International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 11 | Nov - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Sentiment Analysis of Social Media Data

A Machine Learning Approach to Social Media Opinion Mining

Mrs. Said Pranjal Sandip, Mrs. Bhakti Chandrakant Kale

Assistant Professor, Department of Master of Computer Application, JCEI's Jaihind Institute of Management and Research, Vadgaon Sahani Email: sandipsaid6@gmail.com
Kalebhakti56@gmail.com

Abstract - Sentiment analysis is a method that looks at text to find out what people think and feel in posts. Sentiment analysis reads messages, comments and reviews on media to see if the feelings are good, bad or neutral. Instagram, Facebook, X (Twitter) and YouTube are the places where many comments and posts appear. Today a huge amount of data is created by comments, posts, reviews and online chats. Sentiment analysis helps us see what the public thinks about products, services, events and social matters.

The main goal of this project is to build a system that gathers social media text and sorts it into sentiment categories. Sentiment analysis categorizes text based on the emotion or opinion expressed, classifying it as positive, negative, or neutral. The system first cleans the text by removing symbols stop words and other irrelevant bits. Then machine learning and natural language techniques are applied to study the text and give a sentiment score.

Sentiment analysis can help businesses read customer feedback help researchers study opinion and help organizations watch how people react to various subjects. Sentiment analysis also finds trends and patterns, in discussions.. The accuracy of sentiment analysis can drop when sarcasm, slang, spelling errors or many languages are used.

In conclusion sentiment analysis turns a lot of raw social media data into insight. Sentiment analysis helps individuals make informed decisions by identifying public emotions, attitudes, and opinions expressed across online platforms

Keywords: Phishing Detection, Machine Learning, Cybersecurity, URL Analysis, Classification, Website Security.

1. INTRODUCTION

In the modern world, one of the most important trends is the increasing amount of time people spend on multimedia resources: Instagram, Facebook, X (Twitter), YouTube, and many others. Moreover, each person expresses their opinion and thoughts there: reviews, comments, complaints, monologues, and much more. Each day, thousands of messages appear on the Internet that must somehow be processed and analyzed. The task of Sentiment Analysis for Social Media Data is to collect and classify information from social networks for a specific purpose. Sentiment analysis is a technique in natural language processing (NLP), which falls under the broader field of artificial intelligence (AI) and focuses on identifying opinions and emotions expressed in text. To put it another way, it's a computer program that comprehends human emotions and classifies them accordingly. For example, the sentence "I love this product!" will be classified as a positive statement, while the sentence "I'm dissatisfied with the company's work" will be classified as negative. A neutral sentence might sound like "Nothing can be said about this company." Furthermore, the analysis of sentiment allows extracting and quantifying information that is not immediately obvious. In other words, it processes and classifies data mined from online sources. One of the key areas where sentiment analysis is used is in mining customer feedback. As a result, companies receive statistics on the degree of customer satisfaction with specific products or services. Additionally, such a program might be utilized to assess students' reactions to a class, a film, or even a lecture. Moreover, processing of public opinion may be used to study and analyze attitudes towards various social and political events.

1.1 Problem Statement

Sentiment Analysis of Social Media Data is a project intended to identify and extract users' opinions, emotions, and sentiments expressed in media. People express their views on various products, services, films, and events by sharing posts, comments, and reviews. The main challenge of the project is to develop a system, capable of analyzing media texts automatically and classifying them as positive, negative, or neutral. The system will enable the user to examine the contents' sentiment and identify users' attitudes and opinions. Sentiment Analysis of Social Media Data can be applied in diverse fields, such as business, to assess the satisfaction of customers, track public opinion, and identify the trends of the audience's attitude. The project will utilize Natural Language Processing (NLP)

approaches and algorithms of machine learning to analyze the body of text and produce specific results. Sentiment Analysis of Social Media Data aims to reduce the time spent on manual analysis, enable a rapid assessment of opinions, and facilitate the analysis of public opinion by evaluating thousands of documents.

1.2 Objectives

1. Opinion Analysis: Conducting analysis to explore and extract opinions, sentiments, and attitudes of individuals and groups towards specific topics.
2. To identify sentiments: Classifying media posts and comments according to their positive, negative, or neutral emotions and opinions.
3. To process text data: To carry out Natural Language Processing (NLP) procedure to analyze and understand text data.
4. Customer Feedback Analysis: To assess customers' opinions and analyze their feedback to determine their satisfaction levels with various products, services, and brands.
5. To sentiment detection: Saving time and effort by using a machine learning approach to carry out an extensive scan of the social media space.
6. To identify trends: To track the sentiments that emerge from events, new products, and other topics of interest.
7. Decision-Making Support: To help organizations and businesses in their decision-making process by analyzing public opinion and feedback that is available in social media space.

1.3 Contributions

The project deals with recognizing people's sentiments and opinions by analyzing their contents published on various social media networks. The project titled Sentiment Analysis of Social Media Data uses natural language processing (NLP) techniques and machine learning models to examine textual content and determine whether the expressed opinions are positive, negative, or neutral.

The project's strengths include:

1) Automatic Sentiment Classification

The Sentiment Analysis of Social Media Data project automatically detects the emotional tone of textual content, reducing the time and effort required for manual evaluation.

2) Efficient Data Analysis

Sentiment Analysis of Social Media Data reduces the workload of studying the online public opinion and helps get a deeper understanding of the audience quickly and easily.

3) Learning Customer Preferences

Sentiment analysis of social media data helps businesses understand customer preferences, identify areas of dissatisfaction, and evaluate public opinions and perceptions of their brands or products.

4) Public Opinion Monitoring

The project helps assess public opinion on products, services, brands, events, and other topics.

5) Support, for Decision-Making

Sentiment Analysis of Social Media Data allows making informed decisions about the market and developing strategies based on customers' opinions and feedback.

2. LITERATURE REVIEW

Sentiment analysis has become a research area of natural language processing due to the vast amount of information shared on different social media platforms. In addition, people use the various online platforms to express their opinions, emotions, experiences and feedback. The information contained in these expressions plays a central role in helping researchers and organizations determine the reactions of people towards specific topics. Several researchers have proposed different approaches to identify and classify sentiment and classify online text into positive, negative and neutral.

Pang, Lee and Vaithyanathan researched the application of machine learning techniques for sentiment classification in their research paper Thumbs up? Sentiment Classification Using Machine Learning Techniques. They tested different methodologies Bayes, Maximum Entropy and Support Vector Machines to identify whether movie reviews expressed positive or negative opinions. Their research proved that the application of machine learning techniques was useful for opinion classification although the analysis of the emotional content of the text was a difficult task.

Pang and Lee offered an overview of opinion mining and sentiment analysis in their research Opinion Mining and Sentiment Analysis. They discussed various approaches used in identifying information, extracting opinions and classifying sentiments from textual data. Moreover, their work demonstrated the importance of analyzing language and context when mining peoples opinions. This research became the basis of later works on sentiment classification and opinion mining.

Go, Bhayani and Huang researched sentiment classification using Twitter messages in their study Twitter Sentiment Classification Using Distant Supervision. They studied the use of emoticons to automatically generate sentiment labels for training data and applied machine learning algorithms to classify tweets. Although their approach illustrated how large sets of short social media messages could be analyzed, challenges like language variability and non-formal communication of tweets remained an issue.

Liu, in Sentiment Analysis and Opinion Mining explained the concepts, methods and applications of sentiment analysis. The

work studied how opinion could be mined from text and how the extracted sentiment information could be applied to understand customer feedback and public attitudes. It also discussed the various difficulties associated with expressions, ambiguous words and change of meaning of specific terms based on context. This research offered a basis for analyzing the various techniques used in processing opinion based information.

3. METHODOLOGY

The project methodology consists of analyzing social media content to identify and classify users' opinions and sentiments expressed via social media posts and comments. The primary aim is to process text Data and identify if the statement reflects positive, negative, or neutral opinions. The system utilizes Natural Language Processing (NLP) and machine learning algorithms to analyze Data and generate outputs.

The initial step involves collecting relevant textual information such as social media posts, user comments, and online reviews, from available datasets or reliable sources with appropriate permissions. The Data collected should be relevant to the selected research topic. Contain adequate examples for sentiment analysis. The Data is scrutinized to remove irrelevant and duplicate entries that could hinder the quality of the analysis**.**

The Data is preprocessed to prepare it for analysis. The social media content is full of symbols, links, spelling mistakes, abbreviations, and informal language. The text is cleaned and formatted appropriately to enhance the analysis process. Tokenization, lowercasing, and word removal techniques can be utilized at this stage to facilitate the processing of text **Data and generate meaningful insights. This step aims to generate text suitable for the classification model.**

The preprocessed text is transformed into a mathematical representation that can be processed by a machine learning model. Bag of Words or Term Frequency–Inverse Document Frequency (TF-IDF) approaches can be used to represent the words and their frequencies in the dataset. The method selected depends on the type of Data collected and the algorithm utilized for the analysis.

The Data is subjected to the sentiment classification stage. A machine learning algorithm such as Bayes, Logistic Regression, or, Support Vector Machine can be used to train the model using labeled examples to generate opinions or sentiments expressed in the comments or posts. The algorithm analyzes the social media text and identifies the category of each message. The model then labels each comment or post as either negative, positive, or neutral.

The performance of the model is evaluated using training and testing Data. The training Data enables the model to learn the sentiments, while the test Data is used to evaluate the accuracy of the model in classifying new messages. Evaluation metrics such as accuracy, precision, recall, and F1-score can be used to fine-tune the model and optimize its performance in identifying opinions and sentiments expressed in social media posts and comments.

4. RESULTS

The sentiment analysis of social media data aims to identify the opinions, attitudes, and feelings of users through their messages on social media sites. Collected data is processed and analyzed by using NLP and machine learning in order to classify the type of the message as either positive, negative, or neutral.

The result of such analysis helps to get to know better the attitude of users to different products, services, events, and social issues. If the comment is positive, it means that the user feels satisfied or appreciates something; if it is negative, it can suggest the complaint or dissatisfaction of the user. Neutral comments usually have no emotions attached.

Sentiment analysis allows identifying patterns of users' opinions and tracking the sentiments of users over time. Such data might be helpful for various purposes.

The accuracy of results depends on the quality of collected data, the text preprocessing steps, and the performance of classification algorithm. However, some challenges, such as sarcasm, slang usage, spelling mistakes, and mixed emotions, can negatively impact the results of classification.

5. DISCUSSION

Using sentiment analysis of the data from the social media sites is an excellent way to assess the opinion or emotions of the Internet surfers. There is an ample amount of data on the social media sites in the form of comments, messages, feedback, or any other form of communication that one can find online in order to express one's emotions or opinions. Sentiment analysis will allow the researchers to understand what the said person was feeling when expressing their opinion.

The debate over the current project lies in the usage of sentiment analysis techniques to determine the reaction and sentiment of the public toward the issue at hand through online communication. The text can be analyzed using certain

natural language processing and machine learning techniques in order to determine what category it falls under.

These sentiments or attitudes can be used by companies to determine the satisfaction of the customers or other important factors that need improvement.

6. CONCLUSION AND FUTURE WORK

Social Media Sentiment Analysis is an effective approach of understanding the sentiments of people through the use of social media. In the aforementioned project, the analysis will be performed using Natural Language Processing and Machine Learning algorithms.

The sentiment analysis works on the text data of the social media and other online sources including posts, comments, and reviews. Through this process, it becomes possible to obtain useful information from the text data for the purpose of understanding opinions of people towards services, products, events and other matters of interest. In addition, businesses can use such data to know what their clients think about them and improve their performance based on the needs and preferences of their customers. Nevertheless, there are several challenges encountered when performing sentiment analysis on text data some of which include slang language, misspellings, sarcasm, and different contexts of some phrases.

In future, the project can be improved by using advanced Machine Learning and Deep Learning algorithms so that the accuracy rate of sentiment prediction can be increased. Additionally, it might also be important to make the system support other languages such as Marathi and Hindi, as there is also useful data found in these languages than can be used to carry out successful sentiment analysis. The implementation of real-time sentiment analysis can also be integrated into the system.

REFERENCES

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. Foundations and Trends in Information Retrieval, 2 (1-2), 1-135. https://doi.org/10.1561/1500000011

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers. https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media , 8 (1), 216-225. https://doi.org/10.1609/icwsm.v8i1.14550

Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. Stanford University, CS224N Project Report.

Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In Proceedings of the seventh international conference on language resources and evaluation (LREC 2010). https://aclanthology.org/L10-1263/

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2 (1), 1-8. https://doi.org/10.1016/j.jocs.2010.12.007

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. In Proceedings of the fourth international AAAI conference on weblogs and social media.

Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the OMG! In