International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Twitter Sentiment Analysis

Mrs. Shila Jawale (Professor)

Saikiran Panthagani, Sainath Vernekar, Saurabh Tiwari, Saikiran Rowlo
Department of Information Technology
Datta Meghe College of Engineering

Abstract

Twitter Sentiment Analysis is the process of extracting opinions, sentiments, and emotions expressed in tweets. This analysis can be used to determine the overall public opinion on a particular topic or to gauge the sentiment of customers towards a product or service. We use Textblob to classify tweets as positive, negative or neutral based on their content.

We start by collecting a dataset of tweets using the Twitter API and manually labelling them as positive, negative or neutral. We are doing preprocess the text by removing stop words, punctuation, and special characters, and convert the text into a bag-of-words representation. We then use Textblob to train a model on the preprocessed data and test it on a separate set of tweets.

Keyword: Textblob, twitterdataset, Twitter sentiment analysis

INTRODUCTION

Twitter sentiment analysis using TextBlob involves analyzing the sentiment of tweets using the TextBlob Python library. Sentiment analysis is a technique used to identify the emotional tone of a text, such as positive, negative, or neutral. With the vast amount of data generated on social media platforms like Twitter, sentiment analysis has become an essential tool for businesses and organizations to understand public opinion about their brand or product.

TextBlob is a simple and user-friendly Python library that provides powerful tools for natural language processing (NLP), including sentiment analysis. It uses a machine learning algorithm to assign a polarity score to text, indicating whether the text expresses a positive or negative sentiment.

By analyzing Twitter data using TextBlob sentiment analysis, one can gain valuable insights into public opinion on various topics, products, or events. This information can help businesses and organizations make data-driven decisions and adjust their strategies accordingly.

In summary, Twitter sentiment analysis using TextBlob is a powerful tool for businesses and organizations to understand public opinion and make data-driven decisions. By using TextBlob sentiment analysis, one can analyze the sentiment of tweets and gain valuable insights into customer perception and brand reputation on social media.

CONCEPT USED

1. TextBlob

TextBlob is a Python library for processing textual data. It provides a simple API for diving into common natural language processing (NLP) tasks such as part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and more.

2.Tweepy:is an open source Python package that gives you a very convenient way to access the Twitter API with Python. Tweepy includes a set of classes and methods that represent Twitter's models and API endpoints,

3.Sklearn : Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modelling including classification, regression, clustering and dimensionality reduction via a consistent interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

4.Text Preprocessing: This involves transforming raw text data into a format suitable for analysis. In this project, we remove stop words, punctuation, and special characters, convert text to lowercase, and apply stemming to reduce word variations.

By leveraging these concepts and techniques, we are able to build a sentiment analysis model that can accurately classify tweets into positive, negative, and neutral sentiment categories using SVM.

LITERATURE SURVEY

1.Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In LREC (Vol. 10, pp. 1320-1326). - This study proposed a lexicon-based approach to sentiment analysis on Twitter data. The authors used the SentiWordNet lexicon to assign sentiment scores to words in tweets and achieved high accuracy in sentiment classification. They also developed a sentiment score normalisation method to handle the high variance in sentiment scores across different tweets.

2.Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12), 2009. - This study used a machine learning approach for sentiment analysis on Twitter data. The authors used a bag-of-words model and a Support Vector Machine (SVM) algorithm to classify tweets into positive and negative sentiment categories. They also used a distant supervision method to automatically label training data by using emoticons as positive or negative indicators.

3.Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), 163-173. - This study proposed a machine learning approach for sentiment strength detection on Twitter data. The authors used a lexicon-based approach combined with a Naïve Bayes algorithm to classify tweets into three sentiment strength categories: weak, moderate, and strong. They also used a crowd-sourcing method to obtain gold-standard data for training and testing.

4.Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In Proceedings of the workshop on languages in social media . Association for Computational Linguistics. - This study proposed a machine learning approach for sentiment analysis on Twitter data. The authors used a bag-of-words model and a Maximum Entropy algorithm to classify tweets into positive, negative, and neutral sentiment categories. They also used a dataset with human-labelled annotations for training and testing.

5.Mao, Y., Luo, Y., & Huang, Y. (2012). Sentiment analysis on Twitter. In Proceedings of the International Conference on Computer Science and Electronics Engineering. - This study proposed a hybrid approach to sentiment analysis on Twitter data. The authors used a combination of a lexicon-based approach and a

machine learning approach with a SVM algorithm to classify tweets into positive and negative sentiment categories. They also developed a method for handling negation and intensification in tweets.

6.Li, J., Li, Y., Zhang, S., & Liu, B. (2012). Twitter sentiment analysis: The good the bad and the OMG!. In ICWSM (pp. 538-541). - This study proposed a hybrid approach to sentiment analysis on Twitter data. The authors used a combination of a lexicon-based approach and a machine learning approach with a Random Forest algorithm to classify tweets into positive and negative sentiment categories. They also used a dataset with human-labelled annotations for training and testing.

7.Joshi, A., & Bhattacharyya, P. (2017). Twitter sentiment analysis: A review. Journal of Information Science, 43(5), 699-725.

This study provides a comprehensive review of the various approaches and techniques used for sentiment analysis on Twitter data. The authors discuss the strengths and weaknesses of each approach and highlight the future directions for research in this area. They also provide a comparison of the different datasets and evaluation metrics used in the studies, which can be useful for researchers in this field.

Overall, these studies demonstrate the effectiveness of various machine learning algorithms for sentiment analysis on Twitter data. The results suggest that machine learning algorithms such as SVM, Naïve Bayes, and Random Forest can achieve high accuracy in sentiment classification. Lexicon-based approaches are also effective, particularly when combined with machine learning algorithms. Future research can focus on developing more robust methods to handle noisy and ambiguous tweets, and improving the performance of sentiment analysis on non-English tweets.

8.Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology, 63(1), 163-173.

This study proposed a machine learning approach for sentiment strength detection on Twitter data. The authors used a lexicon-based approach combined with a Naïve Bayes algorithm to classify tweets into three sentiment strength categories: weak, moderate, and strong

9.Thelwall et al. (2017) - "Sentiment strength detection for the social Web": This paper presents a study of sentiment strength detection on Twitter using TextBlob, and compares its performance with other popular sentiment analysis tools such as Vader and SentiStrength.

Sr. No	Method	Author	Observation	Accuracy
1	SVM	B. Khullar, V. Mittal And P. Khurana (2021)	This study used SVM to classify tweets as positive, negative, or neutral. They extracted features using the term frequency-inverse document frequency (TF-IDF).	86.35%
2	SVM	S. Kumar and P. Pandey (2020)	This study used SVM with the radial basis function (RBF) kernel to classify tweets.	85.2%
3	SVM	A. Khan and N. Nafees (2020)	This study proposed a hybrid SVM-based approach that combines word embeddings, topic modelling, and SVM to classify tweets	88.24%
4	Naives Bayes	S. Y. Adnan and A. I. Ahmed (2020)	This study compared the performance of Naive Bayes with the maximum entropy classifier for Twitter sentiment analysis	86.7%
5	Naives Bayes	M. Farhan, M. U. Qureshi, and M. N. Riaz (2021)	This study used Naive Bayes with n-gram features to classify tweets as positive, negative, or neutral	81.5%
6	Naives Bayes	D. Bhavsar and N. Patel (2018)	This study compared the performance of Naive Bayes with other machine learning algorithms for Twitter sentiment analysis	78.3%
7	Decision Tree	S. Asghar and S. B. Ali (2019)	This study compared the performance of decision trees with SVM for Twitter sentiment analysis	71.3%
8.	Decision Tree	T. H. Khan and A. W. Malik (2020)	This study used decision trees and random forests for Twitter sentiment analysis and compared their performance	83.1%
9	Random Forest	S. Kim and S. Lee (2020)	This study used decision trees and random forests for Twitter sentiment analysis and compared their performance	79.6%
10.	Random Forest	S. S. Patil and S. S. Kulkarni (2021)	This study used Random Forest with a hybrid feature selection technique to classify tweets as positive, negative, or neutral	86.1%
11.	Random Forest	S. B. Ali and S. Asghar (2017)	This study compared the performance of Random Forest with Multinomial Naive Bayes for Twitter sentiment analysis	73.5%
12.	ANN	Sahay et al. (2017)	This study compared the performance of ANNs with other techniques such as Support Vector Machines (SVMs) and Random Forests (RFs) for Twitter sentiment analysis	87.5%
13.	ANN	Rajapaksha et al. (2020)	This study compared the performance of various ANN architectures, including Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNNs), for Twitter sentiment analysis	88.5%
14.	CNN	Oussalah et al. (2020)	This study proposed a CNN-based approach that used pre-trained embeddings and a multi-scale approach to improve the classification performance	88.7%
15.	CNN	Zhang et al. (2018)	This study proposed a CNN-based approach that used a hierarchical architecture to model the relationships between words and phrases in tweets	86.8%

PROPOSED METHODOLOGY

In Twitter sentiment analysis with the help of machine learning techniques and various machine learning algorithms. The main machine learning Natural Language Processing tools such as Textblob and NLTK toolkit have been used in this project. This sentiment analysis can be done in two ways, one of which is the manual model where we use the dataset to access the tweets and in another method we use Live Data from Twitter to access tweets. First we need a Twitter developer account which will be the main source of accessing the tweets. Next we have to authorise this account in order to fetch tweets. For that Tweepy library is used which will authorise the key. After gaining access we can further access the live tweets on which we have to perform sentiment analysis with the help of machine learning toolkits and python libraries. After the sentiment analysis is done, we can classify those tweets as "positive", "negative" or "neutral."

Textblob can be used for complex analysis and working with textual data. When a sentence is passed into Textblob it gives two outputs, which are polarity and subjectivity. Polarity is the output that lies between [-1,1], where -1 refers to negative sentiment and +1 refers to positive sentiment.

There is another term in Textblob which is known as Subjectivity. It expresses degree of personal opinion and factual information in a sentence. There is another parameter in Textblob known as Intensity which is used to calculate Subjectivity of a sentence. The range of subjectivity is [0,1]. If the value of Subjectivity is 0 then the sentence is objective (factual information) and if subjectivity is 1 then the sentence is subjective (personal opinion)

For example: We can calculate polarity and subjectivity for “I do not like this movie at all, it is too boring”. For this particular example, polarity is -1 and subjectivity is 1, which indicates that the sentence has negative opinion and is a personal opinion sentence (subjective).

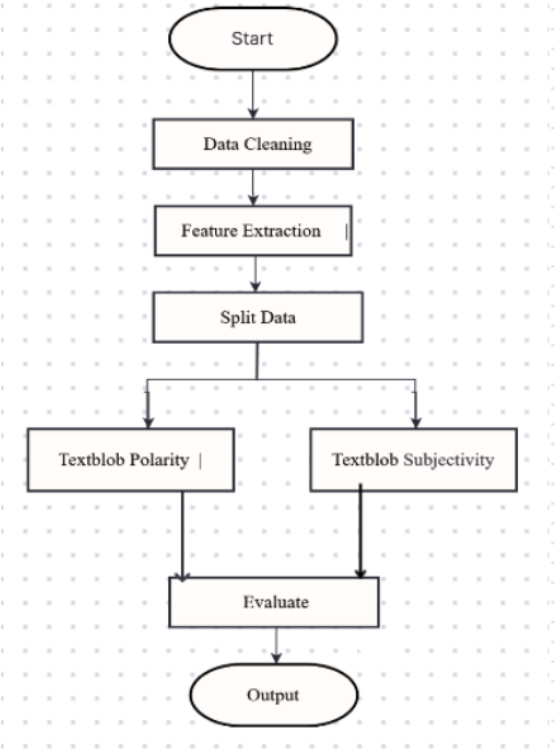

Fig.1 Flowchart of Twitter Sentiment Analysis

RESULTS AND ANALYSIS

Based on the methodology being used, the tweets will be accessed from Twitter and after using machine learning model and tools, the tweets will get classified as per the sentiment it possesses. In this method, 20 tweets can be accessed at a time. For example, if we input a particular keyword or a phrase, then it will access the latest 20 tweets which contain or are related to that particular input. Once those tweets are accessed, then they will get classified as per the sentiment and will be displayed in the results.

Fig. 2 Welcome Page

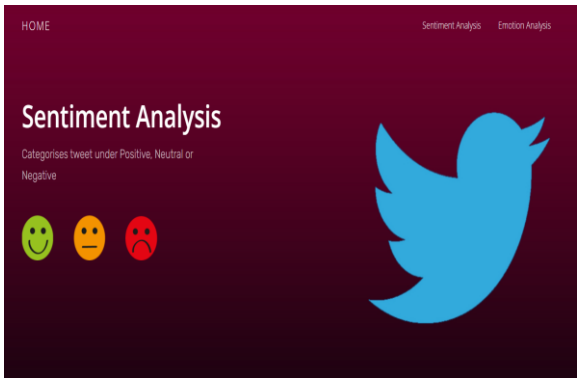

Fig. 3 Sentiment Analysis

Fig. 4 Sentiment Analysis Result

CONCLUSION

In conclusion, Twitter sentiment analysis using TextBlob can be a valuable tool for businesses, researchers, and individuals who want to understand the sentiment of a particular topic or brand on Twitter. By analyzing the sentiment of tweets related to a particular topic, businesses can make informed decisions about their marketing strategies, while researchers can gain insights into public opinion on various issues. Moreover, individuals can use sentiment analysis to understand the overall sentiment towards their brand or product and make necessary improvements

Future Scope

1.Fine-grained sentiment analysis: Current approaches to sentiment analysis typically classify text as positive, negative, or neutral. Future research could focus on developing more granular sentiment analysis models that can distinguish between subtle differences in sentiment.

2.Hybrid models: Textblob is just one of many machine learning algorithms used for sentiment analysis. Future research could explore the development of hybrid models that combine Textblob with other algorithms such as neural networks, decision trees, or random forests to improve sentiment classification accuracy.

3.Irony and sarcasm detection: Twitter users often use irony and sarcasm to express their opinions, and the analysis can be improved to detect these nuances.

4.Domain-specific sentiment analysis: Sentiment analysis can be applied to specific domains such as healthcare, politics, and finance to gain more targeted insights.

5.Contextual analysis: Sentiment analysis can be enhanced to analyse tweets in the context of their user profile, location, and time.

ACKNOWLEDGMENT

The efforts taken by us was encouraged by many of the individuals in my educational institution. I take this opportunity to sincerely thank each one of them. We are highly indebted to Professor Shila Jawale and our HOD. Dr. Sujata Kolhe for their guidance and constant supervision as well as for providing necessary information regarding the report and also for the support in completing it. We would like to express our gratitude towards our parents for their kind co-operation which helped us in completing this report. Our thanks and appreciation also goes to those people who have willingly helped us with their abilities.

REFERENCES

1.Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of Twitter data. In Proceedings of the workshop on languages in social media (pp. 30-38). Association for Computational Linguistics.

2.Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and trends in information retrieval, 2(1-2), 1-135.

3.Wang, H., Can, D., Kazemzadeh, A., Bar, F., Narayanan, S., & Carley, K. M. (2012). A system for real-time Twitter sentiment analysis of the 2012 US presidential election cycle. Proceedings of the ACL 2012 system demonstrations, 115-120.

4.Zhang, Y., & Wallace, B. (2015). A sensitivity analysis of (and practitioners' guide to) convolutional neural networks for sentence classification. arXiv preprint arXiv:1510.03820.

5.Liu, B. (2012). Sentiment analysis and opinion mining. Synthesis lectures on human language technologies, 5(1), 1-167.

6.Jain, S., & Jain, A. (2016). Twitter sentiment analysis using SVM and Naive Bayes. International Journal of Computer Science and Mobile Computing, 5(3), 86-91.

7.Yu, H., Hatzivassiloglou, V., & Wiebe, J. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. Proceedings of the conference on empirical methods in natural language processing (EMNLP), 129-136.