambiguity. Machine learning algorithms can also be used to augment rule-based or lexicon-based approaches by learning additional patterns and features from the data.

While traditional approaches have been widely employed in sentiment analysis of Twitter, they have certain limitations. They may struggle with the informal language, evolving trends, and contextual nuances present in Twitter data. However, traditional methods still provide a solid foundation and can be used as baselines for comparison and benchmarking against more advanced approaches like deep learning models. Additionally, hybrid approaches that combine the strengths of different traditional methods can yield improved performance in sentiment analysis tasks.

3.3 Machine Learning Approach

Machine learning approaches are widely used for Twitter sentiment analysis due to their ability to automatically learn patterns and classify tweets based on their sentiment. Here are some popular machine learning approaches used for Twitter sentiment analysis:

1. Naive Bayes Classifier:

Naive Bayes is a probabilistic classifier that assumes independence between features.

It calculates the likelihood of a tweet belonging to a specific sentiment class based on the occurrence of words/features in the tweet.

2. Support Vector Machines (SVM):

SVM is a powerful supervised learning algorithm used for text classification tasks.

It maps the tweets into a higher-dimensional space and finds a hyperplane that maximally separates tweets belonging to different sentiment classes.

3. Random Forest:

Random Forest is an ensemble learning algorithm that combines multiple decision trees.

It can be used for sentiment analysis by training a forest of decision trees on labeled tweet data to predict sentiment labels for new, unseen tweets.

4. Logistic Regression:

Logistic Regression is a statistical modeling technique used for binary classification tasks.

It assigns probabilities to tweets belonging to different sentiment classes based on a linear combination of features and applies a sigmoid function to predict the sentiment label.

5. Gradient Boosting:

Gradient Boosting is another ensemble learning technique that combines weak learners (decision trees) in a sequential manner.

It builds a strong classifier by iteratively adding decision trees that correct the mistakes made by previous trees.

6. Recurrent Neural Networks (RNN):

RNNs are neural networks designed to handle sequential data, such as text.

They can capture the temporal dependencies in tweets and are useful for sentiment analysis by considering the context of words in a tweet.

6. Convolutional Neural Networks (CNN):

CNNs are widely used for image processing tasks, but they can also be adapted for text analysis, including sentiment analysis.

They apply convolutional filters to capture local features in textual data and extract relevant information for sentiment classification.

7. Long Short-Term Memory (LSTM):

LSTMs are a type of RNN that address the vanishing gradient problem, enabling the modeling of long-range dependencies in text data.

They have been successfully employed for sentiment analysis on Twitter due to their ability to capture the context and sentiment nuances in tweets.

8. BERT (Bidirectional Encoder Representations from Transformers):



BERT is a transformer-based language model that has achieved state-of-the-art performance in various natural language processing tasks, including sentiment analysis. It utilizes a transformer architecture to capture the contextualized representation of words and has been fine-tuned specifically for sentiment analysis.

IV. STEPS OF APPROACH

1. Importing necessary libraries:

- Pandas- to create dataframe
- Numpy- sum and mean of polarity and subjectivity
- Tweepy- to use Twitter API
- Textblob- for doing sentiment analysis
- Regex- for cleaning the tweets
- Streamlit - for a web framework
- Pyplot - for plotting graphs

2. Creating a Twitter App:

To extract tweets, we need to access our Twitter account and create an app. We can get the credentials by creating a Twitter developer’s account.

3. Creating a function to access Twitter API:

Here we authorize tweepy with the credentials and access tokens.

4. Getting Tweets:

Here we extract tweets of a particular username or/and a keyword, sent a language, and set a count.

5. Creating DataFrame:

Extracted the tweets, now converting them into a data frame.

V. RESULT ANALYSIS

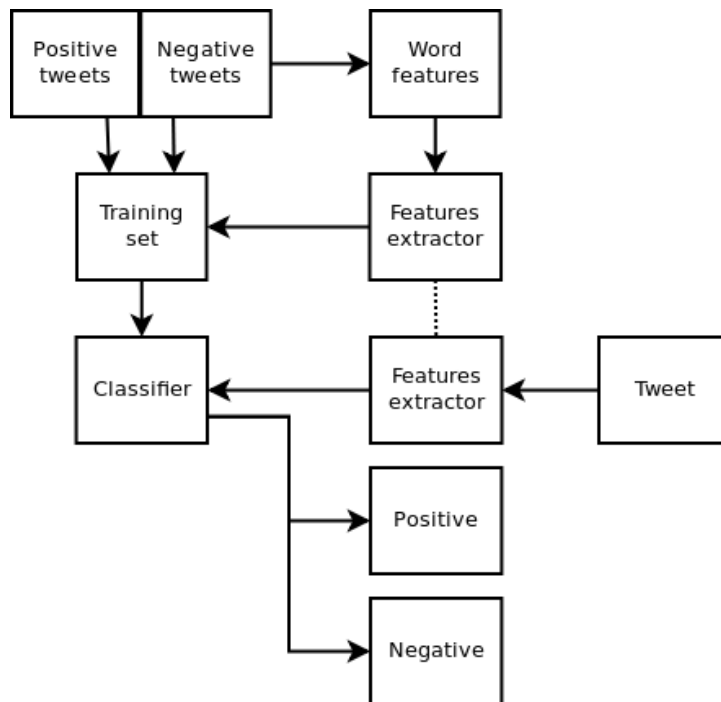


Fig 2: Result overview of sentiment analysis

In Twitter sentiment analysis, the classification of tweets into positive or negative sentiment categories is a fundamental task. It involves determining the polarity or emotional tone expressed in a tweet, whether it is positive or negative. Here's an elaboration on negative and positive tweets in the context of sentiment analysis on Twitter:

Negative Tweets:

Negative tweets express negative sentiments, emotions, or opinions. They often convey dissatisfaction, criticism, disappointment, or frustration.

Examples of negative tweets can include complaints about a product or service, expressing disagreement with a statement or action, venting frustrations, or sharing negative experiences.

Negative tweets may contain negative words, strong emotive language, sarcasm, or expressions of sadness or anger.

Sentiment analysis aims to correctly identify and classify negative tweets to understand public opinion, customer feedback, or social sentiment towards specific topics or entities.

Positive Tweets:

Positive tweets express positive sentiments, emotions, or opinions. They convey satisfaction, happiness, agreement, enthusiasm, or support.

Examples of positive tweets can include expressing joy or excitement, praising a product or service, sharing positive experiences, endorsing a viewpoint, or expressing gratitude.

Positive tweets may contain positive words, words indicating happiness or satisfaction, exclamation marks, emojis, or expressions of appreciation.

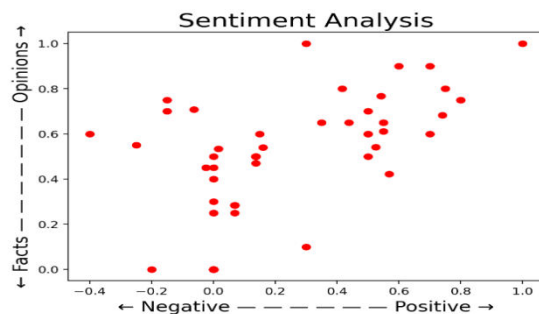


Fig 3 : Facts and opinion graph

Identifying and classifying positive tweets through sentiment analysis is crucial for understanding public sentiment, customer satisfaction, brand perception, or detecting positive trends and sentiments related to specific topics or entities. It's important to note that sentiment analysis is not limited to just negative and positive categories. There can be additional sentiment categories such as neutral, mixed, or even fine-grained sentiment analysis that assigns sentiment scores on a continuous scale.

VI. CONCLUSION

In conclusion, Twitter sentiment analysis is a valuable tool for understanding public sentiment and opinions expressed on the platform. By applying machine learning algorithms and natural language processing techniques, tweets can be classified into positive or negative sentiment categories. This analysis provides insights into public sentiment, helps manage brand reputation, enables customer feedback analysis, aids in market research, facilitates social listening, supports crisis management, and assists in political analysis. Overall, Twitter sentiment analysis empowers businesses and organizations to make informed decisions and engage effectively with their audience based on the sentiment and feedback received on Twitter.

REFERENCES

- [1]. Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12).
- [2]. Mohammad, S., & Turney, P. (2010). Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, 26-34.
- [3]. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. Proceedings of the Workshop on Languages in Social Media (LSM), 30-38.



- [4]. Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, 60(11), 2169-2188.



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.423



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com



www.ijirset.com

Scan to save the contact details