INTERNATIONAL JOURNAL OF
MULTIDISCIPLINARY RESEARCH
IN SCIENCE, ENGINEERING AND TECHNOLOGY

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 7.54

6381 907 438   6381 907 438   ijmrset@gmail.com   @   www.ijmrset.com

Sentiment Analysis for English Language using Dynamic Streaming of Data

Dr. Meenakshi Srivastava

Assistant Professor, Amity Institute of Information Technology, Amity University Uttar Pradesh, India

ABSTRACT: Sentiment analysis is a valuable tool with multiple applications across various industries. Sentiment analysis helps in understanding the feedback of a reader, the requirement of customer as well as opinions, attitudes, and emotions of customers towards their products, services, or brand. By analyzing customer feedback, reviews, and social media conversations, businesses can identify positive or negative sentiments, gaining insights into customer satisfaction levels and making data-driven decisions to improve their offerings. The present manuscript provides a detailed study of web-based sentiment analysis using dynamic streaming of data.

KEYWORDS: Dynamic Streaming of Data, Sentiment Analysis, Social Media Documents, Classification.

I.INTRODUCTION

Sentiment is a temperament, thought, or discernment encouraged by inclination. Slant investigation, which is otherwise called conclusion mining, considers individuals' assumptions towards specific elements. The Web is an ingenious spot for assessment data. The individuals can post through different web-based applications, for example social media sites, web-based journals, or blogs etc[1].

Companies use Twitter Sentiment Analysis to develop their business strategies, to assess customers' feelings towards products or brand, how people respond to their campaigns or product launches and why consumers are not buying certain product.[2] In politics Sentiment Analysis Dataset Twitter is used to keep track of political views, to detect consistency and inconsistency between statements and actions at the government level. Sentiment Analysis Dataset Twitter is also used for analyzing election results: Twitter Sentiment Analysis also is used for monitoring and analyzing social phenomena, for predicting potentially dangerous situations and determining the general mood of the blogosphere[3].

According to a specialist's point of view, numerous web-based social networking objectives deploy their APIs, instigating information accumulation and examination by scientists and designers[4]. For example, Twitter right now has three distinct forms of APIs accessible, specifically - The Search API, REST API and the Streaming API. With the REST API, engineers can assemble status information and client data; the Search API enables designers to inquire explicit Twitter content, though the Streaming API can gather Twitter content continuously[356]. Besides, designers can blend those APIs to make their very own applications. Thus, the assumption examination appears to have a solid fundament with the help of monstrous online information.

Organizations assembling such items have begun to survey these microblogs to get a feeling of general estimation for their item. Many times, these organizations study client responses and answer clients on microblogs. One test is to manufacture innovation to distinguish and outline a general estimation [67]. Our venture Twitter Sentiment Analysis looks like the breakdown of tweets by the people groups on specific results of organizations or marks or performed by political pioneers. To do this we dissected tweets from Twitter. Tweets are a dependable wellspring of data primarily since individuals tweet about everything without exception, they do include purchasing new items and inspecting them[8]. Many companies have already adopted sentimental analysis for the process of betterment. Some of major applications are mentioned as following: -

   i.   **Word of Mouth (WOM)** refers to the transmission of information from one individual to another. This process plays a crucial role in influencing people's decision-making. WOM provides insights into consumers' opinions, attitudes, and reactions towards businesses, services, and products, which can be shared with multiple individuals. This is where Sentiment Analysis becomes relevant.[7]

  ii. **Voice of Voters :-** Political parties typically allocate a significant portion of their funds towards campaigning and influencing voters. Therefore, having access to the opinions, reviews, and suggestions of the public can greatly enhance their effectiveness[8].

iii. **Online Commerce :-** There is a multitude of e-commerce websites, and a significant number of them adopt a feedback policy to gather input from their users and customers. This feedback encompasses various aspects, including service quality, product features, and user experiences, along with any suggestions. Companies collect and aggregate these details and reviews, converting the data into geographical representations while incorporating updates from current online commerce websites that employ similar techniques[5].

iv. **Voice of the Market (VOM) :-** Whenever a product is to be launched by a specific company, the customers would know about the product ratings, reviews, and detailed descriptions about it. Sentiment Analysis can help in analyzing marketing, advertising and for making new strategies for promoting the product. It provides the customer with an opportunity to choose the best among all [6].

v. **Brand Reputation Management (BRM) :-** Sentiment analysis would help to determine how would be a company's brand, service and the service or product would be perceived by the online community. Brand Reputation Management will be concerned about the management of the reputation of the market. It has focuses on the company and product rather than the customer. Thus, the opportunities were created for the purpose of managing and strengthening the brand reputation of the organizations [9].

vi. **Government :-** Sentiment Analysis has helped the administration for the purpose of providing various services to the public. Fair results have to be generated for analyzing the negative and positive points of government [10]. Thus, sentiment analysis is helpful in many fields like decision making policies, recruitments, taxation and evaluating social strategies. Some of the similar techniques provide the citizen-oriented government model where the services and the priorities should be provided as per the citizens [9][11][12][13].

Many of the existing systems work only on the dataset which is constrained to a particular topic. The existing systems also do not determine the measure of impact the results determined can have on the field taken into consideration and it does not allow retrieval of data based on the query entered by the user i.e., it may have constrained scope. We may say that many of the existing systems work on static data rather than dynamic data. Applying unsupervised algorithms like Vector Quantization, are used for data compression, pattern recognition, facial and speech recognition, etc. and therefore cannot be used in determining sentiment in twitter data. The Apriori algorithm fails to handle large datasets and as a result can generate faulty results.

The presented work will help the reader to understand the basics of sentiment analysis. The presented work will provide a detailed insight of steps involved in sentiment analysis and use of machine learning algorithm for classification of sentiment [10][12][13].

II. THE PROPOSED SYSTEM

The framework of the developed system is presented in Fig 1. The goal is to achieve sentiment analysis for data provided from Twitter. We are going to build a classifier which comprises of different ML algorithms. Once our classifier is ready and trained the following steps are followed to classify the data.

Step-1 First we are going to stream tweets in our build classifier with the help of Tweepy library in python.

Step-2 Then we pre-process these tweets, so that they can be fit for mining and feature extraction.

Step-3 After pre-processing we pass this data in our trained classifier, which then classifies them into positive, negative, or neutral class based on trained results.

Since, Twitter is our source of data for analysis. We are going to stream the tweets from twitter in our database. For this we are going to use Twitter Application.
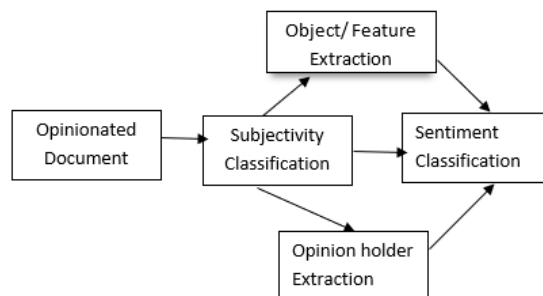

Figure 1 Framework of System

i. **Adaptive Customer Service: -**Your human agents are great at providing flexible service, but it can be difficult to identify the best approach for each customer. With sentiment analysis, it's easier for your team to adapt their service to the mood of the customer early on. Empathetic service makes for a great experience.

ii. **Identifying Key Emotional Trigger: -** Emotional triggers drive our decisions. Using sentiment analysis, you can identify what messages and conversations act as emotive triggers that change customer mood. Perhaps the phrase "Please wait", for example, often triggers customer annoyance. Or perhaps using emojis has a positive effect on the conversation's overall tone. Understanding what messages trigger certain emotions in your customers can help you give better service and is also useful for creating effective marketing materials.

iii. **Identifying Multiple Customers:** - In a chat session, agents can find themselves handling more than one customer at a time. So, another of the benefits of sentiment analysis is its ability to provide a helping hand. Sentiment analysis helps you keep track of the moods of any number of customers for your team.

iv. **Reduce Customer Churn: -** The benefits of sentiment analysis extend into your bottom line. With sentiment analysis, you can identify a dissatisfied customer as and when they're chatting with your team. This enables your agents to offer a smooth service and quick resolution to appease, and ultimately retain, the customer.

v. **Tracking Overall Customer Satisfaction:** - Sentiment analysis scoring puts a quantifiable number on customer satisfaction. It enables you to see the impressions and moods of customers when they approach you, before they get support, and how effective your service is at increasing satisfaction.

Twitter API (Application Programming Interface)
Twitter allows users to gather twitter data with the help of Twitter API. Twitter provides two kinds of APIs: REST API and Streaming API[14]. The differences between these are: REST APIs support connections for short time interval and only limited data can be collected at a time, whereas Streaming API provides tweets in real-time and connection for long time. Streaming API has been used for analysis. For collecting large number of tweets, long-lived connection and no limit data rate is required.

Data Collection
Data of Twitter
To utilize twitter API we should initially have a twitter account. It can be easily created by filling the registration details in twitter.com website. After this you will be given a username and secret key which is used for signing in. Twitter provides a platform from which data cab ne accesses from twitter account and can be used for various purposes.

Data Storage
Once we start getting our data from Twitter API our next step is to store that data so that we can use it for sentiment analysis. We ran our scripts for periods of month and collect the tweets for different political parties. Every time we run the script described in figure a .csv (comma separated values) file is generated which consists of tweets that are extracted from Twitter API. We use .csv format for our collected data files because data consists of several fields. CSV separates each field with a comma, thus making it much easier to access the field which consists of texts. CSV files also provide faster read/write time as compared to other files.

	tweets	len	ID	Date	Source	Likes	RTs
0	Amazing e	140	1.11E+18	########	Twitter fo	23673	6051
1	In the Nor	133	1.11E+18	########	Twitter M	8722	2486
2	What are	140	1.11E+18	########	Twitter M	13680	3700
3	Glimpses	140	1.11E+18	########	Twitter W	7805	2193
4	Its been	140	1.11E+18	########	Twitter M	10305	3084
5	Commun	140	1.11E+18	########	Twitter M	8216	2612
6	à¤à¤¯	113	1.11E+18	########	Twitter fo	28921	5675
7	Hello Man	95	1.11E+18	########	Periscope	10393	3105
8	Saradha.	140	1.11E+18	########	Twitter M	19409	5634
9	Why are	140	1.11E+18	########	Twitter M	12205	3820
10	Speed bre	140	1.11E+18	########	Twitter W	9436	2927
11	Best wish	140	1.11E+18	########	Twitter fo	18719	4038
12	Udaipur ir	94	1.11E+18	########	Periscope	15006	4146
13	On #Worl	144	1.11E+18	########	Twitter fo	22035	4442
14	#WorldHe	140	1.11E+18	########	Twitter fo	29312	5977
15	Great vibr	86	1.11E+18	########	Periscope	15463	4806
16	Be it	140	1.11E+18	########	Twitter M	21015	6010
17	Will be	139	1.11E+18	########	Twitter fo	20018	4505
18	Today's	140	1.11E+18	########	Twitter W	13054	3839
19	à¤¦à¥‡à¤¶	87	1.11E+18	########	Twitter M	18907	5270
20	à¤•à¤¾à¤	118	1.11E+18	########	Twitter M	30393	8013
21	Overwhe	140	1.11E+18	########	Twitter W	9187	2664
22	Nanded is	124	1.11E+18	########	Periscope	17351	4990
23	BJPâ€™s	140	1.11E+18	########	Twitter M	12807	3758

Figure 2: Data Collection

III.RESULTS

First, we use the Algorithm API to pass our topic to the algorithm Retrieve Tweets with Keyword (Retrieve tweets that include keyword anywhere int their text. Limited to 500 tweets per call.) as our input. This will grab tweets containing our phrase. Second, we clear out the retweets so that we don't have duplicate data throwing off our scores. Twitter conveniently includes "RT" at the beginning of each tweet, so we find tweets with that string and remove them from our data set. This leaves us with a convenient set of tweets in the array of no retweets.

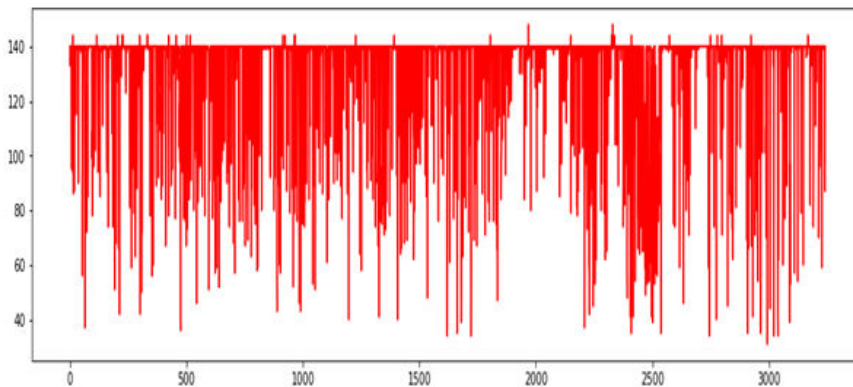|	tweets	len	ID	Date	Source	Likes	RTs
0	Amazing enthusiasm! Hats off to the people who...	140	1114948500071112711	2019-04-07 17:50:11	Twitter for iPhone	23673	6051
1	In the Northeast, Congress has committed too m...	133	1114934730858696705	2019-04-07 16:55:28	Twitter Media Studio	8722	2486
2	What are the compulsions that make Mahamilawat...	140	1114934569663209473	2019-04-07 16:54:50	Twitter Media Studio	13680	3700
3	Glimpses from the lively public meeting in Imp...	140	1114934243161849856	2019-04-07 16:53:32	Twitter Web Client	7805	2193
4	Its been a year since Tripura created history ...	140	1114906611779342337	2019-04-07 15:03:44	Twitter Media Studio	10305	3084
5	Communists claim to standup for the rights of ...	140	1114905875393728513	2019-04-07 15:00:49	Twitter Media Studio	8216	2612
6	जय माँ त्रिपुरा सुंदरी!\n\nFeeling blessed aft...	113	1114905169744072705	2019-04-07 14:58:01	Twitter for iPhone	28921	5675
7	Hello Manipur! Talking about NDA's development...	95	1114858627221708800	2019-04-07 11:53:04	Periscope	10393	3105
8	Saradha.\n\nNarada.\n\nRose Valley. \n\nThese ...	140	1114858077239431168	2019-04-07 11:50:53	Twitter Media Studio	19409	5634
9	Why are TMC and Didi so disturbed when:\n\nInd...	140	1114857525860397057	2019-04-07 11:48:41	Twitter Media Studio	12205	3820

Figure 3: Sample Data

Figure 4 Frequency of Tweeted words

After gathering and cleaning our data set, we are ready to execute the sentiment analysis algorithm on each tweet. Then, we will calculate an average score for all the tweets combined.

	tweets	len	ID	Date	Source	Likes	RTs	SA
0	Amazing enthusiasm! Hats off to the people who...	140	1114948500071112711	2019-04-07 17:50:11	Twitter for iPhone	23673	6051	1
1	In the Northeast, Congress has committed too m...	133	1114934730858696705	2019-04-07 16:55:28	Twitter Media Studio	8722	2486	1
2	What are the compulsions that make Mahamilawat...	140	1114934569663209473	2019-04-07 16:54:50	Twitter Media Studio	13680	3700	0
3	Glimpses from the lively public meeting in Imp...	140	1114934243161849856	2019-04-07 16:53:32	Twitter Web Client	7805	2193	0
4	Its been a year since Tripura created history ...	140	1114906611779342337	2019-04-07 15:03:44	Twitter Media Studio	10305	3084	-1

Figure 4: Tweeted Data

The Positive, Negative and Neutral sentiments are represented with the help of pie-chart. 46.19% are Positive sentiment, 6.11% ar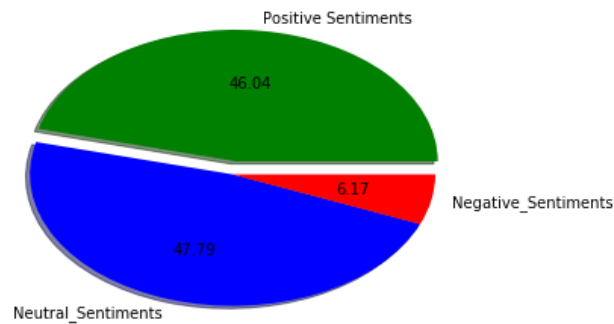e Negative sentiment and 47.70% are Neutral sentiment. Here the Positive, Negative and Neutral sentiments are represented with the help of pie-chart. 46.19% are Positive sentiment, 6.11% are Negative sentiment and 47.70% are Neutral sentiment.

Figure 3: Positive, negative and Neutral Sentiments

IV. CONCLUSION

Twitter sentiment analysis comes under the category of text and opinion mining. It focuses on analyzing the sentiments of the tweets and feeding the data to a machine learning model to train it and then check its accuracy, so that we can use this model for future use according to the results. Some of future scopes that can be included in our research work are an online application can be made for our work in future, building the grouping classifications with the goal that we can show signs of improvement results, dealing with multi dialects like Hindi, Spanish, and Arabic to give slant investigation to increasingly nearby.

REFERENCES

1. David Zimbra, M. Ghiassi and Sean Lee, "Brand-Related Twitter Sentiment Analysis using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks", IEEE 1530-1605, 2016.
2. Varsha Sahayak, Vijaya Shete and Apashabi Pathan, "Sentiment Analysis on Twitter Data", (IJIRAE) ISSN: 2349-2163, January 2015.
3. Peiman Barnaghi, John G. Breslin and Parsa Ghaffari, "Opinion Mining and Sentiment Polarity on Twitter and Correlation between Events and Sentiment", 2016 IEEE Second International Conference on Big Data Computing Service and Applications.
4. Mondher Bouazizi and Tomoaki Ohtsuki, "Sentiment Analysis: from Binary to Multi-Class Classification", IEEE ICC 2016 SAC Social Networking, ISBN 978-1- 4799-6664-6.
5. Nehal Mamgain, Ekta Mehta, Ankush Mittal and Gaurav Bhatt, "Sentiment Analysis of Top Colleges in India Using Twitter Data", (IEEE) ISBN -978-1-5090-0082-1, 2016.
6. Halima Banu S and S Chitrakala, "Trending Topic Analysis Using Novel Sub Topic Detection Model", (IEEE) ISBN- 978-1-4673-9745-2, 2016.
7. Shi Yuan, Junjie Wu, Lihong Wang and Qing Wang, "A Hybrid Method for Multi-class Sentiment Analysis of Micro-blogs", ISBN- 978-1-5090-2842-9, 2016.
8. Apoorv Agarwal, Boyi Xie, Ilia Vovsha, Owen Rambow and Rebecca Passonneau, "Sentiment Analysis of Twitter Data" Proceedings of the Workshop on Language in Social Media (LSM 2011), 2011.
9. Neethu M S and Rajasree R, "Sentiment Analysis in Twitter using Machine Learning Techniques", IEEE – 31661, 4th ICCCNT 2013.
10. Aliza Sarlan, Chayanit Nadam and Shuib Basri, "Twitter Sentiment Analysis", 2014 International Conference on Information Technology and Multimedia (ICIMU), Putrajaya, Malaysia November 18 – 20, 2014.
11. Feature engineering, Wikipedia 2017, https://en.wikipedia.org/wiki/Feature_engineering
Efthymios Kouloumpis and Johanna Moore,IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012
12. S. Batra and D. Rao, "Entity Based Sentiment Analysis on Twitter", Stanford University,2010
13. Saif M.Mohammad and Xiaodan zhu ,Sentiment Analysis on of social media texts:,2014
14. Afroze Ibrahim Baqapuri, Twitter Sentiment Analysis: The Good the Bad and the OMG!, Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media.2011.

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com