Twitter Sentiment Analyzer

[1] Pinky Yadav, [2] Sanchita Rathi, [3] Daksh Malhotra [4] Prateek Bansal

[1] Assistant Professor [234]Students

[1234] Computer Science Engineering Department,

[1234]Dr. Akhilesh Das Gupta Institute of Technology & Management, Delhi, India.

Abstract: Social media as we know plays a vital role in influencing one's opinion and life. We are so wrapped in our virtual world that we want to confess and share every feeling of expression and knowledge through words or pictures and Twitter is one of those platforms that provide us the freedom of speech. In this paper, we are going to analyze the sentiments of such speeches using a support vector machine and random forest in python in a format that the tweets will be categorized in either positive, negative or neutral zones. Using the Django framework allowing us to quickly and efficiently create a quality Web application for both the front end and back end.

***Keywords* - Twitter, Django, Random Forest, Support Vector Machine, Sentiment.**

I. INTRODUCTION

This project addresses the emotions of speech present on Twitter by classifying tweets consistent with the sentiment expressed in them: positive, negative, or neutral. Twitter is a web micro-blogging and social-networking platform which allows users to write down short status updates of a maximum length of 140 characters. It is a rapidly expanding service with over 200 million registered users - out of which 100 million are active users and half of them go online on Twitter on a day to day - generating nearly 250 million tweets per day. Due to this batch of usage, we hope to understand a reflection of public sentiment by analyzing the emotions expressed within the tweets. Sentimental analysis is an automatic process of identifying and classifying subjective information of text data.. This might be someone's opinion, judgment, or a feeling about a particular event, topic, or product used by consumers. This sentimental analysis can be used to monitor the company and find out the opinion of people. Imagine you just launched a new product or service and notice a sudden increase in its mentions on Twitter. Are customers tweeting more because they are happy with the new feature or have complaints? Looking at each of these comments one by one would take way too much time. You could miss out on valuable feedback that could help you improve a customers' experience with the possible modifications. (*bug issues*, *user experience*). By performing sentiment analysis with machine learning, you can quickly understand the sentiments of mentions regarding your product/service.

II. BASIC CONCEPTS OF SENTIMENTAL ANALYSIS

Scalability: Analyze as many tweets mentioning your brand and automate the manual task. Easily scale sentiment analysis tools as your data grow and gain valuable insights on the way ahead.

Real-Time Analysis: Twitter sentiment analysis is essential to monitor sudden reactions of customer moods, detecting if their complaints are on a rise, and taking an action before those problems escalate. With sentiment analysis, monitor brands mentioned on Twitter in real-time and can easily gain actionable insights for further use.

Consistent Criteria: We can easily avoid inconsistencies that can stem from any human error. Customer reps never always agree about which tag to use for every piece of data, so you may end up with some inaccurate results. Instead, machine learning models perform sentiment analysis using one set of rules, so you can ensure that your Twitter data is always tagged consistently.

.

III. RESEARCH METHODOLOGY

The methodology section outlines the plan and method of how the study is conducted.

3.1 Data and Sources of Data

For this study, Multiple datasets have been used in the project. The first dataset is of the words that tell the polarity of the words, i.e., whether the word is positive or negative. The dataset has a polarity of 1 for positive words and a polarity of 0 for negative words.

Here are the listed Negative words **(Fig 1.)** and Positive words **(Fig 2.)** respectively;

anti-social	0
anti-us	0
anti-white	0
antipathy	0
antiquated	0
antithetical	0
anxieties	0
anxiety	0
anxious	0
anxiously	0
anxiousness	0
apathetic	0
apatheticall	0
apathy	0
apocalypse	0
apocalyptic	0

Figure 1.

accurate	1
accurately	1
achievable	1
achievemen	1
achievemen	1
achievible	1
acumen	1
adaptable	1
adaptive	1
adequate	1
adjustable	1
admirable	1
admirably	1
admiration	1
admire	1
admirer	1

Figure 2.

The rest of the words are considered neutral in the project. The second dataset is a set of tweets. That has been downloaded from Twitter using the Twitter API (tweepy) by entering the hashtag and the limit of the number of tweets to be downloaded was also specified. The dataset is stored in the .csv format. The downloaded set is also stored in a .csv format.

Here are the listed top 15 real-time tweets based on the political data;

	A	B	C	D
1	tweet	username	dated	
2	@realDonaldTrump Biden got 81 million votes that is way more than 73m.	CraggsMegyorky	12-12-2020 07:56	
3	@scotland4me_ @realDonaldTrump Just because Rump received 74 millio	StupoTrump	12-12-2020 07:56	
4	|RT @JackPosobiec: The first debate had 73 million viewers and Chris Wallace refused to allowed questions about Hunter Biden			
5	Hunter Biden w...|	JoAnnHaskins	12-12-2020 07:56	
6	|An email obtained by NBCNews indicates President-elect Biden's son was	000 in income... https://t.co/Vt643K	flika_limilki	12-12-2020 07:56
7	@realDonaldTrump You may have been the "sitting president" with the mc	ScatJanDo	12-12-2020 07:56	
8	|RT @AriBerman: We cannot normalize how totally insane it is that Trump	18 state AGs & 106 House Rept	carl_young	12-12-2020 07:56
9	|RT @DeaconBlues0: Trump: I want to see Biden in prison.			
10	Biden: why does Trump think I would visit him in prison?|	CindyMcCauley1	12-12-2020 07:56	
11	|@Donna40323948 @gregkellyusa @JoeBiden It's funny when Trump can	000 people at each rally and Biden w	whskrs98	12-12-2020 07:56
12	@willquim @PreetBharara @realDonaldTrump When Trump told us that v	rwvaughn	12-12-2020 07:56	
13	RT @axcomrade: ICU nurses died after taking care of COVID patients whil	davidxrommal	12-12-2020 07:56	
14	RT @showusyourwork: Cuomo is going to beg Biden to appoint him to sor	rariel81	12-12-2020 07:56	
15	RT @plazynoodles: ATTENTION AMERICANS. PLEASE RETWEET THIS. the r	junevrisrezi	12-12-2020 07:56	

Figure 3.

3.2 Theoretical framework

Django is a python based open source and free web framework which follows the model template views architectural pattern. Django is a high-level python web framework. We used Django because it is versatile, secure, scalable, maintainable, and portable We have used Django in our project to create forms to display the result of the Twitter sentiment analysis in the form of a web page that contains the pie chart to depict the result of the percentage of positive, negative and neutral tweets. We are also displaying the tweets along with the user names and the polarity of the tweet through the random forest and support vector machine algorithm individually in the tabular format.

Figure 4.

3.3 Algorithms :

➢ **Random Forest -**

The fundamental concept which runs behind random forest is simple but powerful. In terms of data science, the reason that the random forest model works properly is an enormous number of relatively uncorrelated models (trees) that operate as a committee outperform any one of the individual constituent models.

The small correlation between such models is its key. A bit like the investments with very small correlations close to making a portfolio that's greater than the sum of all its parts, uncorrelated models can produce predictions that are more accurate than anybody of the individual predictions.. The reason for this awesome effect is that the trees protect one another from any individual errors. While some trees can turn out to be wrong, many of the other trees will give you the right answers, so as a group the trees are always able to move in the correct direction. So the prerequisites for random forest classifier to perform are:

a) There needs to be some actual signal in the features that we add so that models we built by using those features can do better than mere random guessing.

b) The predictions (and the errors) made by any of the individual trees need to have a low correlation with each other.

➢ **Support Vector Machine -**

This paper describes implemented results that consist of Support Vector Machine (SVM) on benchmark datasets to teach a sentiment classifier. The course of procedure is followed up by Splitting the dataset into training and test samples, classifying the predictors and target initializing Support Vector Machine and fitting the training data Predicting the classes for the test set, comparing the particular classes and predictions. An SVM is implemented in a slightly different way than other machine learning algorithms. it's capable of performing classification, regression, and outlier detection. Support Vector Machine may be a discriminative classifier that's formally designed by a separative hyperplane. The most objective of a support vector machine is to segregate the given data in the absolute best way. When the segregation is completed, the space between the closest points is understood because of the margin.

IV. **Results and Conclusion**

Fig 5 shows the result of the Twitter sentiment analysis in the form of pie charts. The first pie chart depicts the result of the sentimental analysis of the tweets through the Random forest algorithm and the second pie chart shows the result of the sentiment analysis of the tweets through SVM. The tweets that are downloaded and analyzed are shown in the table along with the username and the result of the Twitter sentiment analysis of each tweet individually through random forest and SVM.

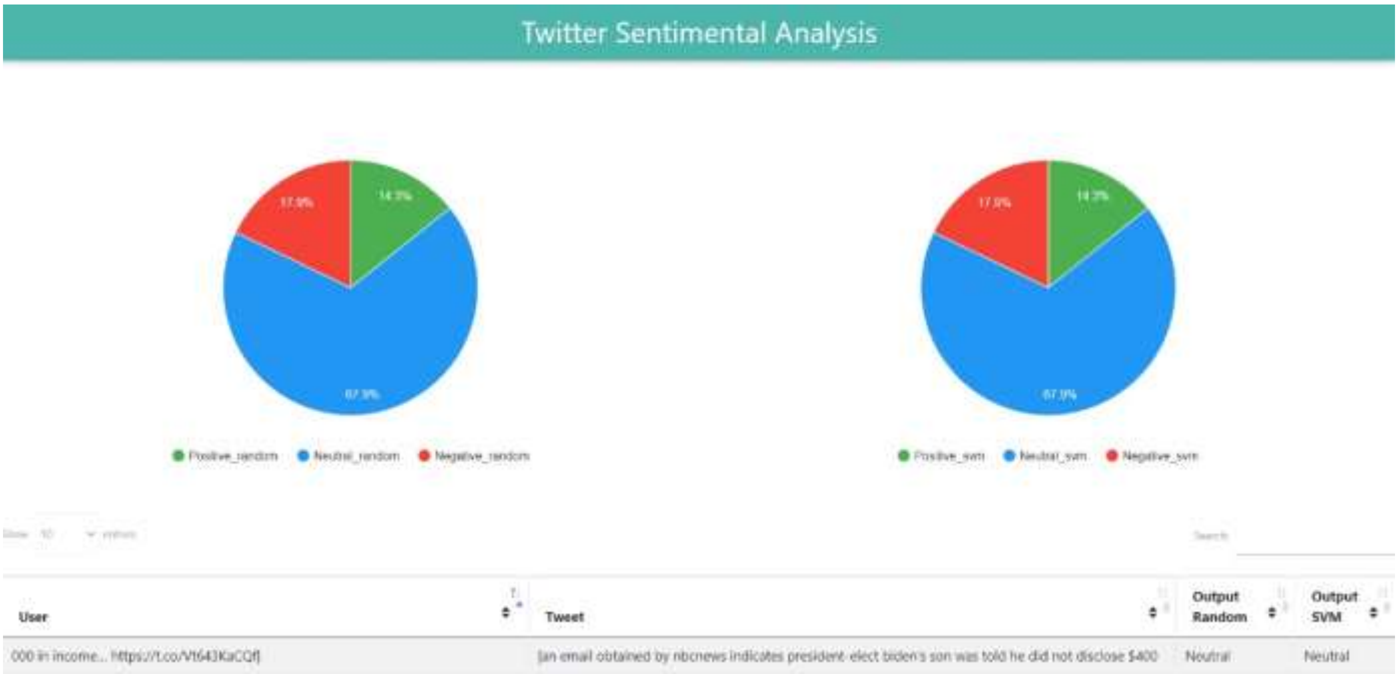

Figure 5.

Fig 6 shows the sentiment analysis of the tweet that has been entered manually by the user. It also has two pie charts. The first pie chart shows the result of the sentiment analysis through the Random forest algorithm and the second pie chart shows the result of the sentiment analysis through the SVM.

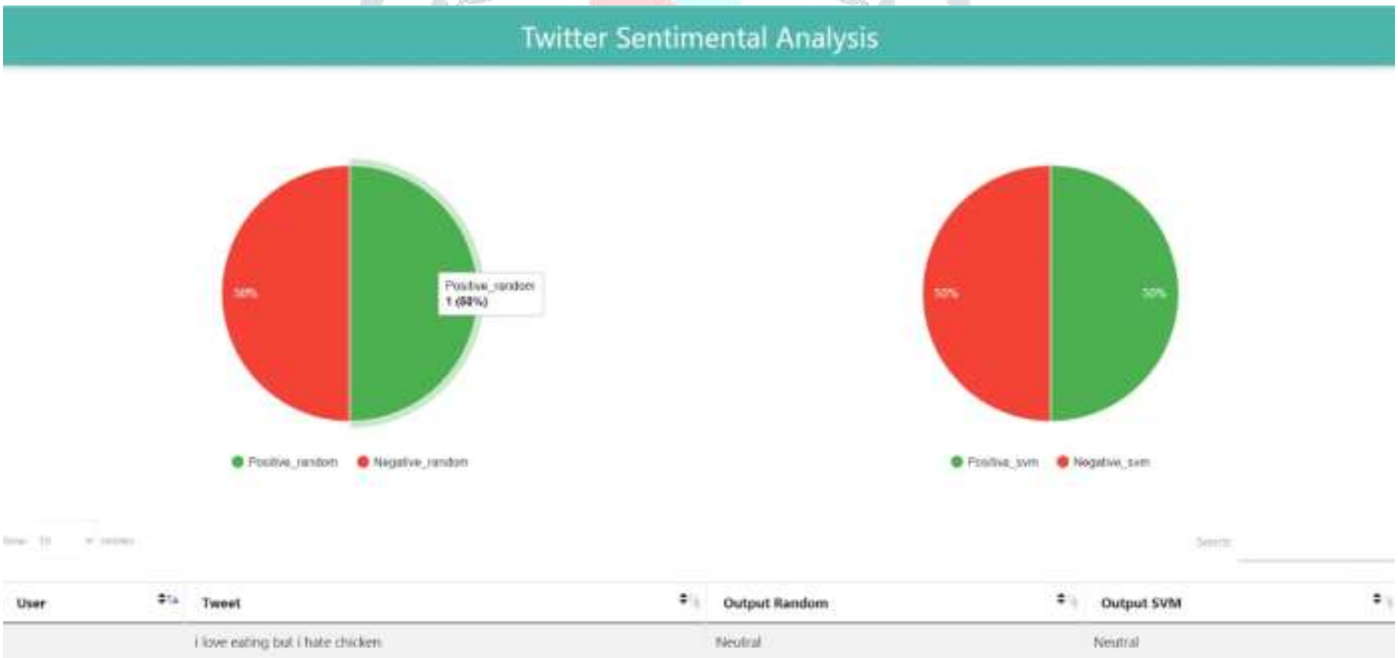

Figure 6.

V. THROUGHOUT THE ANALYSIS

a) Words play a significant role when considering the tweets for evaluating the characteristics of the statements.
b) Along with the extraction of single sentences from random Twitter data sentiments, real-time Twitter data can also be analyzed.
c) The pie charts also include the number of positive, negative, or neutral word a statement contains

VI. FUTURE SCOPE

Sentiment analysis is a very powerful tool for businesses that want to measure the attitudes, feelings, and emotions of customers for their brand. By analyzing customer sentiments, brands can get an inside look at consumer sentiments and, ultimately try to serve their audiences with better products, services, and experiences which they can offer.

Sentiment Analysis is not just a social analytic too. This field is still being studied but not at great lengths due to the intricacy of this analysis. That is a sentimental analysis

Some functions are too complex for machines to understand. The ability to understand sarcastic comments, positive feelings, or negative feelings is very difficult, for machines that lack feelings. Algorithms fail to predict with more than 60% accuracy the feelings displayed by people. Yet with so many limitations this is one field that is growing at a great pace in many industries. Companies want to apply sentiment analysis in the areas of customer feedback, marketing, CRM, etc.

Sentiment analysis is also breaking into new areas of application. While we think of it in terms of the traditional marketing sense, the world is already using it in other areas. Social media analytics helped predict and explain the emotions of concerned parties behind Brexit and therefore the 2016 US election, which has made several non-brand organizations to research how sentiment analysis often wont to predict outcomes and map the emotional thinking of voters. Also, businesses are watching ways in which sentiment analysis are often used outside of their marketing and PR departments. Sentiment analysis is going to be very fashionable in the future.

ACKNOWLEDGMENT

We have taken efforts in this project. However, it would not have been possible without the kind support and help provided by our guide Ms. Pinky ma'am. We are highly indebted to Dr. Akhilesh Das Gupta Institute of Technology & Management for their cooperation and encouragement which help us in the completion of this assignment. Also, our thanks and appreciation go to the people who directly or indirectly helped us out in developing this research paper and project.

REFERENCES

[1] Efthymios Kouloumpis and Johanna Moore, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012

[2] S. Batra and D. Rao, "Entity Based Sentiment Analysis on Twitter", Stanford University,2010

[3] Saif M.Mohammad and Xiaodan zhu, Sentiment Analysis on of social media texts:,2014

[4] Ekaterina kochmar, University of Cambridge, at the Cambridge coding Academy Data Science.2016

[5] Manju Venugopalan and Deepa Gupta, Exploring Sentiment Analysis on Twitter Data, IEEE 2015

[6] Brett Duncan and Yanqing Zhang, Neural Networks for Sentiment Analysis on Twitter.2017

[7] 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

[8] M. Bautin, L. Vijayrenu L, and Skenia., "International sentiment analysis for news and blogs", In Second International Conference on Weblogs and Social Media (ICWSM), 2008.

[9] P. Turney., "Thumbs up or thumbs down?" Semantic orientation applied to unsupervised classification of reviews. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics, pp. 417424, 2002.

[10] B. Pang and L. Lee., "Opinion mining and sentiment analysis", Foundations and Trends in Information Retrieval, vol. 2, no. 1-2, pp. 1-135, 2008.

[11] B. Pang and L. Lee., "Using very simple statistics for review search: An exploration", In Proceedings of the International Conference on Computational Linguistics (COLING), 2008.

[12] K. Dave, S. Lawrence, and D. Pennock., "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews", pp. 519-528, 2003.

[13] N. Godbole, M. Srinivasaiah, and S. Skiena., "Large-scale sentiment analysis for news and blogs," 2007.

[14] A. Kennedy, D. Inkpen, "Sentiment Classification of Movie and Product Reviews Using Contextual Valence Shifters", Computational Intelligence, pp.110125, 2006.

[15] J. Kamps, M. Marx, R. Mokken., "Using WordNet to Measure Semantic Orientation of Adjectives", LREC 2004, vol. IV, pp. 11151118, 2004.

[16] V. Hatzivassiloglou, and J. Wiebe., "Effects of Adjective Orientation and Gradability on Sentence Subjectivity", Proceedings of the 18th International Conference on Computational Linguistics, New Brunswick, NJ, 2000.

[17] A. Andreevskaia, S. Bergler, and M. Urseanu, "All Blogs Are Not Made Equal: Exploring Genre Differences in Sentiment Tagging of Blogs", In International Conference on Weblogs and Social Media (ICWSM-2007), Boulder, CO, 2007.

[18] P. Turney, and M. Littman., "Measuring Praise and Criticism: Inference of Semantic Orientation from Association", ACM Transactions on Information Systems.

[19] G. Biau and L. Devroye. On the layered nearest neighbor estimate, the bagged nearest neighbor estimate and the random forest method in regression and classification. Journal of Multivariate Analysis, 101:2499–2518, 2010.

[20] Bartlett P. and Shawe-Taylor J., "Generalization performance of support vector machines and other pattern classifiers", In C.~Burges B.~Scholkopf, editor, "Advances in Kernel Methods--Support Vector Learning". MITpress, 1998.