



Sentiment Analysis of Twitter Feeds Using Flask Environment: A Superior Application of Data Analysis

Astha Modi¹ · Khelan Shah¹ · Shrey Shah¹ · Samir Patel² · Manan Shah³

Received: 7 December 2021 / Revised: 25 July 2022 / Accepted: 31 August 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

In this challenging world, social media plays a vital role as it is at the pinnacle of data sharing. The advancement in technology has made a huge amount of information available for data analysis and it is on the hotlist nowadays. Opinions of the people are expressed and shared across various social media platforms like Twitter, Facebook, and Instagram. Twitter is a prodigious platform containing an ample amount of data and analyzing the data is of topmost priority. One of the most widely utilized approaches for classifying an individual's emotions displayed in subjective data is sentiment analysis. Sentiment analysis is done using various algorithms of machine learning like Support Vector Machine, Naive Bayes, Long Short-Term Memory, Decision Tree Classifier, and many more, but this paper aims at the generalized way of performing Twitter sentiment analysis using flask environment. Flask environment provides various inbuilt functionalities to analyze the sentiments of text into three different categories: positive, negative, and neutral. Also, it makes API calls to the Twitter Developer account to fetch the Twitter data. After fetching and analyzing the data, the results get displayed on a webpage containing the percentage of positive, negative, and neutral tweets for a phrase in a pie chart. It displays the language analysis for the same phrase. Furthermore, the webpage calls attention to the tweets done on that phrase and reveals the details of the tweets. Considering the major industry runners of three different sectors namely Enterprises, Sports Apparel Industry, and Multimedia Industry, we have analyzed and compared sentiments of two different Multinational companies from each sector.

Keywords Sentiments · Twitter · API · Flask · Python · Webpage · HTML

✉ Samir Patel
Samir.patel@sot.pdpu.ac.in

¹ Department of Information and Communication Technology, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat 382426, India

² Department of Computer Science, School of Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat 382426, India

³ Department of Chemical Engineering, School of Energy Technology, Pandit Deendayal Energy University, Gandhinagar, Gujarat 382426, India

1 Introduction

With the innovation and evolution of online technology, there is a tremendous volume of data available on the web for internet users, as well as a substantial amount of data generated [1]. With this humungous amount of data being generated, it can be processed over the internet leading to a growing demand for computer-assisted data analysis [2]. Social networking sites generate most of the data either in a structured, unstructured, or semi-structured format. With the advancement of time, these networks are accessible at our fingertips at any moment, therefore analyzing the data plays a vital role in the growth of industries.

The analysis of data directly gets connected to two terms: Data Science and Big Data Analytics. Data Science is an incorporative domain that allows retrieving information from raw data using analytical techniques and algorithms. It entails filtering, consolidating, and modifying data to evaluate the data. The field of data science extensively uses data mining. To find patterns and potentially important facts from huge databases is a technique known as Data mining. Sometimes, it is also referred to as knowledge discovery in data [3]. The insights driven by data mining are used by businesses to develop a solution and improvise their business [4]. While Big Data Analytics is described as the use of technological advances, predictive modeling, and analytics to address commercial and industrial challenges [5]. Big data analytics assists firms in making inferences from today's massive data repositories. Massive volumes of data are currently produced by people, companies, and technologies. Examples include social media, cloud applications, and many more [6].

Social networking services such as Twitter, Facebook, Snapchat, and Instagram are fast growing in popularity because they enable individuals to exchange and voice their opinions on many themes, engage them in conversations with various communities, and post messages all over the globe [7]. Twitter has recently emerged as one of the most essential digital media platforms for information dissemination [8]. It is a social networking service that allows users to post and consume 140-character messages known as "Tweets." Users on this platform can utilize brief and direct communications to express their ideas and opinions, promote their findings, and argue with billions of other users. Everyone in the Twitter network does not always have a mutual connection with each other. The relationship in this scenario is either guided or unguided. The enormous data supplied by this microblogging platform, such as tweet messages, user information, and the number of followers/followings in the network, serve an important role in data analysis, prompting most studies to explore and evaluate the various interpretation methods to acquire the most recently used innovations [9]. One such method to analyze textual data to get the polarity of the emotions expressed is Sentiment analysis.

Sentiment analysis generally known as opinion mining is textual context-specific processing that finds and extracts subjective data from the source material, assisting businesses in improving the human emotion of their company, goods, or organization while tracking online conversations. Because of the accelerated expansion of social media, the relevance of sentiment analysis has grown in the current context [10]. The core principle behind sentiment analysis is to identify and classify the polarity of text documents or short phrases [11]. In this project, the sentiments are classified as

positive, negative, or neutral. This classification of polarity facilitates various organizations to record people's sentiments about their services as well as those of their opponents [12]. Moreover, this classification gives insights into the feelings and attitudes of people over the internet which is monitored for enhancing the performance of political parties, government agencies, movies, and so on [13].

Sentiment analysis is widely performed using natural language processing. Natural language processing emphasizes enabling computers to comprehend text and spoken speech in a manner like that of humans. It blends computational semantics (human language rule-based modeling) with statistics, machine learning models, and neural networks. These technologies, when combined, allow computers to process the conversations which may be in the form of text or speech data, and comprehend the true context, replete with the subject's purpose and mood. NLP at various levels is used for Sentiment analysis [14]. Other than NLP, there are various methodologies and algorithms by which one can perform sentiment analysis. A few of them are mentioned as follows:

1.1 LSTM

LSTM layers are used to capture long-term text data dependencies and improve performance. It can detect long-term relationships in sentences of unknown length and can be used to effectively govern information by avoiding a vanishing gradient. The memory cell in an LSTM is used to save the selected data for a longer period without decaying [15].

1.2 Naïve Bayes

The Naïve Bayes Method is a method that can be trained or used on small-scale data and can provide predictive results in real-time. It can also aid in classifying a class, the results of which can be used in parallel to increase the scale of the dataset, particularly in large-scale data case studies [16].

1.3 Decision Tree

The decision tree classification gives the end-user a better option for categorizing positive and negative tweets. It is accomplished by comparing the most frequent items generated by the rules in the training data with the most frequent items in the test data, allowing for simple classification [17].

1.4 SVM

Because SVMs are binary classifiers, they are frequently used for binary sentiment detection. The problem must be transformed into a set of binary classification problems before multi-class classification can be performed [18].

The above-mentioned algorithms have some other disadvantages; therefore, our motivation is to eliminate the task of data preprocessing and training the data with machine learning which consumes more time and is tedious.

The main contribution to this work includes:

- Using the Tweepy library, we analyze the sentiments of tweets as well as language analysis.
- Analysis performed for any desired keyword.
- Visually appealing results redirected to the website using flask environment.

In this paper, we focus to perform Sentiment analysis for a specific keyword on Twitter data using a flask environment. To access Twitter data, one must create a Twitter developer account. Four credentials are provided while creating the account which is necessary to access and analyze data. The proposed method does not require any machine learning algorithms. It is expected that our method would perform better, faster, and more efficiently than other existing works.

2 Literature Survey

Anupama et al. [19] proposed the process of creating software that examines the nature of tweets on a specific topic in this project. The user would be able to type in a keyword (hashtag) and receive the nature of that keyword based on the most recent tweets that contain that keyword. Each tweet was classified depending on whether it had a favorable or negative sentiment. Data was gathered from movie reviews on the IMDB website. The machine learning algorithm Naïve Bayes was utilized. The model's output was put to the test using a variety of testing metrics. Furthermore, our model outperforms the competition when it comes to extracting text from Twitter.

Shahzad et al. [20] aim to forecast the nature of Twitter users' reactions to scientific papers along with looking into what characteristics of research articles aid in such prediction. Scientists will be able to measure a new societal influence of their research articles by analyzing the attitudes of research articles on social media. They tested five sentiment analysis tools to see which ones were best for capturing the sentiment value of a tweet and chose NLTK VADER and Textblob. Followed by, dividing the sentiment value into three categories: negative, positive, and neutral. For research papers containing multiple tweets, they calculate the mean and median sentiment value of the tweets. Following that, constructed machine learning models to anticipate the attitudes of tweets about scientific papers and looked into the key features that guided the algorithms. Using the Random Forest Classifier, they obtained an accuracy of 89%.

Abdullah et al. [21] proposed a method to develop a machine learning model for analyzing Arabic messages on Twitter. They used Word2Vec for word embedding, which is the main source of features in this model. Naive Bayes was utilized as a baseline classifier, and two pretrained continuous bag-of-words (CBOW) models were explored. SMOTE (synthetic minority oversampling approach) was applied with and without many single-based and ensemble-based machine learning classifiers. The experimental results reveal that using word embedding with an ensemble and SMOTE

improved the average F1 score significantly when compared to the baseline classifier and other single- and ensemble-based classifiers without SMOTE.

Kumar Singh et al. [22] have proposed a Twitter sentiment analysis on the coronavirus outbreak using machine learning algorithms. According to the paper, tweets used to classify the sentiments were related to coronavirus, COVID19, Job, and school. The approach was to fetch the tweets by using the Twint library to access Twitter API's and search for hashtags like #school, #job, #coronavirus, and #covid19. The tweets related to four searches were obtained and the data cleaning process was applied to those tweets. Now the cleaned data was used for sentiment identification which aggregates the cleaned data and will extract the necessary features of the text. In the final step classification of sentiment will finalize sentiment polarity in -1 , 0 , 1 and subjectivity in positive negative and neutral.

Deshpande et al. [23] conducted sentiment analysis on Twitter using a machine learning algorithm called support vector machine (SVM). In the proposed approach, the tweepy library was used to fetch the Twitter data using Twitter API keys accessible from the Twitter developer account. After getting the text data, text processing is performed by removing stop words, stemming, tokenization, and negation handling. Thereafter, the useful words were extracted by using feature extraction so that the analysis becomes efficient. Now, the sentiment classification is done using naïve Bayes and an SVM algorithm. As an output, they obtain a sentiment lexicon which is a collection of different words having a score that classifies the positive, negative and neutral nature of a sentence.

Kharde and Sonawane [7] surveyed different techniques used for sentiment analysis of Twitter data. The pre-preprocessing of the data was performed using an ideal method that is stemming, lemmatizing, removing stop words, and feature extraction. After that, they compared four different techniques for sentiment analysis those are machine learning approach, the lexicon-based approach Cross-lingual approach, and the cross-domain approach. In every approach, multiple algorithms gave different accuracies. The analysis tasks consist of evaluating subjectivity classification, sentiment classification, and complimentary tasks, and also there were different levels of sentiment analysis that are word level, sentence level, document level, and feature-based sentiment analysis. In the end, the final results of the survey showed that the machine learning-based approach showed the highest accuracy and also required few efforts in human-labeled documents.

Education [24] proposed an article stating the technique of Twitter sentiment analysis using an adaptive deep recurrent neural network. The methodology was as clear as performing a classical machine learning approach. The tweets were collected from Twitter using API's and were stored in the database and the database is then pre-processed. The pre-processed data was reverted to the database and to classify the data it was modeled using an adaptive deep recurrent neural network classifier. The classification output obtained from modeling the data was then analyzed into positive and negative sentiments. The output obtained was then visualized in a line chart comparing the other machine learning approaches. The line chart concluded that ADRNN gave the highest accuracy and the accuracy also observed a hike when the number of tweets was increased.

Dabade et al. [25] aimed to analyze the sentiments of Twitter data using Deep Learning and machine learning. The main objective of the research was to compare the results obtained from different machine learning algorithms based on the size of the dataset and to identify the best-suited approach for sentiment analysis. The methodology involved data collection, data cleaning, and implementation of the model. The libraries used were pandas, NumPy, sci-kit-learn, matplotlib, text blob, Vader, fast text, flair, and genism, and the model used was logistic regression and support vector machine. Hence, the main conclusion of the research was to focus on fine-grain sentiment analysis of tweets and analyze emoji labels.

Arun and Srinagesh [26] conducted a multilingual sentiment analysis of Twitter using machine learning. The proposed ideology involves pre-processing of data, language translation for each multi-language tweet, and training the data to classify sentiments. The pre-processing of data involves two steps, firstly the English language data is cleaned and secondly, the non-English data is translated into the English language, and then it is cleaned. After that, the cleaned data was trained using various algorithms like multinomial Naïve Bayes, Logistic Regression, Support Vector Machine, Decision Tree, k-Nearest Neighbor, and random forest. The accuracy of all the models was calculated by plotting three metrics that are precision, recall, and F1-score. The best results were obtained by SVM which is 95% accuracy and random forest with 93% accuracy.

Reddy et al. [27] performed sentiment research on data from Twitter. The proposed system performs sentence-level analysis of sentiments. The whole process is done in seven phases that include feeding the input of usernames or hashtags, specifying the number of tweets to be analyzed, retrieving twitter data stored in a database, processing the tweets, feature extraction, classifying the data, and aggregating the score. The emotions of sentiment research are categorized into seven categories and they are strongly positive, positive, weakly positive, neutral, strongly negative, negative, and weakly negative. The result is going to be visualized within the variety of a pie chart that contains seven major emotions.

Gupta et al. [28] presented a study on sentiment analysis of Twitter using machine learning algorithms in python. The proposed approach follows the classical machine learning method. Firstly, the dataset is pre-processed and cleaned. Thereafter, the dataset is split into training and testing data and the feature extraction is performed on the data. Moving further, the training data is trained in the machine learning algorithm and the model is tested using a testing dataset. The testing data set is classified into the required sentiments. The proposed method was performed on various algorithms like DAN2, SVM, Bayesian logistic regression, Naïve Bayes, Random Forest classifier, Neural network, Maximum entropy, and Ensemble classifier. As a result, it was concluded that the Maximum entropy and Ensemble classifier gave the highest accuracy of 90%.

Hasan et al. [29] aimed to propose a machine learning-based analysis of sentiments of a Twitter account. The methodology was to get the Twitter data using Tweepy API and store it in a database. The database was then preprocessed by slang word remover, URL remover, special characters, and symbol elimination apart from that the Urdu tweets were also translated into English. The polarity calculation and sentiment analysis were performed using text blob, sentiwordnet, and W-WSD, and data

validation was done using Weka. The sentiments were classified into three categories positive, negative and neutral. To display the results, a bar chart was plotted for all three approaches of sentiment analysis versus polarity.

Kolchyna et al. [30] proposed a step-by-step method for sentiment analysis of Twitter data. Two approaches were used that are lexicon-based approach and the machine learning-based approach. Firstly, the data was pre-processed by tokenization, N-grams extraction, stemming and lemmatization, stop word removal, and POS tagging. After that, the data is trained using the lexicon approach and machine learning algorithms like WEKA classifiers. Thereafter, a new approach was performed by combining the lexicon and machine learning approach where the lexicon score is used as a feature in ML classification. As a result, the accuracy of the model got increased by 5%.

Table 1 displays the important attributes of the research papers with references displayed in the table.

3 Framework

Twitter is a microblogging social media platform where people can express their views and opinions by sharing texts and images. We have developed a method to determine the sentiments of the data shared by people on Twitter. To fetch this data twitter provides Twitter Developer Account where one can generate the API keys and use them to model the data.

What is Twitter Developer Platform?

The Twitter Developer Platform empowers users to integrate real-time, public and universal data of Twitter into their projects and its applications. This platform not only offers data but also offers materials and API's which assist users to incorporate and amplify twitter's impact by the means of analysis, applications, and other approaches. The Twitter API is a collection of automated interfaces which can be used for understanding or constructing twitter's discourse. It also enables users to find, access, interact with or generate a range of resources. These resources mainly include tweets, direct messages, trends, users, media, and many more.

3.1 System Design

Figure 1 presents the basic flow of the methods carried out to complete the work.

1. Login into your Twitter account if you already have one, else create a Twitter account and log in. Next, create a Twitter Developer account. One will receive the approval of their developer account within a few days. After receiving the approval, create a new project. Name the new project "Twitter Sentiment Analysis-1". Figure 2 presents the dashboard of the developer portal of the Twitter account. Our developer account has elevated access which provides us with 3 environments per project and access to 2 million tweets per project.
2. Generate keys in the developer account and obtain the following four keys: "consumer_key", "consumer_secret", "access_token", and "access_token_secret". These four keys are an integral part to carry out Twitter sentiment analysis and

Table 1 Literature review of different machine learning methods

Algorithm used	Accuracy	Limitation	Future work	References
XGB	89.81%	Their work is specific and does not look at the mood and emotions of the people	Plan to analyze public sentiments toward other essential topics, healthcare facilities by government, offline examination, and mental health by using DL algorithms to increase their performance on the dataset	Jalil et al. [31]
Machine Learning	98%	Only English language-based databases have been analyzed. No analysis on event-based or emotion-specific papers	Work on the disadvantages and research gaps found during their analysis	Raisa [32]
LSTM, CNN	83, 83.34%	Full stops, commas, and exclamation marks were discarded which could be helpful for sentiment analysis	Tweets always do not have a sentiment; therefore, they can be divided into an emotion ranging from -2 to +2	Kariya and Khodke [33]
ULMFiT-SVM	99.78%	Sentiment analysis was restricted to document level	Plan to take sentiments needs to be taken into account at the aspect level	AlBadani [34]
Naïve Bayes, Support Vector Machine (SVM)	78.01, 78.37% respectively	Cannot perform well on increasing the number of classes and the model is also not tested on any specific domain	The model can be used for sentiment analysis in different domains according to the need	Kumar [34]

Table 1 (continued)

Algorithm used	Accuracy	Limitation	Future work	References
Linear SVC, Logistic Regression	62, 76% respectively	Takes a longer time, when a model is not pretrained	Project a simple and human-centric approach for high-quality results	Pathak [35]
Partial Tree Kernel	2-way classification- 73.93% 3-way classification-60.60%	They have compared Twitter Sentiment analysis with sentiment analysis performed on other platforms but have not performed sentiment analysis on other platforms	Work with even richer linguistic analysis	Munson et al. [36]
SVM	82.9%	Twitter messages are short as displayed in emojis so obstacles in analyzing the accurate meaning of sentiments and detecting the suitable sentiment polarity	Sentiment analysis can be used for social media monitoring which provides a quantitative point of view and can explain real-time processing based on people's opinions	Bagheri and Islam [37]
Naïve Bayes algorithm	NA	Analysis and comparison of only one brand are displayed, while more comparisons would be appreciated	Will take the datasets of multiple apparel brands and implement a hybrid sentiment classification to conduct better accuracy by constructing an apparel domain dictionary	Rasool et al. [38]

they can be revoked if expired or lost. These keys are unique for every different user and are used to make API calls.

3. Creating a Flask environment in Spyder and importing libraries. Spyder is an expansive, cross-platform and fully accessible integrated developer environment (IDE) used mainly for scientific programming in python. Flask is a microweb

Table 1 (continued)

Algorithm used	Accuracy	Limitation	Future work	References
LSTM	Training accuracy: 88.1% Testing accuracy: 73.6%	Need to work on more sentiment classification other than anger, joy, fear, sadness	Will work on more complex sentiment investigation and work on improving accuracy and precision results	Pathak [35]
Naive Bayes classifier	80%	The bad performance was displayed during 3 classes ("positive", "negative" and "neutral")	Plan to collect Twitter data in multilingual languages and compare its characteristics	Pak and Paroubek [39]

technology based on python. It's termed a microframework since it doesn't require the use of any specialized methods or modules.

- Flask_restful is used to access the API and it also helps in completing the process.
 - Tweepy is a python library used for accessing the Twitter API.
 - Textblob is a python library for processing data and is also used for Natural Language Processing (NLP) tasks.
 - Flask_Cors is a flask extension used for resource sharing all over the globe.
4. Simultaneously, we created a webpage with the help of HTML/CSS and hosted the webpage on the server. After running the code, it will get redirected to the Twitter developer account. After searching the word in a webpage, it will then get redirected and all the tweets related to the word will be shown. The tweets shown on the screen will give us some information such as the Twitter handle of the person who has tweeted and the date and time when the person has tweeted. The tweets shown will also depict the percentage of tweets in Positive, Negative, and Neutral Sentiments related to that tweet. The results also display different types of languages and the percentage of the language of tweets in which Twitter users have commented. The output is shown in Fig. 3a.

By clicking on the tweet of the user, it gets redirected to the Twitter handle of the user and reflects its tweet which is shown in Fig. 3b.

3.2 Data Analysis

In today's world, many multinational companies play a major role in India's economy. Being in a limelight for many decades and making growth Year over Year, we have done data analysis of four such worldwide enterprises namely Reliance, Adani Group, Nike, and Adidas. Apart from them, during the pandemic time usage of OTT platforms

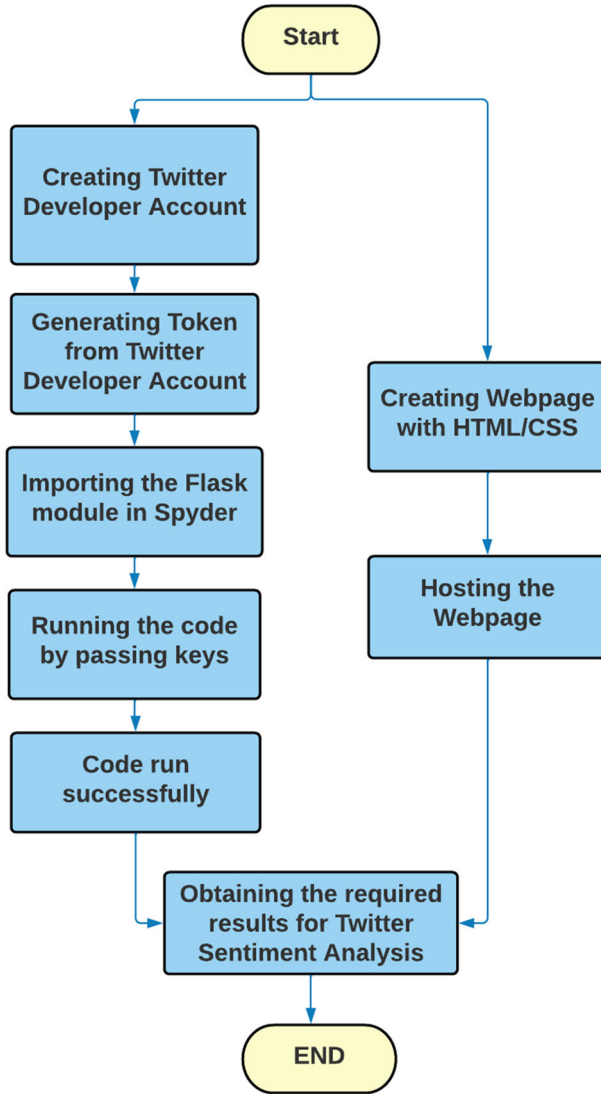


Fig. 1 Work flow diagram

has increased tremendously, so we have also considered sentiment analysis for two famous multimedia platforms, Netflix and Amazon Prime. The language code used for analysis is according to the ISO standard [40].

3.2.1 Scenario 1

In this scenario, the two biggest businesses in the Indian industry, Reliance and Adani are considered for sentiment analysis. For the year 2021, Reliance Industries accounted

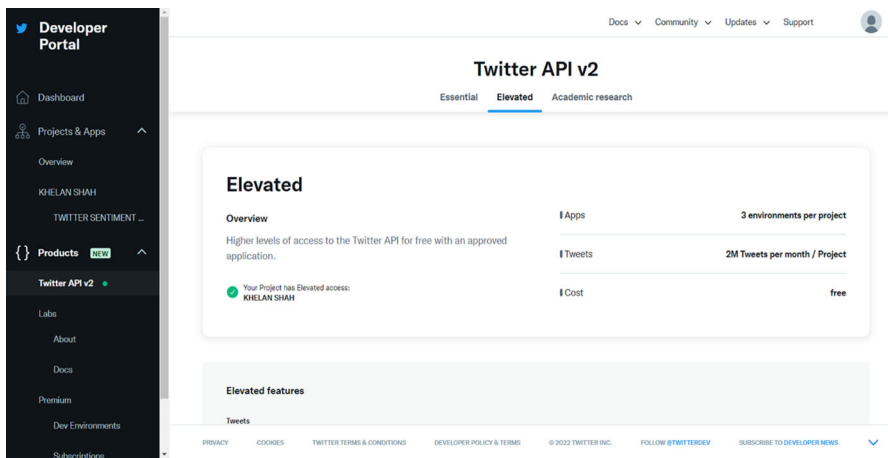


Fig. 2 Dashboard of developer portal

revenue of ₹2,60,485 Cr [41] while Adani enterprises accounted for ₹13,750 Cr [42]. Thus, considering the biggest industry runners, we obtained the below sentiment analysis.

Figures 4a, b, and 5a, b show the data analysis of Reliance Industries and Adani Enterprises respectively, which are the top two industry runners in India. While performing the sentiment analysis of the two companies as shown in Table 2, we found that 68.3% of tweets of Reliance and 67.5% of tweets of Adani Group are neutral. Even for the positive tweets, there is tough competition between Reliance Industries and Adani Enterprise, and Reliance Industries leads by 0.8%. For language analysis, 85% of tweets related to Reliance and 60% of tweets related to Adani Group are in English.

3.2.2 Scenario 2

In this scenario, the two biggest multinational footwear companies, Nike and Adidas are examined for sentiment analysis. For the year 2021, Nike accounted total revenue of 17.36 billion US Dollars [43] while Adidas accounted for 23.83 billion US Dollars [44]. Thus, considering the major footwear conglomerates, we performed the below sentiment analysis.

In the apparels section, we have performed sentiment analysis on Nike and Adidas which are shown in Figs. 6a, b, and 7a, b. The results mentioned in Table 3 show that there are 66.3% tweets in a neutral context, 21.30% tweets in a positive context, and the rest in a negative context for Nike. On the other hand, Adidas has 64.7% tweets in a neutral context, 21% tweets in a positive context, and the rest in a negative context. For Nike, the language analysis points out that 20% of tweets are in Japanese and 65% tweets in English, and the rest 15% for 3 different languages with 5% each, whereas for Adidas there are tweets in several languages. The major part is covered by English

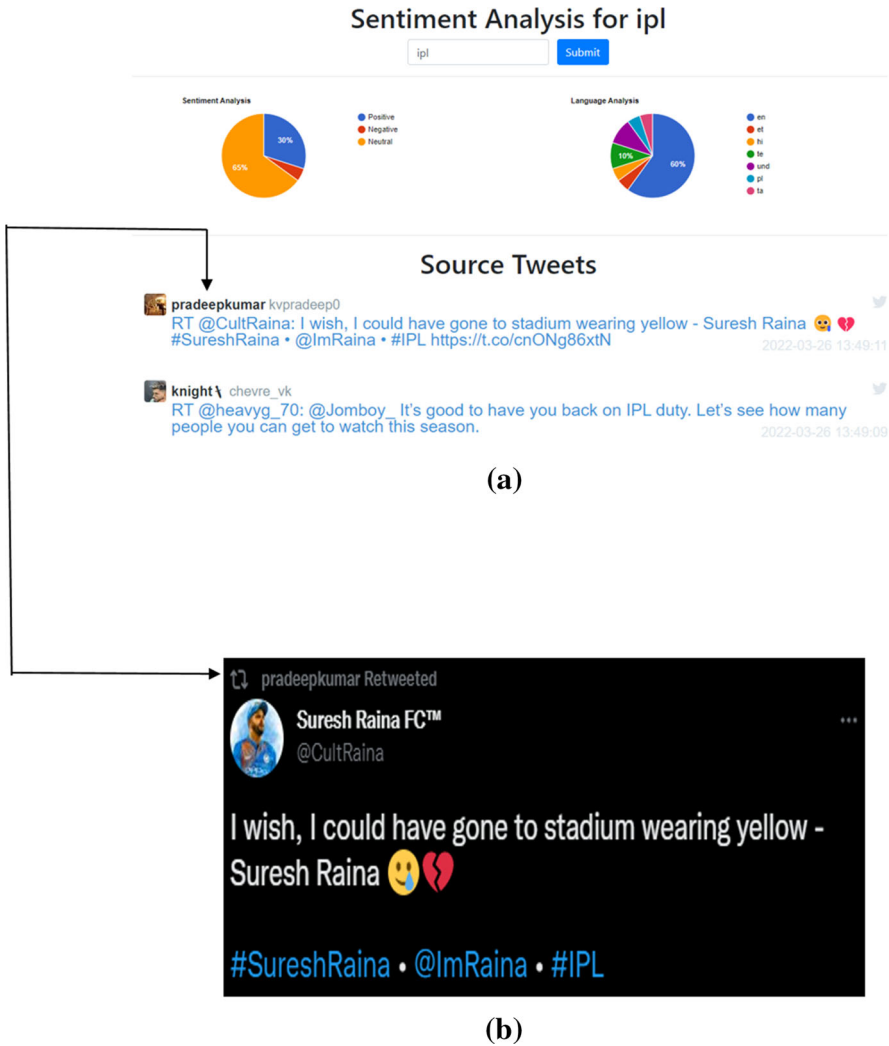
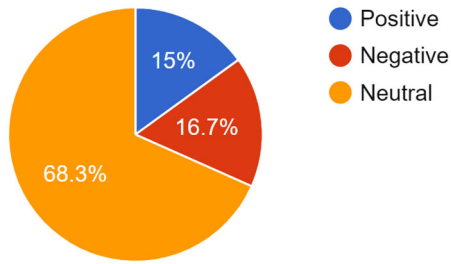


Fig. 3 a Output in the webpage. b Redirected the Twitter handle of the user

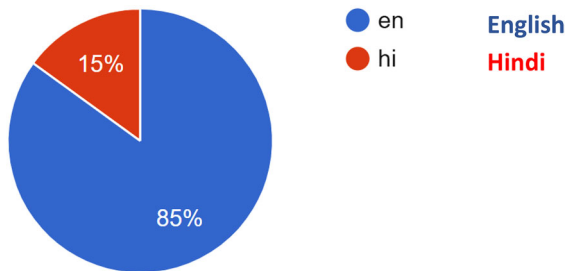
– 63.2%, the percentage of tweets in Japanese is 21.1% and the rest 15.7% is covered by other languages such as French, Spanish and Indonesian.

3.2.3 Scenario 3

In this scenario, the two largest OTT service providers, Netflix and Amazon Prime are considered for sentiment analysis. For the year 2021, Amazon Prime accounted for 11.9 billion US Dollars [45] while Netflix accounted for 7.7 billion US Dollars [46]. So, examining the major OTT platforms, we made the below sentiment analysis.

Sentiment Analysis

(a)

Language Analysis

(b)

Fig. 4 a Sentiment analysis of Reliance Industries. **b** Language analysis of Reliance Industries

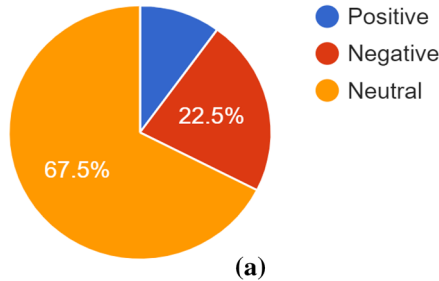
In the entertainment section, we have analyzed the rivalry between Netflix and Amazon Prime as shown in Figs. 8a, b and 9a, b. Table 4 shows the results that 66.7% of the total tweets for Netflix are neutral, 22% of total tweets are positive and 11.3% of total tweets are negative. For Amazon Prime, 66.2% of total tweets are neutral, 21.6% of total tweets are positive and 12.2% of total tweets are negative. For Netflix, 50% of total tweets are tweeted in English, 20% for Japanese, and 15% each for Thai and Indonesian but for Amazon Prime, the majority of tweets are posted in Japanese—45% and then comes English—40%.

Figure 10 represents the detailed sentiment analysis of all the scenarios under one bar chart. The bar chart shows the positive, negative, and neutral sentiments of all 6 companies where the positive sentiment is visualized using a blue bar, negative sentiment using a red bar, and neutral sentiment using a green bar. Apart from that, the table is also visualized in the chart which has the values of each sentiment in percentage for each company under the horizontal axis.

4 Challenges and Future Work

Sentiment analysis is an incredibly tricky task. We had to overcome several hurdles to complete this task. Some of them are discussed as follows. The foremost step in

Sentiment Analysis



Language Analysis

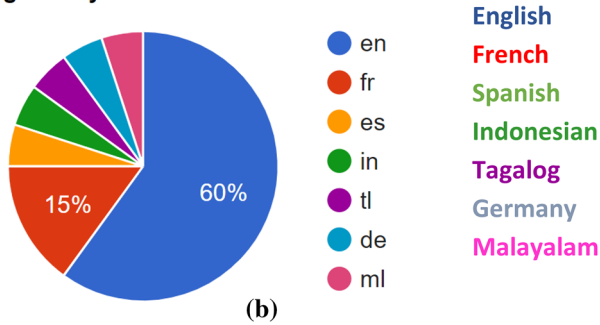


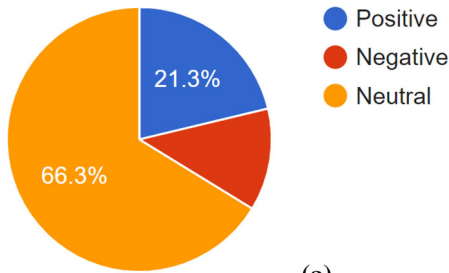
Fig. 5 a Sentiment analysis of Adani Enterprises. **b** Language analysis of Adani Enterprises

Table 2 Comparison of sentiment analysis of Multinational Enterprises

Company	Positive (%)	Negative (%)	Neutral (%)
Reliance Industries	15	16.70	68.30
Adani Enterprises	10	22.50	67.50

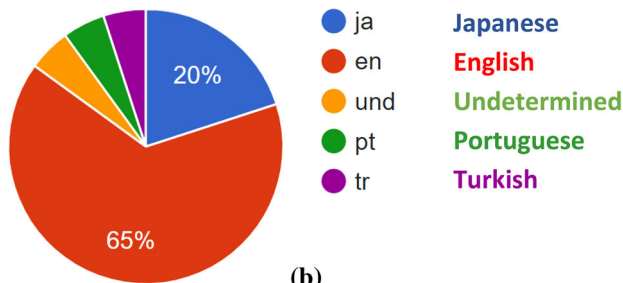
this project is to create a Twitter developer account. Without the account, the user will not be able to access the data. There might be chances that while creating the account Twitter denies access to the account and the user might not be able to proceed further with the project. The Twitter developer account allows to access only 500,000 tweets for analysis. This makes our dataset limited which can lead to a lack of information. This lack of information can produce biased results which affect the overall analysis providing skewed results. Moreover, this can be a roadblock in sentiment analysis as extracting information can become tedious and time-consuming. In addition to extracting data, data has to be verified and cleaned to be used. The output screen displays only twenty tweets. In addition to the limited number of tweets, there is no solid output. The output is in the form of visualizations which sometimes becomes difficult to interpret. Another challenge faced is opinion spamming. Competitors might utilize sentiment

Sentiment Analysis



(a)

Language Analysis



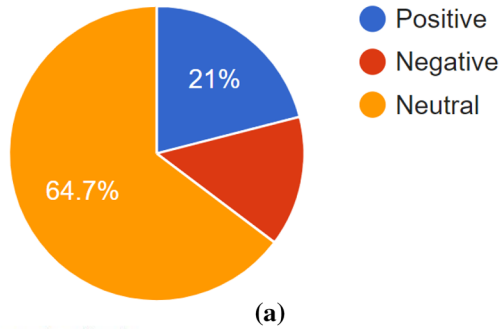
(b)

Fig. 6 a Sentiment analysis of Nike. **b** Language analysis of Nike

analysis to convey a poor picture of a firm. Once sentiment analysis becomes popular as a tool for gauging market success and corporate reputation, such methods may become quite widespread, resulting in a drop in the popularity of sentiment analysis.

This project provides us with polarity detection as well as language detection for a specific keyword but there can be some further enhancements that can be done in this project. The scope of this project is limited to textual data obtained from tweets on Twitter but can be further extended to pictures, videos, and many other formats. Another future suggestion would be to employ hashtags at the top of tweet emotions as text categorization characteristics. The code can be refurbished and can be written to reduce the complexity as well as make it more efficient. It will also provide analysis that can be easily comprehended and will have better visualizations. The static webpage can be converted to a dynamic webpage. Moreover, after the results are displayed the output can be converted to.pdf or.jpg formats so that they can be downloaded and shared among people. This analysis can be linked with other platforms such as Tableau, R, or Power BI which will amplify the visualizations making them more comprehensible and attractive to users. This project can be implemented in Spark as it is a growing technology. Spark also has some data visualization libraries which can make visualizations more appealing and can provide better accuracy. In the future, all of these attributes would be taken into account.

Sentiment Analysis



Language Analysis

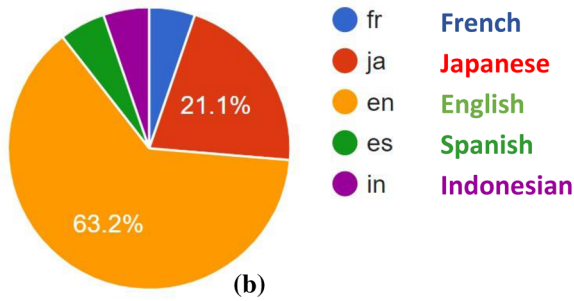


Fig. 7 a Sentiment analysis of Adidas. **b** Language analysis of Adidas

Table 3 Comparison of sentiment analysis of Apparels Brand

Company	Positive (%)	Negative (%)	Neutral (%)
Nike	21.30	12.40	66.30
Adidas	21	14.30	64.70

5 Conclusion

Twitter sentiment analysis is achieving greater heights in the analysis segment. People accessing Twitter globally and making opinions, and interpreting the data are salient. In this paper, we have developed a solitary way to analyze Twitter tweets using the flask environment. The tweets have been accumulated with the help of Twitter API and the Tweepy library. We have extricated the tweets and categorized them into positive, negative, and neutral. The results also mention the languages in which the tweets have been posted. The webpage has also been created which is linked to the python code and the results can be obtained by passing the word or phrase in a webpage. The desired results are obtained and also the Twitter handle of the user and the date and time of the tweet get known. We have used the Textblob library which is specially used in the analysis part as the portion of preprocessing of data gets mitigated. The primary

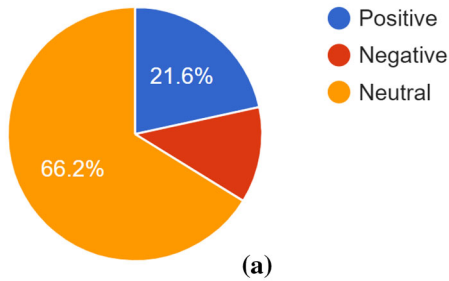
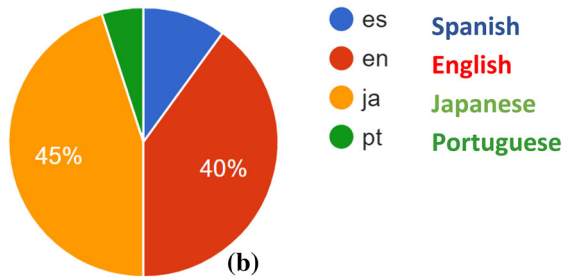
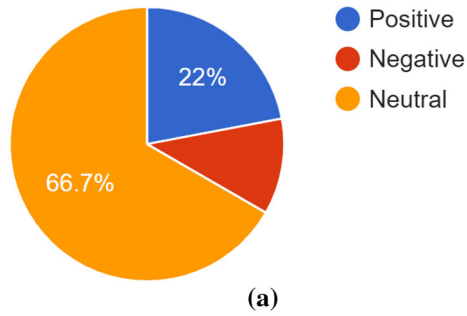
Sentiment Analysis**Language Analysis**

Fig. 8 a Sentiment analysis of Amazon Prime. b Language analysis of Amazon Prime

motive for developing using the flask environment is not to get involved in machine learning techniques. Working using a flask environment, we provide better results as the training and testing part of the model will be terminated. By this, the output will be 100% efficient as it does not require a machine learning algorithm and depends on its accuracy. The idea can take real-time analysis and methodologies to the next level. A different perspective to analyze the tweets can be obtained. Well, the industry is open to all new advancements that can make the work experience more exciting for both the players as well as the spectators. The system proposed is pivotal and striking. It can be easily deployed in various sectors, so companies can take maximum benefit from the proposed work.

Sentiment Analysis



Language Analysis

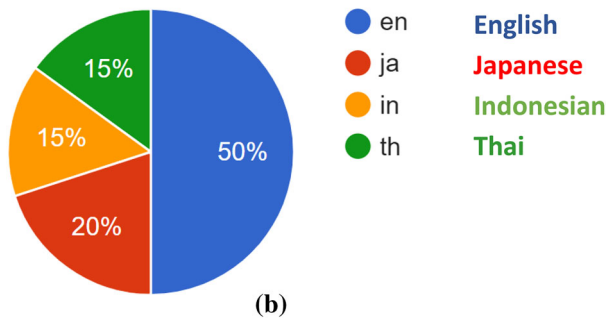


Fig. 9 a Sentiment analysis of Netflix. **b** Sentiment analysis of Netflix

Table 4 Comparison of sentiment analysis of OTT platforms

Company	Positive (%)	Negative (%)	Neutral (%)
Amazon Prime	21.60	12.20	66.20
Netflix	22	11.30	66.70

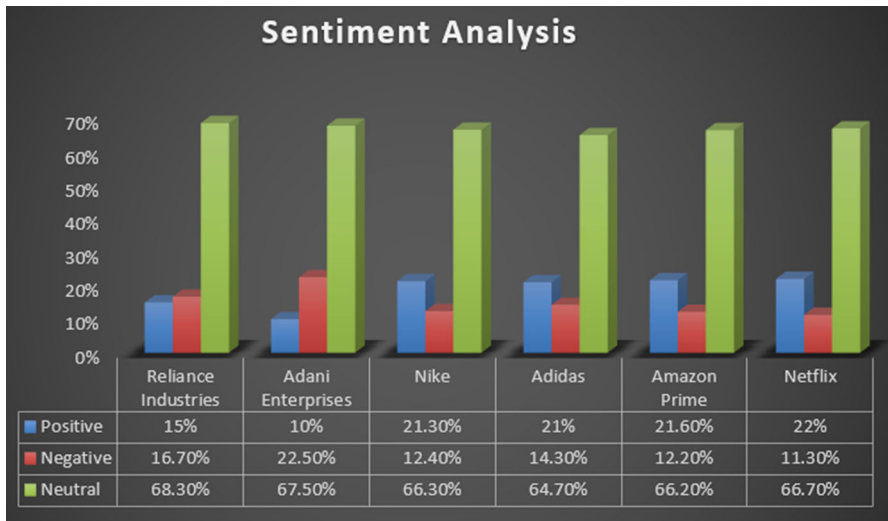


Fig. 10 Detailed comparison of sentiment analysis of all scenarios

Acknowledgements The authors are grateful to Department of Computer Science and Department of Information and Communication Technology, School of Technology and Department of Chemical Engineering, School of Energy Technology, Pandit Deendayal Energy University for the permission to publish this research.

Author Contributions All the authors make substantial contribution in this manuscript. AM, KS, SS, SP and MS participated in drafting the manuscript. AM, KS, SS wrote the main manuscript, all the authors discussed the results and implication on the manuscript at all stages.

Funding Not applicable.

Availability of Data and Materials All relevant data and material are presented in the main paper.

Declarations

Conflict of interest The authors declare that they have no conflict of interests.

Consent for Publication Not applicable.

Ethical Approval and Consent to Participate Not applicable.

Code Availability Not Available.

References

1. Wang Q, Ma Y, Zhao K, Tian Y (2022) A comprehensive survey of loss functions in machine learning. *Ann Data Sci* 9:187–212. <https://doi.org/10.1007/s40745-020-00253-5>
2. Shi Y, Tian YJ, Kou G, Peng Y, Li JP (2011) *Optimization-based data mining: theory and applications*. Springer, Berlin

3. Olson DL, Shi Y (2007) Introduction to business data mining. McGraw-Hill/Irwin, New York
4. Shi Y (2022) Advances in big data analytics: theory, algorithm, and practice. Springer, Singapore
5. Güven ZA, Diri B, Çakaloğlu T. Comparison Method for Emotion Detection of Twitter Users. In2019 Innovations in Intelligent Systems and Applications Conference (ASYU) 2019 (pp. 1-5). IEEE.
6. Tien JM (2017) Internet of things, real-time decision making, and artificial intelligence. *Ann Data Sci* 4:149–178. <https://doi.org/10.1007/s40745-017-0112-5>
7. Kharde P, Sonawane S (2016) Sentiment analysis of Twitter data: a survey of techniques. *Int J Comput Appl* 139:5–15. <https://doi.org/10.5120/ijca2016908625>
8. Ortega JL (2017) The presence of academic journals on Twitter and its relationship with dissemination (tweets) and research impact (citations). *Aslib J Inf Manag* 69:674–687. <https://doi.org/10.1108/AJIM-02-2017-0055>
9. Anber H, Salah A, El-Aziz AAA (2016) A literature review on Twitter data analysis. *Int J Comput Electr Eng* 8:241–249. <https://doi.org/10.17706/ijcee.2016.8.3.241-249>
10. Singh M, Goyal V, Raj S (2021) Sentiment analysis of social media Tweets on Farmer Bills 2020. *J Sci Res* 65:156–162. <https://doi.org/10.37398/jsr.2021.650319>
11. Alsaeedi A, Khan MZ (2019) A study on sentiment analysis techniques of Twitter data. *Int J Adv Comput Sci Appl* 10:361–374. <https://doi.org/10.14569/ijacsa.2019.0100248>
12. Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. 4:177–181. <https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf>
13. Sharma A, Ghose U (2020) Sentimental analysis of Twitter data with respect to general elections in India. *Procedia Comput Sci* 173:325–334. <https://doi.org/10.1016/j.procs.2020.06.038>
14. Sarlan A, Nadam C, Basri S (2015) Twitter sentiment analysis. InProceedings of the 6th International conference on Information Technology and Multimedia 2014 (pp 212–216). DOI: <https://doi.org/10.1109/ICIMU.2014.7066632>
15. Khan L, Amjad A, Afaq KM, Chang H (2022) Deep sentiment analysis using CNN-LSTM architecture of English and Roman Urdu text shared in social media. *Appl Sci*. <https://doi.org/10.3390/app12052694>
16. Wongkar M, Angdresey A (2019) Sentiment analysis using Naive Bayes algorithm of the data crawler: Twitter. InProceedings 2019 4th International Conference Informatics Computing ICIC 2019. <https://doi.org/10.1109/ICIC47613.2019.8985884>
17. Kasthuri S, Jebaseeli AN (2020) An efficient decision tree algorithm for analyzing the twitter sentiment analysis. *J Crit Rev* 7(4):1010–1018
18. Naw N (2018) Twitter sentiment analysis using support vector machine and K-NN classifiers. *Int J Sci Res Publ*. <https://doi.org/10.29322/ijrsrp.8.10.2018.p8252>
19. Anupama BS (2020) Real time Twitter sentiment analysis using natural language processing. *Int J Eng Res* 9:1107–1112. <https://doi.org/10.17577/ijertv9is070406>
20. Shahzad M, Alhoori H (2022) Public reaction to scientific research via Twitter sentiment prediction. *J Data Inf Sci* 7:97–124. <https://doi.org/10.2478/jdis-2022-0003>
21. Al-Hashedi A, Al-Fuhaidi B, Mohsen AM et al (2022) Ensemble classifiers for Arabic sentiment analysis of social network (Twitter data) towards COVID-19-related conspiracy theories. *Appl Comput Intell Soft Comput*. <https://doi.org/10.1155/2022/6614730>
22. Kumar Singh S, Verma P, Kumar P, Abdul A (2020) J Crit Rev Sentim Anal Covid-19 Epidemic Mach Learn Algorithms Twitter 7:2020
23. Deshpande P, Joshi P, Madekar D et al (2019) A survey on: classification of Twitter data using sentiment analysis. *Asian J Converg Technol* 5:34–37
24. Kavitha P (2021) Twitter sentiment analysis based on adaptive deep recurrent neural network. *Turk J Comput Math Educ (TURCOMAT)* 12:2449–2457
25. Dabak MS (2021) Sentiment analysis of Twitter data by using deep learning And machine learning. *Turk J Comput Math Educ*. 12(6):962–970. <https://doi.org/10.17762/turcomat.v12i6.2375>
26. Arun K, Srinagesh A (2020) Multi-lingual Twitter sentiment analysis using machine learning. *Int J Electr Comput Eng* 10:5992–6000. <https://doi.org/10.11591/ijece.v10i6.pp5992-6000>
27. Reddy AB, Vasundhara DN, Subhash P (2019) Sentiment research on Twitter data. *Int J Recent Technol Eng* 8:1068–1070. <https://doi.org/10.35940/ijrte.B1181.0982S1119>
28. Gupta B, Negi N, Vishwakarma K et al (2017) Study of Twitter sentiment analysis using machine learning algorithms on Python. *Int J Comput Appl* 165:29–34. <https://doi.org/10.5120/ijca2017914022>
29. Hasan A, Moin S, Karim A, Shamshirband S (2018) Machine learning-based sentiment analysis for Twitter accounts. *Math Comput Appl* 23:11. <https://doi.org/10.3390/mca23010011>

30. Kolchyna O, Souza TTP, Treleven P, Aste T (2015) Twitter sentiment analysis: Lexicon method, machine learning method and their combination
31. Jalil Z, Abbasi A, Javed AR et al (2022) COVID-19 related sentiment analysis using state-of-the-art machine learning and deep learning techniques. *Front Public Heal* 9:1–14. <https://doi.org/10.3389/fpubh.2021.812735>
32. Raisa JF, Ulfat M, Al Mueed A, Reza SS. A review on Twitter sentiment analysis approaches. In 2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD) 2021 (pp. 375-379). IEEE. <https://doi.org/10.1109/ICICT4SD50815.2021.9396915>
33. Kariya C (2020) Khodke P (2020) Twitter sentiment analysis. *Int Conf Emerg Technol INCET 2020*:1–17. <https://doi.org/10.1109/INCET49848.2020.9154143>
34. AlBadani B, Shi R, Dong J (2022) A novel machine learning approach for Sentiment analysis on Twitter incorporating the universal language model fine-tuning and SVM. *Appl Syst Innov*. <https://doi.org/10.3390/asi5010013>
35. Pathak S (2020) Twitter Sentiment analysis using different algorithms. *Int J Res Appl Sci Eng Technol* 8:1023–1026. <https://doi.org/10.22214/ijraset.2020.31647>
36. Munson E, Smith C, Boehmke B, Freels J (2019) Sentiment analysis of Twitter data (SAOTD). *J Open Source Softw* 4:764. <https://doi.org/10.21105/joss.00764>
37. Bagheri H, Islam MJ (2017) Twitter sentiment analysis. 8:1–2. DOI: <https://doi.org/10.31219/osf.io/6xc4y>
38. Rasool A, Tao R, Marjan K, Naveed T (2019) Twitter sentiment analysis: a case study for apparel brands. *J Phys Conf Ser*. <https://doi.org/10.1088/1742-6596/1176/2/022015>
39. 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'10) 2010*: 1320–1326. DOI: <https://doi.org/10.17148/ijarcce.2016.51274>
40. <https://www.andiamo.co.uk/resources/iso-language-codes/>. Accessed 2nd Oct 2022
41. <https://www.moneycontrol.com/financials/relianceindustries/profit-lossVI/RI>. Accessed 1st Oct 2022
42. <https://www.moneycontrol.com/financials/adanienterprises/profit-lossVI/AE13>. Accessed 10th Oct 2022
43. <https://www.statista.com/statistics/888676/nikes-revenue-in-theus/#:~:text=In%202021%2C%20Nike's%20U.S.%20revenue,about%2017.36%20billion%20U.S.%20dollars>. Accessed 9th Oct 2022
44. <https://companiesmarketcap.com/adidas/revenue/>. Accessed 9th Oct 2022
45. <https://backlinko.com/amazon-prime-users>. Accessed 11th Oct 2022
46. <https://www.statista.com/statistics/273883/netflixs-quarterly-revenue/>. Accessed 11th Oct 2022

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.