Predicting Public Opinions and Elections Result by Sentiment Analysis: A Machine Learning Framework

Madhu Shrivastava[1], Bhavana Gupta[2]

[1]*M.TechScholar,* [2]*Associate Professor*

[1,2]*CSE, Sagar Institute of Science & Technology , Bhopal, INDIA*

Abstract- in late years, online Social Networks has given end clients an amazing stage to voice their sentiments. The judgment of human beingsissues areenormouscompact to divide how the stimulating of statisticsimpressions the mankind in a online Social networksakin Twitter, Facebook etc. Organizations need to recognize the extremity of these sentiments to comprehend client direction and accordingly settle on more brilliant choices. One such relevance is in the domain of senatorial matters, where governmental elements shouldfigure out general assessment to decide their crusading technique. Twitter is in fact utilized widely for political consideration. We find that the simple number of messages referencing a gathering mirrors the political decision result. This information is utilized to foresee the result of the political decision throughSentiment Analysis (SA). SA of the tweets decides the extremity and tendency of the huge populace towards a particular point, thing, or substance. Famous Text mining methods like NaiveBayes and Support Vector Machine etcare Supervised Learning methods which involve a preparation informational gathering to execute Sentiment extraction. These calculations are used to fabricate a classifier and characterize the test information as sure, negative, and impartial. In this work we used dataset contains the General Election dataset of AAP, BJP, and INC of the year 2019. These datasets are having both positive and negative audits that are conveyed by viewers on sites. After applying string to vector filtration for the preprocessing of information, include choice techniques are utilized to choose the most important highlights. Improved Naive Bayes calculation utilizing TF-IDF calculations is applied for the reviews/tweets and VADER for sentiment analysis. We are utilizing four parameters for contrasting the outcomes that are Accuracy, Root Mean Square Error, Precision, and F-Measure.

Keywords- Social Media, Tweets , Political Reviews, Opinion Mining, VADER.

I. INTRODUCTION

Sentiment analysis is only the healthy region of exploration and mining strategy with respect to feelings of individuals, assessment, and their state of mind on specific realities and occasions. It is only the center region of normal language preparation that profoundly manages sentences. With the mammoth investigation of online media in the most recent decade and the blast of web innovation an extremely enormous no of individuals today with no dread set forward their assessment and remain on a specific point by the different methods for tweets, post, remarks, likes, and despises and even by sharing specific news identified with some occasion to their locale. And all such exercises really contain a great deal of shrouded data and besides breaking down such content [1] can yield immense data about the conduct of the individuals that can be useful in any part of life.

Arising online Social media gave them enough potential and capacity to spread and make their assumptions on some politically persuaded occasions, current issues on sites, and different occasions occurring everywhere in the world particularly on the different interpersonal interaction locales [2]. At first, the idea of Sentiment analysis and its phrasing was put sent by Nasukawa and Yi [3], while on the opposite side opinion mining was proposed by Dave et al. [4]. Here one thing is outstanding that absolute first and early examination in assessment grouping and investigation was subject to the computation of the measure of utilization of descriptors and single word phrases in the sentence or any post just to register the conclusion estimation of the sentence in any case, in ongoing exploration considers it has been seen that action words, modifier, and two word phrases [5] [6] [7] likewise contribute a great deal to comprehend the general assessment of the sentences.

Aside from this, at present numerous sorts of examination, creators, and examiner have begun to study and discovering a portion of the standard based slant investigation and applying the idea of AI draws near, characteristic language handling to discover the genuine significance of semantic importance [8][9][10]. In this way, it causes colossal examination in the politically arranged outcomes and afterward to investigate and broadcast the aftereffect of any political decision. These sorts of exploration are firmly examined. We have fundamentally investigated all such strategies for foreseeing the consequences of political exercises in various situations related to a huge example of the

populace and their convictions through opinion mining. Here, we will likewise view a portion of the well known work done by various analysts in the territory of Sentiment analysis.

II. RELATED WORK

In the wake of getting essential and fundamental information about estimation and assessment examination we have enough information and thought that this territory of exploration has an all the more wide degree in breaking down a huge populace's assumption and mind-set and this is the reason it is broadly utilized in political situations commonly. What's more, as of late this custom is step by step expanding with a quick rate

Initially, OmaimaAlmatrafi et al. [11] in this paper they analyzed the dataset of twitter for Sentiment analysis and executed the position based SA approximately 0.6 million tweets and then understand the basic inclinations and molds about the Indian General elections. They proclaim that emotions, both negativeand positive, modify from one position to another. Moreover they put this fact forward that some definitecommunityoccasions' mightactivate a quickgrow in both positiveandnegative emotions about a particular political party.

Ali et al. [12] in this paper named 'SAaaS' (Sentiment Analysis as a Service') model categorized the community information assistance depends on itswide range of statistics attributes. They projected an assistance symphony method that subsists of various assistances for SAdepends on communal data assistance categorization. It also commenced a novel assistance element framework to eliminate the error from communal data assistance and carried outresearches on the dataset from real-world.

Gupta et al [13] in the paper named "spam and Sentiment Analysis Model for Twitter Data using Statistical Learning" observed that the positivereviews express greater polarity than the negative reviews. They even concluded this fact that that AAP (political party of India) has contains more negative tweets as compared to other two parties (i.e. INCand BJP-Akali). They have built their classifier in Python that can beutilized for any rationale centered on tweets. In their paper it was suggested that From the LIWC result all the parties showed more positive emotions ascompared to negative words in the tweets.

Parul et. al [14] in his paper named "prediction of Indianelection using sentiment analysis on Hindi twitter"found out that it is incrediblycomplicated to calculate the outcomes of general elections utilizingother schemes, together withopensurveys, and with themountingincidence of online social networks, such as Twitter andFacebook, so they determined to employSA oftweets from Twitter to calculate the outcomes of the general election of India. And subsequent to this they estimated the accuracy and recall of the algorithm like NaiveBayesfor their analysis.

Birmingham and Smeaton [15] also worked extensively in this field and they basically analyzed two different approachesnamed multinomial NaiveBayesmodel and support vector modelrespectively and applied then on to internet pages and blogs. Their result showed that multinomial NaiveBayes model generally outperforms the support vector model on scaled areas when short substance is applied or used.

Another famous political scenario, the US presidential was studied by Wang and Can et al. [16] in which they built a reliable structure for recover political assumptions at occupation utilizingtweets of Twitter. In which from twitter database the real time tweets are used and even tweet's location was used to increase the reliability of result as one filter to analyzing people'ssentiments.

Singhal et al. [17] worked on the Indian General and showed those publics aregradually more utilizing online Social networks to convey their sentiments. And, Twitter one of the main areas to explorethese community sentimentsmainlythroughoutgeneral election moment.Monitoring the outcomes they said that allowing for its emotionsolely couldoutcome in awarding a commoninitiative about the general election outcomes. They introduced many protocolsdepends on linguistic formation of the statement. And their preliminary outcomes showed the efficiency of the projectedmethod over other presentedschemes. They were even equippedcalculate the voting dividend of some parties with adequate percentage of error.

One of the remarkable result were came from Yueet al. [18] and his survey paper that binds all dynamics of role of social media in opining mining that comprehensively discussed all such types of the Sentiment analysis. In this review, a succession of the modern journalisms has been analyzed. Particularly, this studydivided and categorizedSAexplores from numerousstandpoints, i.e., ob-oriented, atomize-oriented, andmethod-oriented. Moreover they investigated different categories of statistics and tools that couldbe utilized in SAexplore and recommended their potency and boundaries. They offered an outline of MSA (multimodal sentiment analysis) and even suggested there are noteworthy and sensibleprospects for potentialstudy in themulti-punitive domain of multimodal synthesis. This studylaid afoundationfor advancedresearch in sentiment analysis.

III. PROPOSED WORK

The principle approaches towards SA are utilized in the related workdepend on discreteopinions categorization and multiclass categorization. In binary categorization, each assessment is categorized into 2groups i.e. good and bad.

Alternatively in multi-class categorization, it is cited to as good, very good, impartial, bad, and very bad. Generally binary categorization is utilized for evaluation of 2emotions like “glad” and “depressing” etc.

The exampleproposedassembles on political appraises that are loaded in an unstructured textual design. Unstructured statistics is additionallyexchanged into significantconfessionviapertaining ML algorithms. Conventionalschemes of ML algorithms were recycled by investigators but when it arrives to bulky datasets with the speedystreaming of statisticsconstantlygrowinggradually it's tough to evaluate with the only MLmethods. Meanwhile, we had proposed machine learning algorithms for our main approach.

Sentiment analysis is the automated process of recognizing ansentimentconcerning to a specifiedtopic from text or verbalspeech. It is needed so much in today’s world due to online reviews taken from different users regarding any product, services and any types of issues. So sentiment analysis has to be done accurately otherwise inaccurate results will lead to us in a wrong path. In this work we have shown the sentiment analysis on the reviews taken on the behalf of political parties like BJP, Congress and AAP. People have given their reviews towards the political party in India. Our proposed sentiment analysis is based on Valence Aware Dictionary and sEntiment Reasoner (VADER) which shows improved classification than other tradition natural processing tool.

3.1 Proposed Algorithm

1. Import all the python libraries(i.e. numpy, pandas, sklearn etc)
2. Import the political dataset from kaggle data repository (i.e. BJP, Congress, AAP)
3. Preprocess the input data
    - Stop word removal
    - Lemmatization
    - Dimension reduction
    - Stemming
4. Set the input and label with polarity, feature extraction
5. Getting processed dataset for evaluation
6. Splitting the dataset into training and testing dataset
7. Analyzing the sentiment

   VADER = Sentiment Intensity Analyzer ()

   Analyzer = Sentiment Analyzer ()
8. Calculate the result (Accuracy, precision, recall ,F1 score and confusion matrix)

3.2 Proposed Flowchart

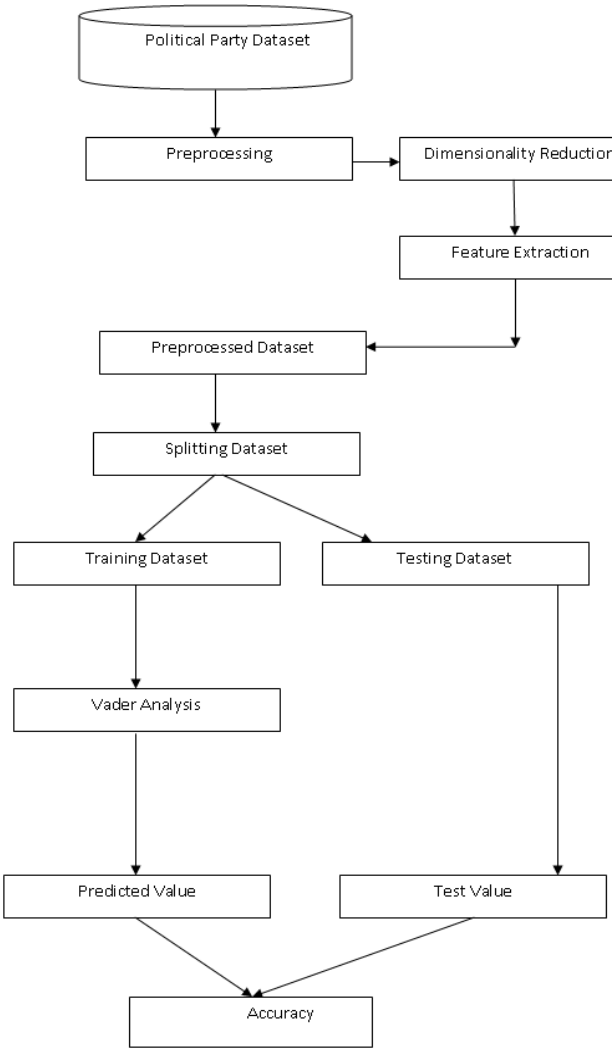

Fig.1: flow graph of proposed method

IV. RESULTS AND DISCUSSION

The parameters which we will consider for evaluation are as follows:

1. Confusion matrix

		True class	
		Positive	Negative
Predicted class	Positive	True Positive Count (TP)	False Positive Count (FP)
	Negative	False Negative Count (FN)	True Negative Count (TN)

Table 1: Confusion matrix table

For evaluating statistical classification model, confusion matrix is used in machine learning which is also known as

error matrix. The predicted class of instances is represented in the row whereas actual classes are represented in columns.

2. **Accuracy**: $\frac{TP+FN}{TP+TN+FP+FN}$

Accuracy can be defined as the percentage of the total number of predictions that are correct is known as Accuracy.

3. **Precision:** $\frac{TP}{TP+FP}$

Precision can be defined as the percentage of positive predictions that are correct is known as Precision.

4. **Recall:** $\frac{TP}{TP+FN}$

Recall can be defined as the percentage of positive cases that are predicted as positive is known as Recall.

5. **F1 Score:** $\frac{2TP}{2TP+FP+FN}$

F1 scores can be defined as the Harmonic mean of precision and recall. Also known as F1-measures**.**

4.1 Result Analysis

The implementation with different aspects and to analyze those parameters is the significant part of result analysis in thesis. In this section we will compare the performance metrics like accuracy, precision, recall and f1 scores with the graphical representation and the comparative table. The observation had done mainly with the domain of supervised learning approach of machine learning classifier namely NaiveBayes, Support Vector Machines, Staring with comparative graph on different evaluation parameters are as follows:

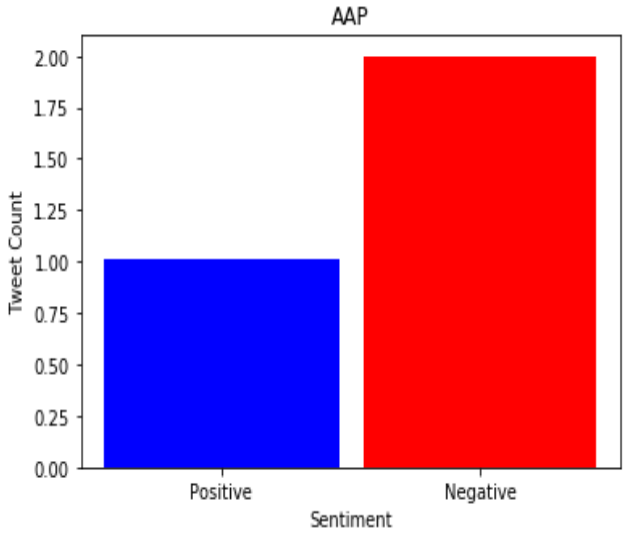

Fig. 2: Results of AamAadmi Party Dataset

Analysis of the AAP Dataset

Proposed Accuracy: 1.0

Proposed Precision: 1.0

Proposed Recall: 1.0

Proposed F1 measure: 1.0

Result analysis of BJP Dataset

Accuracy: 0.7333333333333333

Precision: 0.7142857142857143

Recall: 1.0

F1 measure: 0.8333333333333333

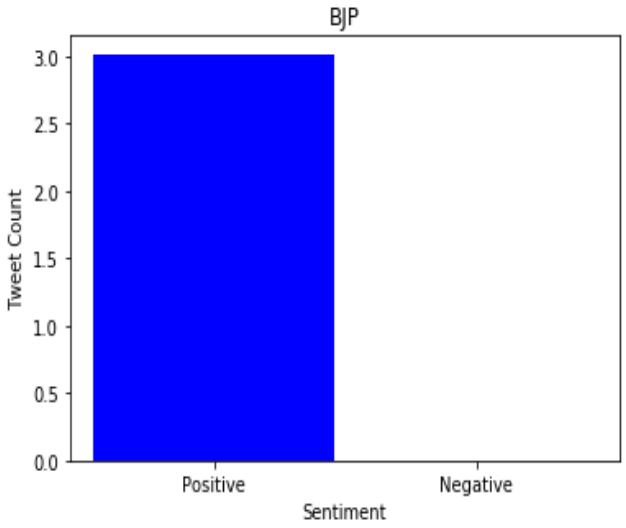

Fig. 3: Results of BJP Dataset

Result analysis of Congress Dataset

Accuracy: 0.6

Precision: 0.7142857142857143

Recall: 0.5555555555555556

F1 measure: 0.6250000000000001

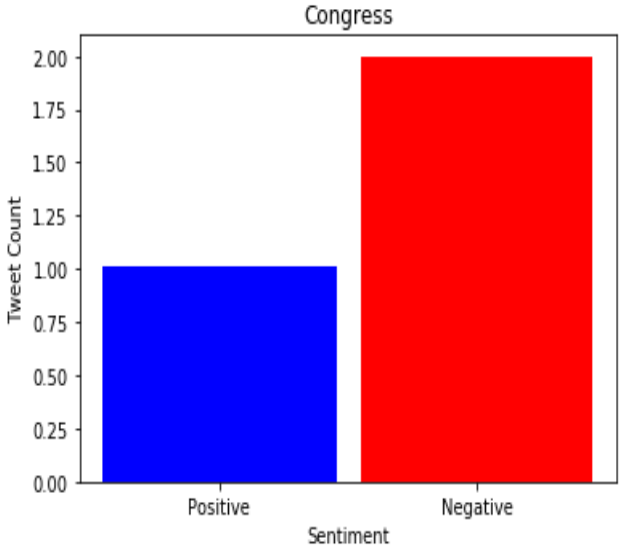

Fig. 4: Results of Congress Dataset

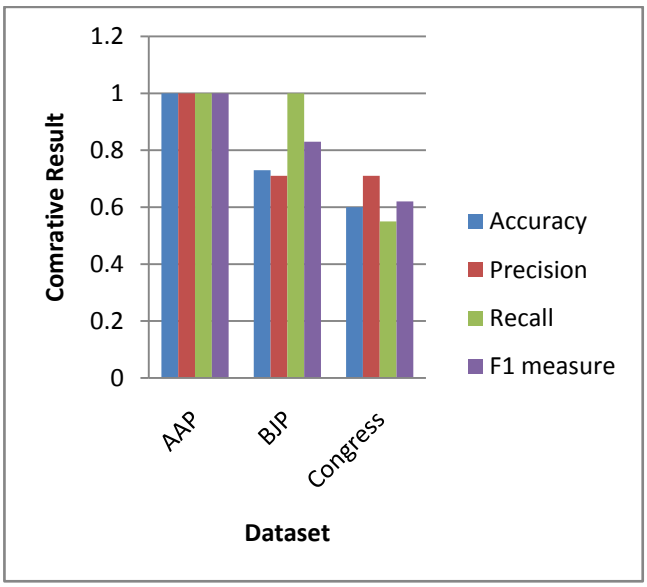

Fig. 5: Comparative Results of All Datasets

Table 2: Comparative Analysis of Results

Dataset	Accuracy	Precision	Recall	F1 measure
AAM AADMI PARTY	1	1	1	1
BJP	0.73	0.71	1	0.83
Congress	0.6	0.71	0.55	0.62

V. CONCLUSION

Sentiment analysis is a rapid and low-costmethod for general election supervision and itsestimations. Diverse sort's information is created from online Social media organizations that should be prepared and to watch an individual's point of view toward items, objects, political reviews, and so on as it is realized that the issue with data gain is the qualities with an enormous number of qualities. In this work, we extricated new highlights that strongly affect deciding the extremity of the political reviews and applied calculation etymological techniques for the preprocessing of the information. In our proposed work we have improved the accuracy by utilizing VADER and we have additionally demonstrated that expanding the size of the dataset will likewise improve the precision and different boundaries. Later on, we might want to assess the competence of the proposed sentiment analysis and methods for different undertakings, for example, conclusion order. We might want to apply inside and out ideas of NLP for a better forecast of the extremity of the report. We might likewise want to stretch out this procedure to different areas of sentiment mining likes textmining, political conversation discussions, and so forth. This investigation has trade withonly the English speech. Whilea lot of of the perceptions would pertain to other speeches, particularly roman speeches, this stays untried. Further domain-definiteemotion lexicons in another fields could also be investigated, which might be utilized in revolve for advanceinvestigation so that the restrictions of the this research in the domain of SA could be determined forward.

REFERENCES

[1] Blenn, N., Charalampidou, K., Doerr, C.: Context-sensitive sentiment classification of short colloquial text. In: Proceedings of IFIP'12, pp. 97–108, Prague, Czech Republic (2012)

[2] Mittal, N., Agarwal, B., Agarwal, S., Agarwal, S., Gupta, P.: A hybrid approach for twitter sentiment analysis. In: 10th International Conference on Natural Language Processing (ICON), pp. 116–120 (2013).

[3] T. Nasukawa, "Sentiment Analysis: Capturing Favorability Using Natural Language Processing Definition of Sentiment Expressions," pp. 70–77, 2003.

[4] K. Dave, I. Way, S. Lawrence, and D. M. Pennock, "Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews," 2003.

[5] Agarwal, B., Mittal, N.: Prominent feature extraction for review analysis: an empirical study. J. Exp. Theor. Artif. Intell. (2014). doi:10.1080/0952813X.2014.977830

[6] Subrahmanian, V.S., Reforgiato, D.: Ava: adjective-verb-adverb combinations for sentiment analysis. Intell. Syst. 23(4), 43–50 (2008)

[7] Turney, P.: Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of 40th Meeting of the Association for Computational Linguistics, pp. 417–424, Philadelphia, PA (2002)

[8] Romanyshyn, M.: Rule-based sentiment analysis of ukrainian reviews. Int. J. Artif. Intell. Appl. 4(4), 103–111 (2013)

[9] Kessler, J.S., Nicolov, N.: Targeting sentiment expressions through supervised ranking of linguistic configurations. In: 3rd International AAAI Conference on Weblogs and Social Media (2009)

[10] Bandyopadhyay, S., Mallick, K.: A new path based hybrid measure for gene ontology similarity.

IEEE/ACM Trans. Comput. Biol. Bioinform. 11(1), 116–127 (2014).

[11] Almatrafi, Omaima & Parack, Suhem & Chavan, Bravim. (2015). Application of location-based sentiment analysis using Twitter for identifying trends towards Indian general elections 2014. 10.1145/2701126.2701129.

[12] K. Ali, H. Dong, A. Bouguettaya, A. Erradi and R. Hadjidj, "Sentiment Analysis as a Service: A Social Media Based Sentiment Analysis Framework," *2017 IEEE International Conference on Web Services (ICWS)*, Honolulu, HI, 2017, pp. 660-667. doi: 10.1109/ICWS.2017.79

[13] Gupta, Deepak. (2019). Spam and Sentiment Analysis Model for Twitter Data using Statistical Learning.

[14] Sharma, Parul & Moh, Teng-Sheng. (2016). Prediction of Indian election using sentiment analysis on Hindi Twitter. 1966-1971. 10.1109/BigData.2016.7840818.

[15] A. Bermingham, and A. F. Smeaton, "Classifying sentiment in microblogs: Is brevity an advantage?," Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 1833–1836, Oct. 2010

[16] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan, "A system for real-time twitter sentiment analysis of 2012 us presidential election cycle," Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics, pp 115–120, July 2012.

[17] Singhal, Kartik & Agarwal, Basant & Mittal, Namita. (2015). Modeling Indian General Elections: Sentiment Analysis of Political Twitter Data. 10.1007/978-81-322-2250-7_46.

[18] Yue, L., Chen, W., Li, X. *et al.* A survey of sentiment analysis in social media. *Knowl Inf Syst* **60,** 617–663 (2019)