

Filtering Political Sentiment on Social Media

^[1]Indrajeet Garde, ^[2]Shreyas Lolage, ^[3]Shekhar Sabale, ^[4]Ajay Patil, ^[5]Preeti Joshi

^{[1][2][3][4][5]}Department of Information Technology, MMCOE, Pune, India

Abstract--- Social Media is an incredible source of appraisals, political or otherwise, therefore the domain of sentiment analysis has seen an increase in interest over the span of the last couple of years. And given the current rise of populist leaders all around the world, gauging the public's response to any political institution or candidate is vital. Sentiment analysis mines opinions at word, sentence, and document levels, and offers input polarities and characteristics of articles. We propose a system that filters out political data from the social network Twitter, using typical Machine Learning algorithms, and then performs sentiment analysis. Thus, we shall better understand the general public's opinion on certain political topics.

Keywords— Machine Learning, Politics, Sentiment Analysis, Twitter

I. INTRODUCTION

1.1 Motivation

Politics of the present day have become more and more contentious. As debates become very polarized, conversations encompassing politics, particularly regarding the future, are more or less impossible to avoid. Whether online, in the classroom or simply watching the news, the terms "left" and "right" find a way to creep into any dialogue. Though typically repetitive, these conversations are vital, and it might be unaccountable to steer away from them. Thus, as a result of the academic opportunities we tend to all have as a community, it's imperative to acknowledge its importance on our lives.

Currently, the world has witnessed the rise of several populist leaders. Webster's dictionary defines a Populist as a member of a political party claiming to represent the common people. Various leaders such as India's Narendra Modi, Pakistan's Imran Khan, the Brazilian Jair Bolsonaro, and quite famously, the former United States' President Donald Trump have risen and gained a lot of influence. Hence, it is of further importance that we pay attention to political topics because like it or not, the government plays a huge role in our daily lives. Be it the rate of taxes, the price of a light bulb or even what opportunities you may get. Having this information could help us determine what is the best course of action to help ourselves.

Social media, albeit a toxic cesspool can be a very useful tool when it comes to data. A person expresses everything, from how their day was to whether they liked a movie, from which celebrity they like to their beliefs on certain topics. Various papers have credited the success of political movements like Occupy Wall Street, Nirbhaya Justice and the Brazilian protests to social media. The involvement of Cambridge Analytica during the

forementioned Donald Trump's campaign during the 2016 US Presidential elections is well documented as well. Therefore, it could be quite important to understand what the general masses express on social media platforms.

1.2 Problem Definition

Gathering public opinion over political matters is an important part of political decision-making. Currently, there are no easy ways to gather public opinion over the data. Polls and Surveys take a lot of time to process and may also be unreliable.

Social media allows people to share their opinions freely and voluntarily. People voluntarily give their opinions about all kinds of Subject matter on social media. It is also available for everyone to view. This data also faces the same problem of unreliability as Surveys and Polls but the Social Media data is readily available. It can be viewed anytime and it is available in real time. Political parties can study this data and it can help to make better decisions. At this point in time, this task is manual. When done manually, there are people involved and people cost more and there is a chance of manual error. Hence, an automated system that gathers public opinion over a subject and gives us an easy-to-read analysis of the data will definitely make this task easier to accomplish.

II. LITERATURE SURVEY

2.1 Literature Review

1] A model is developed that captures the comments of individuals and forwards them to the preprocessing module. It contains URL removal, stopword removal, tokenization, abbreviation expansion, and stemming processes. Furthermore, the result obtained is passed on to a neural network. The final result obtained is binary classification and representation of opinions. Their future aim is to use a single model for preprocessing as well.

2] To reduce the risk of selecting an inappropriate classifier, the authors (Rincy Jose et al) are combining the outputs of a set of classifiers. Thus, in this paper, they use an approach that automatically classifies the sentiment of tweets. Here positivity or negativity of each tweet is determined by using the majority voting principle on the result of three classifiers. They used a sentiment classifier for finding political sentiment from real-time tweets. They got an improved accuracy in sentiment analysis using the classifier ensemble approach.

3] Sentiment analysis was performed on a sample of data collected from the social media network Twitter on the previous USA president Donald Trump. The potential of the tool RapidMiner was exploited in order to complete all phases of text mining and arrive at the desired results. The results show the ratio between positive, negative and neutral opinions about the president.

4] This paper introduced a technique of dynamic vocabulary, in which the vocabulary develops as the training is done. The experimental result shows the performance of the proposed technique is satisfactory. One limitation that they faced was that when training the model, memory becomes an issue if limited. So, extra techniques to reduce the number of features can be tried so that the size of the input vector can be reduced and so the size of the vocabulary.

5] Social associations may ask the opinion of the individuals and their conclusion on current civil arguments. This data can be fetched from social blogs like twitter, since the twitter users express the feelings on the various parts of the life consistently. Here the authors (Patil et al) use the tools for sentiment analysis of the data by providing some of the tweet data as input and obtaining the respective scores as output.

3	Collection and Sentiment Analysis of Twitter Data on the Political Atmosphere.	Merima Čišija, Emir Žunić, Dženana Đonko
4	Twitter Sentiment Analysis using Dynamic Vocabulary	Hrithik Katiyar, Monika, Parveen Kumar and Ambalika Sharma
5	Sentiment analysis by identifying the speaker's polarity in Twitter data	Rashmi Patil, Siddu Algur

Table 1. Literature Review

III. SYSTEM ARCHITECTURE

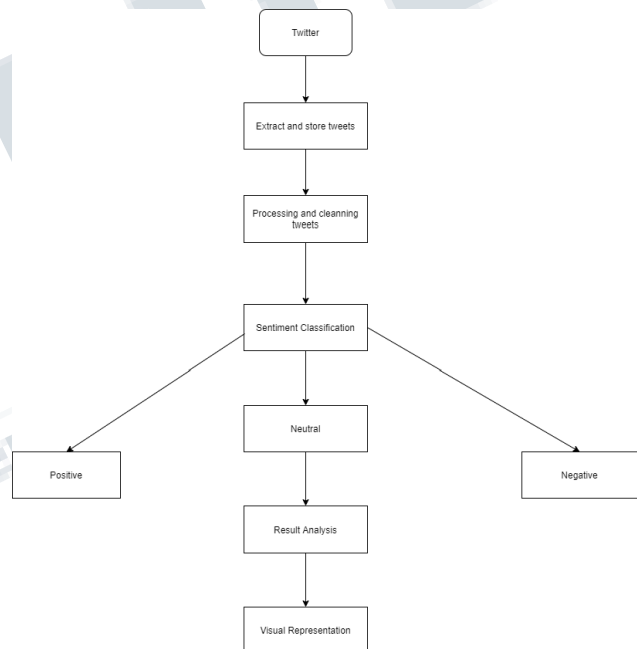


Fig. 1. Proposed System Architecture

A. Naive-Bayes

It is a classification approach primarily based totally on Bayes' Theorem with an assumption of independence amongst predictors. In easy terms, a Naive Bayes classifier assumes that the presence of a selected function in a class is unrelated to the presence of some other function. The Naive-Bayes version is straightforward to construct and in particular beneficial for extremely big records sets. Along with simplicity, Naive Bayes is thought to outperform even quite state-of-the-art classification methods. Bayes

Sr. No.	Title	Authors
1	Political Sentiment Analysis through Social Media.	Saurabh Dorle, Dr. Nitin Pise
2	Prediction of Election Result by Enhanced Sentiment Analysis on Twitter Data using Classifier Ensemble Approach	Rincy Jose, Varghese Chooralil

theorem affords a manner of calculating posterior probability $P(c|x)$ from $P(c)$, $P(x)$, and $P(x|c)$.

$$p(C_k | \mathbf{x}) = \frac{p(C_k) p(\mathbf{x} | C_k)}{p(\mathbf{x})}$$

B. VADER

VADER is a sentiment analysis tool based on vocabulary and predefined rules. VADER not only tells us about the ratio of positive and negative, but also tells us how positive or negative a concept is. VADER can even score points with emojis. He also places great emphasis on capital letters and punctuation (such as "!") because they often emphasize emotions. In addition, different rating modifiers will receive different ratings. For example, "this is a great place!" is more favorable than "this is a good place", a

IV. RESULTS

A. Classifier

For the classifier, we experimented with various classification models to observe the metrics of the same. The dataset was split in the ratio 80:20. Bernoulli Naïve-Bayes was implemented using the Sci-kit learn package in python. Accuracy was computed, as were all the other remaining metrics. Table 2 showcases the various metrics of the Bernoulli Naïve-Bayes classifier.

Category	Accuracy	Precision	Recall	F-1
Non-Political	0.99	0.97	0.98	0.98
Political	0.97	0.98	0.97	0.98

Table 2. Results of classification

B. VADER

To test the performance of VADER, we put a few tweets manually retrieved from Twitter through both, the classifier and VADER. It assigns a positive score for text which emotes positive sentiment and negative score for negative sentiments. Tweets such as, "Just as India sent aid to the United States as our hospitals were strained early in the pandemic, we are determined to help India in its time of need.", "Do you know how mad it is that Mason Mount is already making is 100th appearance for Chelsea? He's not even completed two full seasons yet. What a great player" were accurately judged to be positive, while tweets such as, "ARE GOVT. IS IN FAVOUR OF SEEING THESE TYPES OF AWFUL CONDITIONS WITH STUDENTS!?" and "STOP THE FRAUD!" were

accurately assigned negative scores. Table 3 showcases the various tweets, their classification result and the sentiment scored assigned by VADER.

Sr. No.	Tweet text	Alignment	Sentiment Score
1	Just as India sent aid to the United States as our hospitals were strained early in the pandemic, we are determined to help India in its time of need	Political	0.64
2	Do you know how mad it is that Mason Mount is already making is 100th appearance for Chelsea? He's not even completed two full seasons yet. What a great player.	Non-Political	0.23
3	Bengal - BJP Assam - BJP TamilNadu - DMK Kerala - Left Puducherry - BJP	Political	0.0
4	ARE GOVT. IS IN FAVOUR OF SEEING THESE TYPES OF AWFUL CONDITIONS WITH STUDENTS!????	Political	-0.33

V. CONCLUSION AND FUTURE WORK

To understand the public's opinion and to be aware of how they feel about political topics is quite essential. It can help in educating them or ourselves on both sides of the spectrum on a topic. This will help in encouraging conversation and fostering the growth of democracy. Diversifying the training data is the main concern while implementing such a system, as an imbalanced dataset will skew the results.

REFERENCES

- [1] Ying Fang, Hai Tan, Jun Zhang, "Multi-Strategy Sentiment Analysis of Consumer Reviews Based on Semantic Fuzziness", IEEE Access-2020.

-
- [2] Shihab Elbagir and Jing Yang, "Twitter Sentiment Analysis based on ordinal Regression", IEEE Access-2020.
- [3] Huyen Trang Phan, Van Cuong Tran, Ngoc Thanh Nguyen, and Dosam Hwang, "Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model", IEEE Access-2020.
- [4] Farkhund Iqbal, Jahanzeb Maqbool, Benjamin C. M. Fung, Rabia Batool, Asad Masood Khattak, Saiqa Aleem, and Patrick C. K. Hung, "A Hybrid Framework for Sentiment Analysis using Genetic Algorithm based Feature Reduction", IEEE Access-2020.
- [5] Rosy Indah Permatasari, M. Ali Fauzi, Putra Pandu Adikara, "Twitter Sentiment Analysis of Movie Reviews using Naïve Bayes", IEEE Access-2020.
- [6] Saurabh Dorle, Dr.Nitin Pise, "Political Sentiment Analysis through Social Media", Proceedings of the Second International Conference on Computing Methodologies and Communication
- [7] M. Govindarajan, "Analysis of bagged ensemble classifiers for blogger data", IEEE Access-2020.
- [8] Xian Fan, Xiaoge Li, Feihong Du, Xin Li, Mian Wei, "Apply word vectors for sentiment analysis of APP reviews", ICSAI 2016.
- [9] Shihab Elbagir and Jing Yang, "Twitter Sentiment Analysis based on ordinal Regression", IEEE Access-2020.
- [10] Huyen Trang Phan, Van Cuong Tran, Ngoc Thanh Nguyen, and Dosam Hwang, "Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model", IEEE Access-2020.
- [11] Farkhund Iqbal, Jahanzeb Maqbool, Benjamin C. M. Fung, Rabia Batool, Asad Masood Khattak, Saiqa Aleem, and Patrick C. K. Hung, "A Hybrid Framework for Sentiment Analysis using Genetic Algorithm based Feature Reduction", IEEE Access-2020.
- [12] Robert Hillmann and Matthias Trier. 2012. Dissemination patterns and associated network effects of sentiments in social networks. In Proceedings of the IEEE International Conference on Advances in Social Networks Analysis and Mining.
- [13] Stefan Stieglitz and Linh Dang-Xuan. 2013. Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *J. Manag. Inf. Syst.* 29, 4 (2013), 217–248
- [14] Jie Tang and A. C. M. Fong. 2013. Sentiment diffusion in large scale social networks. In Proceedings of the IEEE International.
- [15] Maria Karakiza. 2015. The impact of social media in the public sector. *Procedia Soc. Behav. Sci.* 175(2 2015). DOI: <https://doi.org/10.1016/j.sbspro.2015.01.1214>
- [16] Saifuddin Ahmed, Kokil Jaidka, and Jaeho Cho. 2017. Tweeting India's Nirbhaya protest: A study of emotional dynamics in an online social movement. *Soc. Mov. Stud.* 16, 4 (2017), 447–465.
- [17] Yung-Ming Li and Ya-Lin Shiu. 2012. A diffusion mechanism for social advertising over microblogs. *December Supp. Syst.* 54, 1(2012), 9–22.
- [18] Hung T. Nguyen, Preetam Ghosh, Michael L. Mayo, and Thang N. Dinh. 2017. Social influence spectrum at scale: Near optimal solutions for multiple budgets at once. *ACM Trans. Inf. Syst.* 36, 2 (2017), 14.
- [19] Josemar A. Caetano, Hélder S. Lima, Mateus F. Santos and Humberto T. Marques-Neto, "Using sentiment analysis to define twitter political users' classes and their homophily during the 2016 American presidential election", SpringerOpen, *Journal of Internet Services and Applications* 2018. DOI: <https://doi.org/10.1186/s13174-018-0089-0>
- [20] Debashis Naskar, Sanasam Ranbir Singh, Durgesh Kumar, Sukumar Nandi, and Eva Onaindia de la Rivaherrera. 2020. Emotion Dynamics of Public Opinions on Twitter. *ACM Trans. Inf. Syst.* 38, 2, Article 18 (March 2020)
-