

## Opinion Mining Using Social Media

**K.Ruth Ramya<sup>1</sup>, Chandrika.D<sup>2</sup>, Nikhitha. K<sup>3</sup>, Niteesha.T<sup>4</sup>, Pravallika.P<sup>5</sup>**

<sup>1</sup> Assistant Professor, Department of CSE, KoneruLakshmaiahEducation Foundation, Vaddeswaram, India-522502

<sup>2,3,4,5</sup> Student, Department of CSE, KoneruLakshmaiahEducation Foundation, Vaddeswaram, India-522502

### ABSTRACT

A significant piece of our data gathering conduct has consistently been to discover what others think. With the creating availability and noticeable quality of feeling rich resources, for instance, online review areas and individual web diaries, new possibilities and challenges arise as people at present can, and do, successfully use information developments to look out and appreciate the evaluations of others. The sudden eruption of exercise in the realm of sentiment evaluation, which offers with the computational remedy of opinion, sentiment, and polarity in text, has thus occurred no less than a direct response to the search of curiosity that offers instant fruitful results with opinions as a foremost entity. Telegram Messenger is one of the best instant and expressive platforms for messaging and communicating. The telegram bot is operated by software that has AI features where they can be used for broadcasting and also helps to integrate with other services. This paper is our attempt to use a Telegram bot that interacts with the users when asked about the review on any product or movie or celebrity or personality or disease and many more.

**Keywords:** BotFather, Natural Language Processing, Python Telegram Bot, Sentimental Analysis, TextBlob, Tweepy

### I. INTRODUCTION

In the contemporary days, humans cannot conclude a decision without the opinion of others. For a given volume of reviews, the reviewer has to experience numerous individual reviews which require an ample amount of time as well as tremendous human processing which might or might not lead to an efficient conclusion. Instead, more interesting applications and developments came into existence in the area of Artificial Intelligence to make computers able to analyze the study of human emotions to extract an analyzed and reliable opinion on any discussion[1].

Promoting on social media is one of the predominant techniques an organization opts to advertise its product. It now has been so valuable for these organizations to know how their products are carrying out or how people are responding to them and as it's important to connect with customers, social listening helps in keeping track of conversations that happen on social media platforms[2]. The instant thought we had in our brains once we began, was to build a working model to foretell the polarity of the text and the sentiment whose attributes are positive, negative, and neutral in reach of every individual. Twitter, LinkedIn, Instagram, etc., occur to be a tremendous platform for folks to go and write about something that's how they genuinely feel or what their view of a convincing product or personality is.

Instant messaging and content production websites, such as Twitter is one of the extremely essential sources for feedback. Since this API is clean and comes with rich developer resources, Twitter proves to be a great source of data for interpreting. The knowledge is rich in data and has a data format appropriate for analysis[3]. For anyone with fair utilization rights, Twitter data is accessible. And the fact that all people post their opinions on a variety of topics, discuss current issues, complaints, and exhibit positive sentiment for products that they use in daily life,

although hashtags,[4] likes and shares would predominantly explain how well your product is growing cannot be deniable.

Sentiment Analysis is the procedure of deriving reliable information from the textual context to get reviews about numerous aspects like films, stardoms, products, political conversations, and many more[5]. Opinion Mining also uses Natural Language Processing and Machine Learning to help organizations look far beyond just the number of tweets they get on the trending topics. This helps us to understand the accurate score or review of any particular domain which gets updated for each and every second.

## II. LITERATURE SURVEY

The globe has been globalized with the presence of cyberspace. Anyone can access any information they need without moving from their place. There are many online networks where people share their thoughts, views, emotions and opinions openly. From this data many had started working on Sentimental analysis or opinion mining as in N. Yadav et.al., 2020 [6]. This technical paper had used the twitter dataset collected from kaggle to evaluate the emotional AI in order to give feedback of products from the tweets. This model was designed using the classifier algorithms which analyzes the products to know the behaviour or status which helps both the clients and the customers to know the feedback about sales and products in the present market. Huge number of tweets about numerous products have been used to analyze them if they have a positive sentiment or neutral sentiment or negative sentiment by comparing all the classifier techniques. This existing model haven't applied conventional text classification and utilized sentiment features to give high precision.

El Rahman, S. A et.al., 2019 [7] concentrates on processing tweets available in English. The data extraction from twitter is accomplished via R programming language and Twitter API. The "wordcloud" package in R programming language is utilised in this paper to trigger a word cloud from the data extracted as text from twitter which makes more significance about the routinely used terms on a particular subject. The extracted unstructured data is cleaned by transition of the text into a dataframe which abolishes text URLs, eliminates pause phrases, numbers, unnecessary spaces, usernames and punctuations. A lexicon based model is used which consists of two text files which contains a list of positive and negative words to identify the sentiment of the tweet. Various supervised and unsupervised machine learning techniques are used for training and testing of the dataset.

The study of emotions and analyzing the texts has made us enthusiastic to work on to try something new with the technology to help everyone. We scrutinized the ideas in order to advance the existing models using the latest data with help of Twitter API's which collects most recent tweets every time a person checks reviews. Not only the review and score of products like in previous model one can use this enhanced model to know the overall opinions and sentiment/status of the input. The main add on is that since telegram has become more popular these days people can check the reviews about their desired search from telegram itself. We had created a bot which searches the users required information without opening twitter. The persons who have no twitter to know the opinions of others can also use this bot which helps them to identify their reviews.

Dashtipour. K et.al., 2016 [8] proposed a model which compares with various other previous models to predict the sentiment scores of the information which is not restricted to only one language. This multi-lingual analysis is applicable for all other languages such as German, Arabic, Spanish Z. Lin et.al., 2014 [9] etc., The disadvantage of our model is that our proposed telegram bot takes only English language as it's input since it would be difficult to work on both telegram and twitter if we need to perform multi-dialect opinion mining. And the other thing we noticed from our model is that if the required search of the specific user has no tweets it shows the result as neutral. Our main aim is to make opinion mining easier who uses telegram to know the feedback of their desired search since English is used as Universal language.

### III. THEORETICAL ANALYSIS

The results of analyzing the sentiment is divided into several subtasks which are needed to be performed using various approaches and techniques. In this paper such terms would include tweepy, textblob, NLP and Opinion Mining which are explained in detail.

#### 3.1 TWEETPY

This feature enables Python to use its API to browse the Twitter platform/database. Tweepy is an open source software that offers you a very convenient way to access the Web Twitter Data. Tweepy provides a collection of modules and procedures that depict Twitter templates and API end devices, and manages different mitigation information, including such: data encoding and decoding, in a timely basis. Tweepy promotes connectivity to Tweets through Standard Encryption and the modern OAuth framework. Twitter has refused to allow Simple Authentication, so now OAuth is the only way to access the Twitter Data. The key drawback of the simple authentication is that the Twitter data is achievable through a username and a password which cannot have high security chances. So, after 2010 Twitter began to mandate OAuth which works with access keys that are highly secure. The incentives of the OAuth implementation are feasible. This model does not disclose the user credentials rendering it more reliable. Licenses are easier to configure that is a collection of tokens and keys can be generated that only works from timeframes, so in the circumstance that anyone possesses these keys, they won't be capable of writing or sending instant messages thus mitigating the danger. The implementation does not react to login credentials, so although the user replaces it the implementation will still launch.

#### 3.2 TEXTBLOB

TextBlob is a library that is present in Python for handling textual data. TextBlob is constructed upon NLTK and provides a straightforward to use interface to the NLTK library. It helps in diving into familiar natural language processing (NLP) jobs like part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation, and many more.

We can compute the sentiment using TextBlob/Vader. Depending on the polarity and subjectivity, one can evaluate whether the tweet is a positive text or negative text or neutral text. A good issue regarding TextBlob is that they're rather like python strings. So, you'll be able to remodel and play with it in the same way like we have a tendency to do in python. For TextBlob, if the polarity is  $>0$ , it is observed as positive,  $<0$  is examined as negative and  $=0$  is considered neutral.

#### 3.3 NATURAL LANGUAGE PROCESSING

The preprocessed data that is collected from the twitter is majorly labelled and divided into four phases. Generally the framework includes removing spaces, punctuation, normalizing the data into a single case. Preprocessing of the data is almost labelling the sentence segmentation which includes formatting necessary tagging and summarization where a given text is divided into sentences, tokenization involves a task where the sentence texts are improved into words that are distinct by splitting the words that has spaces, blanks or any punctuation marks such as colon, comma etc. are the primary attributes that are divided into tokens, removing stop words as they give no meaning to the whole sentence when framed. It can include various conjunctions or unnecessary prepositions or terms which eliminates the suffixes and word stemming.

#### 3.4 OPINION MINING

Opinion mining, or sentiment analysis, could be a text analysis technique that uses linguistics and language process to mechanically establish and extract sentiment or opinion from among text (positive, negative, neutral, etc). In order to produce commodities and products that suit the user requirements, you need to get into the minds of the consumers and figure out their likes and dislikes. Using certain machine learning techniques, opinion mining can be performed on a highly unstructured data with less human intervention. Sentiment analysis can analyze multiple sites, tweets, comments, or polls for significant results in a short span of time.

### IV. PROPOSED METHODOLOGY

For a producer, it is important to know how the consumers feel about their product for achieving profits and customer's expectations. In the case of politics, it is predominant to perceive the democratic will of the public. Accordingly, a survey on the sentimental score of their product or political party is necessary. This paper defines an own sentimental analyzer that is built using a telegram bot by feeding the topic into the bot and executes the sentiment of the topic from the latest tweets. To extract the sentiment of a topic from the social media platform,

Twitter is opted because politicians, notable figures and indeed common people regularly update their opinions in the form of tweets.

In an attempt to provide the users with reliable outputs, up-to-date tweets are extracted as a dataset through Twitter API. The Twitter API is an application programming interface of twitter that enables authenticated users to retrieve massive information on the needed subject. In terms of being able to acquire the twitter dataset, the initial step to be followed is to sign up for a developer account on twitter. A questionnaire will then be shown. Fill up the essential fields and transmit the information. This phase reroutes to the API Console of the twitter platform. Choose a project or build a new one from the givenFrom the API keys projects. Select Create credentials and then API keys which will redirect to the 'Consumer Keys' column, which contains the link to twitter data access i.e API key or consumer key, API secret key or consumer secret key, access token and access token secret that will be generated after you create any project. Secure the produced keys. Make a link to access the twitter dataset through an authenticated definition with the inputs namely API key, API secret key, access token and access token secret.

Figure 1: Twitter API Keys Generation

Figure 1 depicts the page where the consumer keys of a project are obtained. The page involves API key and API secret keys, Bearer token and Access token and Access Secret token.

In order to build a message bot using telegram, we need to trigger tokens for activation of the telegram API. The telegram API is a free open-source to create any applications on their platform. It provides an interface to the users to freely create programs and bots. We can run our own bot API server but instead we choose to use BotFather, a transfer bot ownership where we send or shift the existing bot to another telegram account. Thus, by generating a token to authenticate the bot, users can send requests to the bot API. Secure the generated token to update the obtained outcome to the telegram bot.

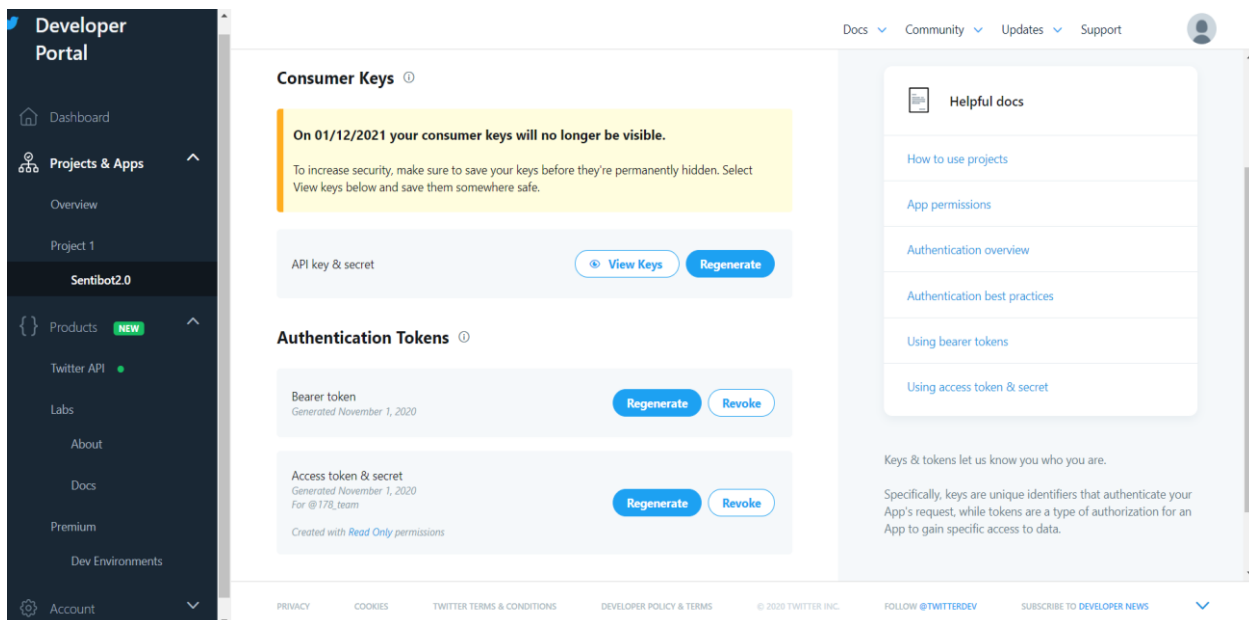




Figure 2: Create a new account in Telegram Bot

Figure 2 depicts the page where the API key of the Bot is generated through BotFather in Telegram. The page involves the unique API key which is created for Telegram Bot where the username is labelled as “OpinionMining\_bot”.

Before we analyze the tweet sentiment of the tweet accessed through the Twitter API keys we need to pre-process the data in other terms can also be called as cleansing of the tweets, it is the process of removing incorrect or unnecessary data, so that the result does not produce false analysis or wrong decisions. This stage implies numerous phases which eliminates the links, usernames and numbers and lastly, translation of uppercase scripts to lowercase scripts and removing spaces using the nltk and regex libraries and word tokenizer. For achieving the sentiment of a tweet that is polarity of a tweet which lies in the range of -1 to 1, where less than or equal to -0.25 is a negative tweet and less than or equal to 0.25 is a neutral tweet else a positive tweet in the case of this paper using Textblob library. To produce a reliable outcome, the average polarity of all the latest tweets including the discussion are calculated.

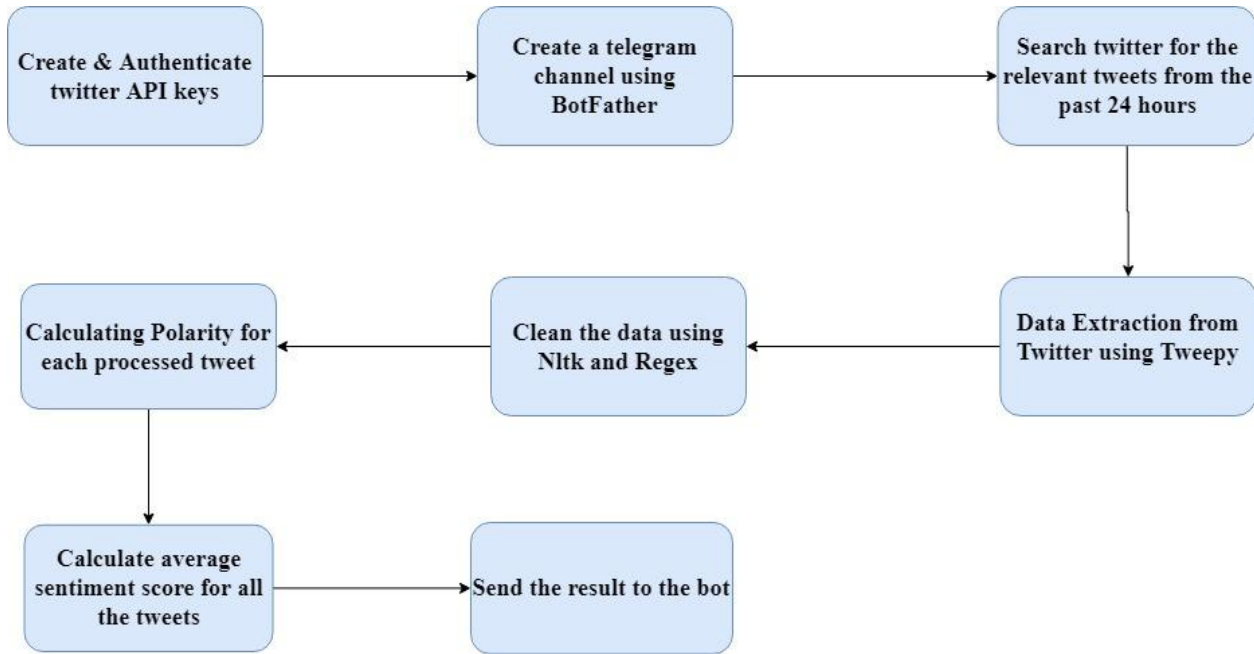


Figure 3: WorkFlow to connect Telegram bot and Twitter developer account

Figure 3 workflow depicts the overall development of the opinion mining model. Given any keyword this model determines whether the given subject is positive, negative or neutral.

To obtain the calculated sentiment score and the sentiment of a discussion in the telegram bot Updater, MessageHandler, Filters and CommandHandler classes from the telegram.ext class are applied. The Updater class supplies the software developer with a presentation layer, so they can concentrate solely on the bot's scripting. Its aim is to acquire Telegram updates and transmit them to the dispatcher. The dispatcher facilitates handlers for disparate data types: Telegram alerts, simple text commands, and even arbitrarily defined types. The purpose of MessageHandler Class is to manage telegram messages. They may contain modifications to text, media or profile. The mission of the CommandHandler class is to supervise telegram commands that initiate with a '\ ' and selectable approach by '@ bot's name' and/or any supplementary script. The filter argument is used with the MessageHandler class.

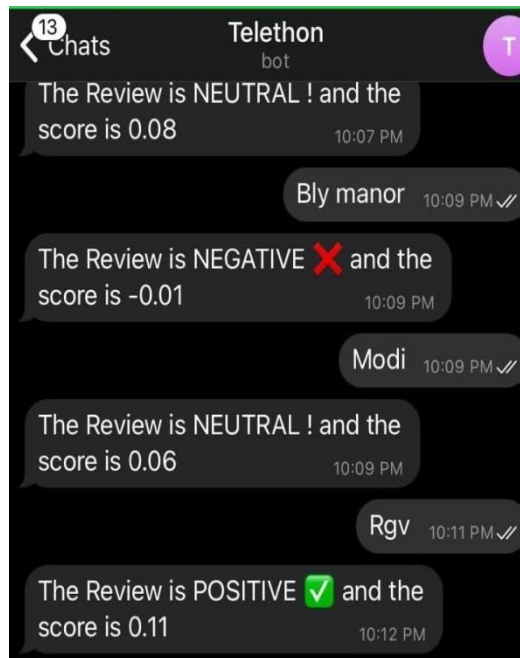
After executing the algorithm, flip over to the telegram bot created using the step mentioned in the methodology and provide any subject of interest as the input. This provides an ultimate end result which provides the users with an effective and reliable service. The outcome of the given input will be the sentiment score and the sentiment of the given input keyword i.e either negative or neutral or positive.

## V. RESULTS

As our working model considers a keyword on the area of the interest of the client, it produces a reliable outcome as the sentiment of the input based on the latest available tweets from twitter. The sentiment is classified into three attributes namely positive, negative and neutral based on the calculated emotional score derived from the tweets. The out-turn produced in the telegram bot displays both score and sentiment of the given input.

The result obtained in the telegram bot is the sentiment of the input provided by the user. The initial stage is the cleansing of tweets related to the keyword given by the user which includes the removal of numbers, spaces, links and usernames through the natural language toolkit. The sentiment that is the polarity of the given keyword is calculated by using the TextBlob library. The sentiment calculated is

further categorized into three types for user understandability. The output can either be positive, negative or neutral based on the range of the sentiment score. Thus the displayed output contains the polarity that is the sentiment score and sentiment of the input keyword.

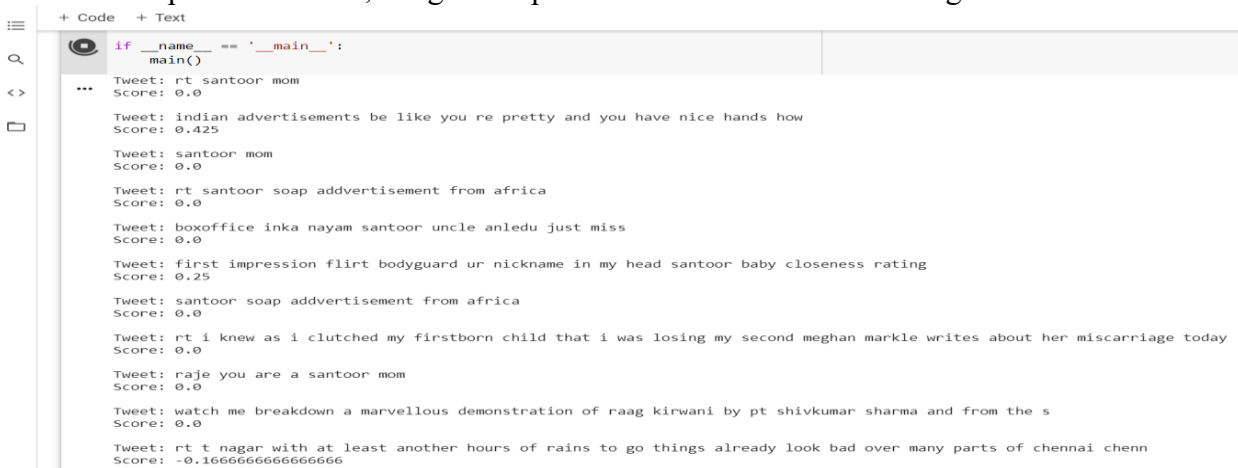


**Figure 4: Results displayed in Telethon Bot**

Figure 4 presents the output of the user's desired input keyword which displays the sentiment score and sentiment.

### 5.1 For a positive review

To achieve a positive review, the given input's sentiment score should range between 0.25 and 1.



**Figure 5(a) :Individual Tweet Score**

Figure 5(a) displays all the tweets which includes the input and the score of each tweet.

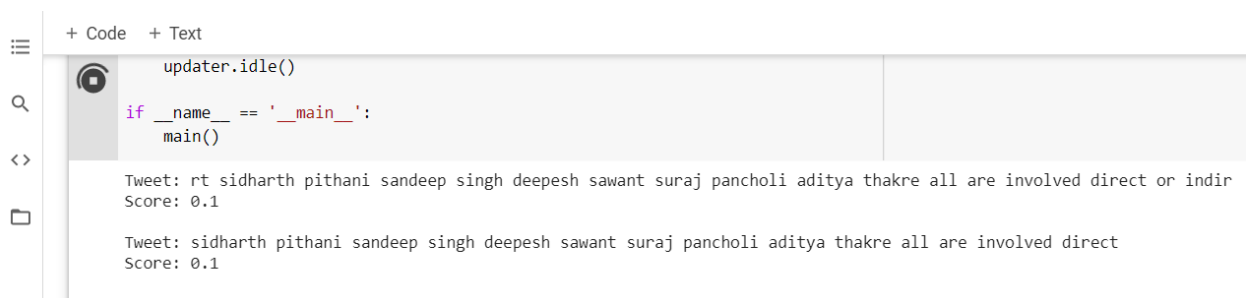


**Figure 5(b) : Final score of the input**

Figure 5(b) shows the user required output for the search keyword “Santoor”

### 5.2 For a negative review

To achieve a negative review, the given input’s sentiment score should range between -0.25 to -1

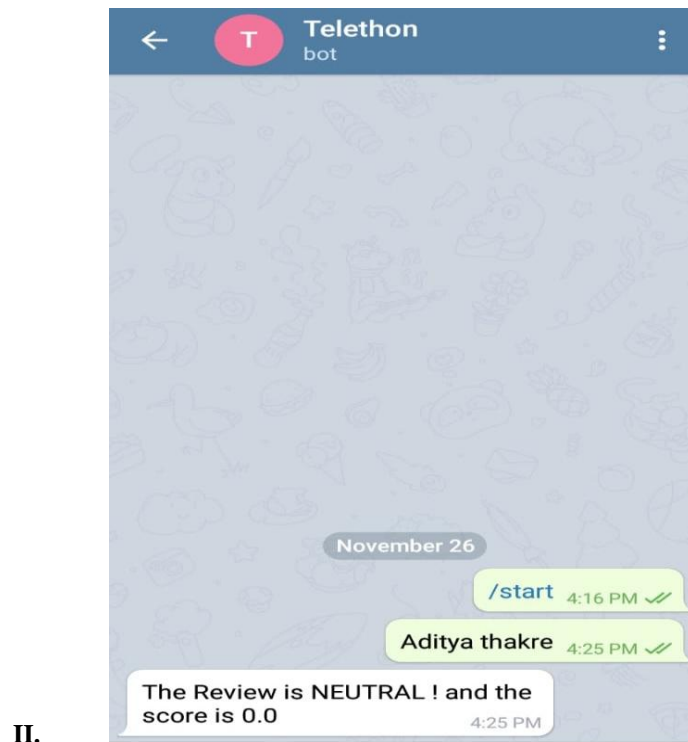


I.

**Figure 6(a) : Individual Tweet Score**

Figure 6(a) displays all the tweets which includes the input and the score of each tweet.



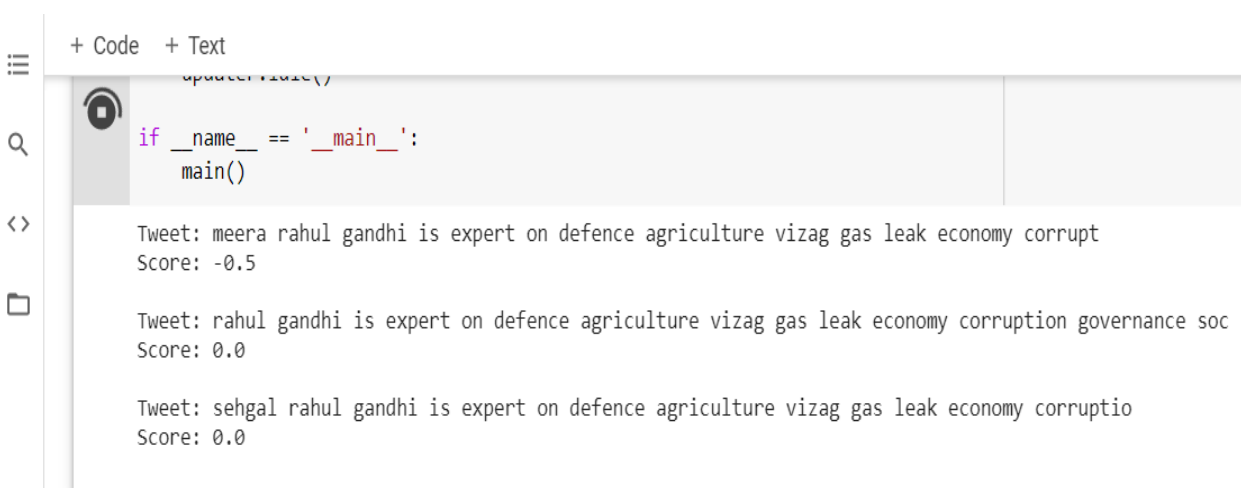


**Figure 6(b) : Final score of the input**

Figure 6(b) shows the user required output for the search keyword “Aditya Thakre”.

### 5.3 For a neutral review

To achieve a neutral review, the given input’s sentiment score should range between



**Figure 7(a) : Individual Tweet Score**

Figure 7(a) displays all the tweets which includes the input and the score of each tweet.



**Figure 7(b) : Final score of the input**

Figure 7(b) shows the user required output for the search keyword “Vizag gas leak”

## VI. CONCLUSION & FUTURE SCOPE

Recently, people have commenced conveying their views on the Web that enhanced the need of scrutinizing the texts of users for several here-and-now applications. In this project, we created a telegram bot inter-connecting the python which helps users to get the desired review of the trending issues, products, personalities, or about anything from Twitter’s latest data or tweets using Twitter API’s every time they want to check out the opinion of what others feel. It’s been observed that the review and score vary from time to time for the same search as there is no constant dataset since it gets updated every time. The future enhancements may include to add advanced features which accepts numerous languages to analyze the text and to give accurate score.

## REFERENCES

1. S. Shayaet al., "Sentiment Analysis of Big Data: Methods, Applications, and Open Challenges," in IEEE Access, vol. 6, pp. 37807-37827, 2018, doi: 10.1109/ACCESS.2018.2851311.
2. M. Desai and M. A. Mehta, "Techniques for sentiment analysis of Twitter data: A comprehensive survey," 2016 International Conference on Computing, Communication and Automation (ICCCA), Noida, 2016, pp. 149-154, doi: 10.1109/CCAA.2016.7813707.
3. Abhilash Mittal and Sanjay Patidar, "Sentiment Analysis on Twitter Data: A Survey," 2019 7th International Conference on Computer and Communications Management, Association for Computing Machinery, New York, NY, USA, 91–95. DOI:https://doi: 10.1145/3348445.3348466.
4. Herlawati, R. T. Handayanto, D. Setiyadi and E. Retnoningsih, "Corpus Usage for Sentiment Analysis of a Hashtag Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985772.
5. S. WidyaSihwi, I. PrasetyaJati and R. Anggrainingsih, "Twitter Sentiment Analysis of Movie Reviews Using Information Gain and Naïve Bayes Classifier," 2018 International Seminar on Application for Technology of Information and Communication, Semarang, 2018, pp. 190-195, doi: 10.1109/ISEMANTIC.2018.8549757.
6. N. Yadav, O. Kudale, S. Gupta, A. Rao and A. Shitole, "Twitter Sentiment Analysis Using Machine Learning For Product Evaluation," 2020 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 2020, pp. 181-185, doi: 10.1109/ICICT48043.2020.9112381.
7. S. A. El Rahman, F. A. AlOtaibi and W. A. AlShehri, "Sentiment Analysis of Twitter Data," 2019 International Conference on Computer and Information Sciences (ICIS), Sakaka, Saudi Arabia, 2019, pp. 1-4, doi: 10.1109/ICISci.2019.8716464.

8. Dashtipour, K., Poria, S., Hussain, A. et al. Multilingual Sentiment Analysis: State of the Art and Independent Comparison of Techniques. *CognComput* 8, 757–771 (2016). <https://doi.org/10.1007/s12559-016-9415-7>.
9. Z. Lin, X. Jin, X. Xu, Y. Wang, S. Tan and X. Cheng, "Make It Possible: Multilingual Sentiment Analysis without Much Prior Knowledge," 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, 2014, pp. 79-86, doi: 10.1109/WI-IAT.2014.83.