International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Exploring the Power of Social Media in Election Predictions

Abdul Manan Koli, Muqeem Ahmed

Abstract*: *The forecasting of election's outcome remained prevailed in prominence from pre-historic times and is still a delightful topic of the current era. The predictions of election results have been started from traditional methods to economic indicators and now is being swung by social media especially sentimental analysis. The present studies discuss the election forecasting methods carried out in diverse nations by the number of researchers till now. Furthermore, different number of approaches for electoral prediction using social media and economic dimensions has been investigated based on previous literature work. The main focus of this work is to study and examine various techniques, methods and parameters used for election predictions in distinct areas. Finally, we suggest some intelligent techniques which will be based upon some parameters such as the development agenda, party type and religionism etc for further modification in election prediction system, so as to enhance the accuracy of political forecasting globally

Index Terms: Big data, election predictions, data mining, and forecasting.

I. INTRODUCTION

The election is a process by which general public directly cast their ballots and choose their representative which hold public office for a specified time. There are total 195 countries in the world and out of these 195 countries, 123 have the democratic type of government [4]. The basal feature of democracy is election efficacy. Different countries have their own constitutions with mentioned norms for elections process like indirect election (eg. the United State of America), direct elections (eg. European Parliament) and also have different tenures like four years or five years, respectively. Therefore we can say that majority of the world's populations are now governed by democratic authority and regency, which is described as systems with citizen political participation, constraints on the power of the executive, and a guarantee of civil liberties and rights. Before the coming of scientific polling in 1936, betting odds in the United States related strongly to vote results [1]. Since 1936, persuasion public opinion poll through survey or direct communication has been a basic theatrical role of political forecasting. More recently, prediction markets have been formed, starting in 1988 with Iowa Electronic Markets[1] But with the advent of statistical and computational techniques especially Machine learning and data mining, electoral data have become increasingly easy to handle. It is no surprise, then, that election forecasting has become a big business, for polling firms, news organizations, and betting markets as well as for research scholars. Research scholar uses different techniques and methods for forecasting elections, some uses economics parameters [2] while other included social media[3] as a mere dimension for predictions with some computational techniques like classification regression and Mean Absolute error[4]. The era of predictions markets becomes more momentous when Mr. Barack Obama strategically used Social media chiefly twitter in his elections process [5]. Then following his footprint numerous political parties in different countries and research scholar applied this technique in their elections process and research work respectively, which is still a leading dimension in prediction markets. The advent of web 2.0 technologies provides the developer with a path to develop large no of social media applications, out of them the most trending are Twitter, Facebook, Flickr, Instagram and Whatsapp, etc. A huge number of citizens are interacting on these social media platform with respect to any issue be it economy, marketing, sports, cinemas or politics. The dissimilation and the vacillations between different social media platforms would make any system shaky and hard to anticipate anything [6]. This evolution of technologies especially with electronic devices gives social media user a platform where they can post their emotions or ideas in the form of tweets or likes within seconds and billions of users' read them in one single click world-wide. This has led the human detached from their immediate surroundings and more addict to social media sites. The national election is undoubtedly one of the most powerful events for any nation or state during which people express their thoughts or views on social media regarding different political parties, the researcher collects the opinions of peoples from such social media and makes predictions or assumptions for the outcome of the election. The invention of web 3.0 brought the revolution in web applications, as it is one of the most exceptional and transformative application developed so far because, web 3.0 will be more connected, open, and intelligent, with semantic Web technologies, distributed databases, natural language processing, machine learning, machine reasoning, and autonomous agents, one must define web 3.0 as a semantic web technologies integrated into large-scale web applications [7].

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Abdul Manan Koli*, School of Computer Science and Information Technology, MANUU (A central University) Hyderabad, India.
Muqeem Ahmed School of Computer Science and Information Technology, MANUU (A central University) Hyderabad, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

Similarly, the technology evolutions continue towards more advanced levels and developed one of the most intelligent techniques of era known as Big Data analytics.

Today we observe an enormous amount of data is being generated from different sources such as the mobile phone, social media applications, radio frequency identification and heavy machine etc at an extremely high speed. These data set ranges from gigabyte to zettabytes with different nature like structured Un-structured or semi-structured. Because of their complexity, volume and size traditional tools like RDMS fizzle to analyze and handle such data type. To overcome such challenges big data technology came into prominence, which can efficiently analyze and stored such data type[12]. So, big data technology is mainly associated, where advanced analytic techniques operate on Metadata set. Analyzing big data analytics is a challenging task as it involves large distributed files system, which should be fault tolerant, flexible and scalable[13]. Hadoop, Hive, Map reduce and spark are some of the most prominent tools of big data technology, which mostly deals with analytics work[14]. The analytics may be health care, product reviews, banking or education sector and political forecasting etc.

Table 1. Difference between web 2.0 and web 3.0

Web 2.0[8],[9]	Web 3.0[10],[11]
2004 to 2016	2016+
Tim O Reilly	Sir Tim Berners Lee
The social web	The semantic web
2nd generation web	3rd generation web
The document web	The metadata web
Google as catalytic	Semantic web organization as catalytic or netvibes.
It contains an abundance of information	It has intelligent information.
Mainly uses the read-write web	Mainly uses semantic web
Mainly focuses on communities	Mainly focuses on the individual.
Two ways web pages, 2D portals, wikis, videos etc	More mobile friendly, 3D, AI. Integrated games, education and business.
It mainly deals with blogs	Its main concern is with lifestream.
It uses Google keyword search for information retrieval.	It uses Databases with metadata for information retrieval.
Interactive advertisement	Behavioral advertisement

In the rest of this research work first, we will study some research work pertaining to election prediction carried out by original research scholar till now, with various number of tools and algorithms. Then finding of research work will be discussed in details. Finally, conclude this work by suggesting some techniques and parameters for predicting election results globally particularly for developing nations.

I. RELATED WORK

The authors in [15] used Twitter tool for predicting the prominence of legislators using LDA SVA and other logistic regression. The authors in [16] have classified twitter data into Positive and negative tweets for predicting poll trends of U.S Presidential held in 2016 using time-series model, LTS, and the multi-variable time-series approach. The researchers in [17] have accomplished sentiment analysis of tweets collected from Twitter using Tweetinvi API for 2016 Spanish General Election. While some authors suggest that indeed Politician exploited the potential of social network sites for their political crusades [18] [19] and [20]. A comprehensive and definite summary of previous literature survey with their corresponding works, methods or tools used with outcomes is discussed below in table 2.

Table 2. A Comprehensive Summary of previous literature survey from 2008 to 2018.

Years	Authors	Methodology	Outcomes
2008	G. S. Gill[21]	Testing and training with survey and election data using Neural Network algorithm.	The researcher builds a Neural Network for forecasting India Lok Sabha election using testing and training phase with the survey from general public and data of previous nine elections. After scrutinizing properly the investigator asserted that this model can be used for predicting election outcomes.

2008	Robert S. Erikson and Christopher Wlezien [2]	Economic and Non-economic parameters like unemployment, inflation and president approval rate were used with root mean squared error.	This model was build using economic and non-economic parameters for forecasting the presidential elections. After a precise analysis of parameters (both economics and non-economics), the researcher concluded it that the proposed methods can be exploited for gauging election outcomes.
2008	Jose Manuel Pavía, Beatriz Larraz & Jose Marí Montero[22]	Spatiotemporal Models with two geostatistical techniques i.e ordinary kriging and ordinary cokriging were used were used for predictions.	The researcher developed this model by considering geographical factor with other traditional dimension like economic or development work for forecasting elections outcomes. The main theme of this research work is first, to analyze the importance of geographical factor and then recommending that geographical factor must be considered for election forecasting with other traditional parameters.
2009	Gregg r. Murray Chris riley Anthony scime [23]	Iterative Expert Data Mining techniques with Decision tree Algorithms (CHAID) and SPSS tool were used.	This exploration work forecast presidential election by choosing only two variable such as the intention to vote and previous presidential vote. After using the Iterative Expert Data Mining technique to classify likely voters, this model predicts the election outcomes with 78% accuracy rate.
2009	Steven E. Rigdon et. al, [24]	Bayesian estimation and Dynamic programming algorithm.	The researcher proposed a methodology for predicting the U.S. presidential election held in 2004. The authors collected state-level election data and generated prophecies for presidential level by applying Bayesian algorithms. After proper scrutinizing the state level poll results with Presidential level, they revealed that this method can be used for election forecasting purpose.
2010	J. Scott Armstrong et. al, [25]	Rating of facial expression was used for making predictions among politicians. The methods applied for rating are as follows: 1. Online betting market. 2. Intrade.com. 3. Facial competence.	The researcher built this model for gauging the united states presidential election, using the facial expression of different legislators. After mining the collected sample of facial expression, the authors revealed that Mr Obama had received the highest facial rating and has the maximum chance of winning the election.
2010	Helmut Norpotha and Thomas Gschwendb[26]	Importance of vital parameters such as 1) the popularity of the incumbent chancellor, (2) the long-term partisan balance in the German electorate. (3) The cost of ruling and measure their coefficient since 1953 to 2005 were investigated for predictions with mean absolute error and standard error.	The investigators build a chancellor model for forecasting Germany state election by testing the three predictors on previous 15 elections from 1953 to 2005. After proper mining, they realized that the chancellor model predicts election outcomes accurately, out-of-sample forecast with an average error rate less than (1.3%).
2011	Michael Steven Lewis-Beck a, Richard Nadeau [27]	Economical dimensions viz (valence, position, and patrimony) were tested with Logistic Regression equations, Bivariate correlation and Robust standard errors	This model was built for forecasting U.S presidential election using Economical dimensions. After scrutinizing properly the economic dimensions, the researcher revealed that economic factor like valence, position and patrimony mostly favour Mr. Obama and thus he has the maximum chance of winning the election.
2011	Michael S. Lewis-Beck et. al.,[28]	Opinion polls with statistical methods with R squared were used for forecasting.	The researcher built an election prediction model based upon opinion survey. After analyzed the collected sample of opinion survey with previous election results the investigators asserted that the proposed model may be used for political forecasting

2012	Lei Shi et. al.,[29]	Sentimental analysis of Twitter data was carried out by using Lasso (Least Absolute Shrinkage and Selection Operator) regression algorithm.	The authors propose a prediction model for forecasting Republican presidential election held in 2011 using Twitter data. Then the collected results were compared with the Realclearpolitics website and finally, they revealed that this method can be used for forecasting election outcome.
2012	Prof. T. M. Kodinariya and Mr. Ravi Seta [30]	Visualizing data mining techniques by creating a data warehouse of election data, which consist of an 8-Dimensional cube such as Word, Candidate, Time, Voter Education, religion, age, and Session.	The researcher developed an election Awareness model using 8 parameters such as Word, Candidate, Time, voter, Education, religion, age, and session with one developing nations India, So as to strengthen awareness among the voters in developing countries.
2013	Stefan Dahlberg[31]	Exploring the impact of context matter on voter perceptions by collecting opinion survey for predictions by applying Gallagher's least-square index and Bivariate analysis.	The researcher developed a model using a small number of hypothesis with only three variable like system related, party related and individual related factors for predicting voter perceptions. After examined properly the authors conclude that system related variables have a small impact on voters' perceptions while the party level and individual-level variables exerted the greatest impact on voter perceptions.
2013	Fernanda Pimenta et.al,[3]	Comparative Study of Social Media Platform viz (Blogs, Facebook, Twitter and YouTube) with data mining techniques such as mean absolute error was used for election predictions.	An analysis was performed solely on Republican candidates and exclude the Democratic party candidates for forecasting election outcomes. Then the potentiality of social media was correlated using only three metric i.e. volume, attention and popularity. Finally, it was concluded that the blogs were the most stable and reliable predictor for making the final conclusion.
2013	Spyros E. Polykalas et.al, [32]	The potential of Google trends was used in analyzing the elections outcomes with data mining algorithm and mean absolute error.	The main theme of this research work is to explore the potential of Google Trends in analyzing the election outcomes of Germany state. The researcher selected two major political parties i.e (Christian democratic party) and (social democratic party) with election viz (2005) (2009) and (2013) for analysis. After the proper investigation, they revealed that the web-search engine behaviour of potential voters be connected with the final election results. used for making the prediction of the forth coming elections.
2014	Andrea Ceron et. al,.[33]	The sentimental analysis in form of tweets was performed on American and Italian Presidential candidates using Hopkins and King methods with Mean absolute errors.	The researcher built an election prediction model using Twitter data. The investigators collected the samples of Twitter data and compared it to traditional polls and finally, they revealed that the proposed system has better prediction results as compared to other systems. Therefore it can be used for making election forecasting.
2014	Malhar Anjaria, Ram Mohana, and Reddy Guddeti[34]	Sentimental analysis of Twitter data was performed using supervised machine-learning techniques such as support vector machines (SVM), Naive Bayes, maximum entropy and artificial neural network with unigram, bigram for feature extraction.	The authors propose an election prediction models using twitter data with supervised machine-learning techniques. After comparing the results of all the three algorithms viz (SVM), (Naïve Bayes) and (ANN) they revealed that SVM outperformed in comparisons to other algorithms and have higher accuracy in election predictions.
2015	Yoonjae Nam, Yeon-Ok Lee, and HanWoo Park. [35]	Network analysis and linear regression were performed on Twitter, facebook, online news and blogs for election forecasting with one way ANOVA test	The primary objective of this research work is to use web ecology from social media platform in analyzing the Korean 2012 election. After properly mined the dataset from four social media, the researcher discovered that Twitter produced most biased results while online news remained least prejudiced.

2015	Rahman Ullah, Abdur Rashid Khan and Muhamma d Irfan [36]	Supervised learning technique such as Naïve Bayes was used for performing sentimental analysis of tweets in form of Positive, Negative and Neutral tweets.	The researcher used twitter data in anticipating the election results held in Pakistan. After mining the twitter data properly the researcher admitted that this method can be used for forecasting election outcomes.
2015	Quanzeng You, et. al [37]	Competitive Vector Auto Regression with multifaceted signals including textual and visual information, obtained from Flickr data was used for forecasting.	The authors proposed a model namely Competitive Vector Auto Regression (CVAR) for forecasting U.S Presidential and U.S House race elections held in 2012 and 2014 respectively. After performing different multifaceted signals including textual and visual information associated with the online shared photos in Flickr. The researcher found that the CVAR model outperformed as compared to other models like AR and VAR in predicting the U.S. election results.
2015	Adam Tsakalidis, Symeon Papadopou los, et.al,[4]	Lexicon based approach for fetching sentiments of peoples in form of tweets with linear regression, Gaussian process, and sequential minimal optimization for regression.	The researcher built this model for predicting the elections outcomes of three countries by considering twitter data. After Proper mining the data set, the authors revealed that Gaussian process achieved the lowest MAE (1.31), followed by sequential minimal optimization (1.35) and linear regression.
2016	Varsha D. Jadhav, and Sachin N. Deshmukh [38]	Sentimental analysis of Twitter data was carried out using Naïve Bayes classifier and Rtool with mean Absolute Error.	The investigator developed election prediction models by taking the Twitter data in the form of positive negative and neutral tweets. After Proper mining, they reveal that National Democratic Alliance got more Positive
			Tweets as compared to Grand Alliance. So National Democratic Alliance may win maximum seats in elections.
2016	Kabir Ismail Umar, and Fatima Chiroma[3 9]	Sentimental analysis of Twitter data using Polarity lexicon method with RTool.	The researcher built a forecasting model for U.S Presidential candidates held in 2016. They selected two leading competitors namely Donald Trump, and Hillary Clinton for their analysis work using Twitter data. After proper analyzing the user sentiments, the researcher acknowledges that Mr. Trump has a better chance of winning election as compared to Mrs. Clinton.
2016	Ming-Hun g Wang and Chin-Laun g Lei[40]	Analysis of Twitter data was carried out using sentimental analysis, Volume count and peer to peer rating with time series analysis and regression analysis.	The researcher builds a hybrid model using three indicators via, peer-to-peer ratings, sentiment scores, and candidate mentioning volumes for predicting 2014 Taiwanese local elections. After investigating properly the analyst discovered that the proposed regression model surpasses all other models in term of prediction, with more final winners (5 from 6) and an accuracy rate of 83%.
2017	Maurice Vergeer[41]	Comparative Analysis of Twitter data was carried out for multiple countries using negative binomial regression as well as Poisson regression restraint.	The main purpose of this research work is to identify how the political candidates raise their online popularity using twitter data. After a proper calculation of tweets dataset, the authors acknowledged that signing up Twitter as early as possible increase the candidate's follower as compared to signing up late.
2017	Ruth Dassonnev ille et. al,[42]	Three parameters from Economic perspective viz GDP, Unemployment, previous success and length of time in office with Ordinary least squares regression and mean absolute error were used for prediction	The Authors proposed a structural model for predicting the Netherland PM's party performance. After applying the Least Square Regression the writers conclude that their model performed fairly well in comparison to other poll methods with minimal least Square error.

2017	Vinay K. Jain, and Shishir Kumar [43]	Comparative analysis of news article and polarity of tweets was conducted with supervised learner classifiers like Naïve Bayes, Support vector machines random forest and decision tree.	The authors proposed a model in predicting the Delhi state election result held in 2015. They collected data using RSS (Rich Site Summary) feeds of news articles published in the leading newspapers and trending keywords from Twitter. After applying supervised learning algorithms, the proposed model predicts a clear victory with majority for AAM party with 79% accuracy.
2017	Mohammad Zolghadr[44]	Comparative study of learning algorithms such as (support vector machine and Artificial neural network) was carried out with some independent variable like GDP, unemployment rate etc in	The researcher developed this model for forecasting U.S. Presidential elections. A comparison between the learning algorithms (SVR and ANN) and Linear regression was performed and eventually, it was revealed that Support vector machine has better predictions results in contrast to other algorithms.
		order to increase the efficiency of prediction.	
2017	Jonathan Mellon and Christopher Prosser. [45]	Twitter and facebook data were used for accessing the attitudinal scale between social media user and non-social media using linear regression methods.	The researcher builds the model for measuring the attitudinal scale between social media user and non-social media user. After proper mining the data set, the researcher reveals that social media users are younger and better educated than non-social media users, and they are more liberal and pay more attention to politics.
2017	Xu Zhishuai Liu Wei [46]	Opinion mining was carried out by conducting survey and opinion poll from the general public with Apriori Algorithms.	The main aim of this research work is to improve the accuracy result of the forecasting model by applying the delegating and Weighting operations with Apriori algorithm for Hong Kong Legislative council election 2016. After mining the dataset collected from survey and opinion poll, the researcher concluded that this model can precisely anticipate the election results with 82.5% accuracy
2017	Harsh Ranjan and Motahar Reza [47]	Sentiment Analysis of Twitter data was carried out using Word2Vec Model with deep learning approach, decision tree and python language.	The researcher built this model for forecasting Gujarat state election, using Twitter data. After applying deep learning approach and decision tree they acknowledged it that this model can be employed for predictions purpose.
2018	Neetu N Narwal Kavitta Pabreja[48]	Sentimental analysis of Twitter data was performed for election predictions using K-means clustering algorithms and RTool.	The researcher developed an election forecasting model for Delhi MCD election using twitter data. After analyzing the twitter data for three parties namely BJP, Cong and AAP they revealed that BJP received maximum positive tweets and has the maximum chance of winning the election.
2018	Ana Cristina Bicharra et. al., [49]	Opinion mining was carried out from online leading newspapers using Multivariate linear regression technique with Weka tool and M5 methods.	The researcher built this model for forecasting Brazilian 2016 municipality election. The investigators performed opinion mining from the online newspaper and depict that more positive comments and likes in newspapers article mean candidates have the maximum chance of winning the elections.
2018	Ali Hasan et.al [50]	Sentimental analysis of Twitter was carried by comparing three sentimental analyzers viz, W-WSD, SentiWordNet,TextBlob with support vector classifier, Naïve Bayes and python language.	The researcher builds a classification model using Naïve Bayes and Support vector machine in weka tool. After Pre-processing the tweets in all the three sentimental analyzers, it was admitted that Text-Blob has better accuracy rate in contrast to other sentimental analyzer.
2018	Imane El Alaoui et. al,[51]	Sentimental analysis of Twitter data was performed for predicting the election outcomes using big data techniques like Apache Kafka, HDFC and Spark.	The authors proposed a novel adaptable approach that relies on social media posts and big data architecture to analyze the user views for 2016 U.S Presidential elections. After comparing the results of the proposed model with other models, they recognized that this model has higher prediction results.

2018	Bedour Sharar, Mostafa Abd-El-Barr [52]	Sentimental analysis of Twitter data and online survey such as survey monkey was carried out with the Statistical tool (SPSS) to explore the people's views regarding political participation.	The researcher built this model to scrutinize the use of social media and its impact on the political participation of Kuwaitis populations. After getting the data sample they acknowledged that the Kuwaitis population used social media frequently for sharing their thoughts to any political situation.
2018	Pritom Mazumder [6]	Principle Component Analysis and Adaptive Neuro-Fuzzy Inference System with Matlab and python language were used for forecasting popularity of presidential candidates.	The Authors proposed a model for forecasting the popularity of presidential candidates in a single day using twitter data. After a proper analysis, the researcher asserted that this model can be used for predicting popularity with 83% accuracy rate.
2018	Pranay Patel [53]	Psychometric Big Data Analytic which include a third-party firm Cambridge analytica adopted in Donald Trump Campaign.	The main theme of this research work is to explore the campaign strategies employed by Donald Trump in the U.S presidential election of 2014. The researcher had found that the real reason behind Mr. Trump victory was just to adopt the Psychometric Big Data Analytic a third-party firm in his election campaigns. This Psychometric Big Data Analytic firm strategically guides Mr. Trump and his teams during the election campaign, which ultimately leads to his victory in the election.

III FINDING OF THE LITERATURE SURVEY:

From the above survey, we did not find any suitable election forecasting models. However, many of them attempt the good, but the level of predictions still remained low and fluctuating. Mostly all of them done sentimental analysis using Social media data specifically Twitter, Facebook and Flickr data. However, some others used critical parameters like previous development records, GDP, Unemployment rate, number of time in office etc with data mining algorithms such as Decision tree, support vector machine and neural network etc. The prime problem which we encountered with the sentimental analysis is, it produces biased results in developed nations [54] and erroneous predictions in developing nations [55]. Research scholars like [56], [57] also suggests that one cannot predicts election results based on social media, as they are not suitable predictors of political forecasting. Some limitation of social media data pertaining to sentimental analysis are listed below:

1. Fake ids.
2. Concepts of bots.
3. Mostly used by elite class (Twitter).
4. Random sampling is performed on a selected portion of populations who used it and does not include every human being who has the right to vote.
5. Presence of rumors and misleading information, not all the social media posts are necessarily trustworthy.

The technicalities mostly adopted for sentimental analysis using social media data is listed below in table3:

Table3. Algorithms, Tools and methods used in existing survey.

Algorithms	Tools	Error Methods
Decision tree: I Iterative Dichotomiser 3.(ID3) ii Classification and Regression Tree.(CART) iii Chi-squared Automatic Interaction Detector. (CHAID)	Weka	Mean Absolute Error.
Support Vector Machine: i..C-SVM classification. ii.nu SVM classification.	Rtool	Root Mean Squared Error.

iii epsilon-SVM regression. iv nu-SVM regression		
Naïve Bayes: I Gaussian Naive Bayes classifier. ii Multinomial Naive Bayes. iii Bernoulli Naive Bayes.	Hadoop	Root Mean Squared Error.
K-Means clustering Algorithms	Rapid miner	Ordinary Least Square methods.
Artificial Neural Network: I Feed Forward Neural Network Artificial Neuron. ii Radial basis function Neural Network. iii Convolutional Neural Network. iv Modular Neural Network.	Matlab	Standard Error.

A detail and comprehensive study of various Tools and Algorithms used in the literature review are listed below in Table 4 and Table 5 respectively.

Table 4. Comparative studies of various Tools.

Tool Name	Matlab [58]	Rtool [59], [60]	Rapid miner [61]	Weka [62], [63]	Hadoop [12], [14], [64]
Release data	1984	1993	2006	1993	2006
Area	Computational analysis	Statistical	Statistical and analytic	Machine learning	Predictive analytics and machine learning
Availability	Not open sourced	Open source	Not open source	Open source	Open source
License	Proprietary commercial software	GNU General Public License	AGPL Proprietary	GNU General Public License	GNU General Public License
Language	Matlab	R	Language independent	Java	Java
Interface	GUI	Both CLI and GUI	GUI	GUI	GUI
Deep learning supported	Yes	Yes	Yes	No	Yes

Table 5. Comparison of various Algorithms.

Algorithms	General Accuracy	Speed of classification	Parameter handling	Capacity to handle Noise data	Training speed
Decision tree [65], [66]	Medium	Very High	High	Medium	Very High
Neural network [67], [68]	Medium	Very High	Low	Medium	Medium
Naïve Bayes [69]	Low	Very High	Very High	High	Very High
K-NN [70], [71]	Medium	Low	High	Low	Very High
Support vector machine [72], [73]	Very High	Very High	Low	Medium	Medium

The above literature also suggests that sentimental analysis is usually performed in developed nations may be used for predictions, but in case of developing nations, it may lead to vague forecasting. A comparative study of Developed and Developing Nations is listed below in fig 1.

Table 6. A comparative study of Developed and Developing Nations.

	Developed [74]	Developing[75]
Literacy rate	***	**
Human development indexed rate.	***	*
Industrial worker	***	*
Agricultural worker	*	***
Technology usage	***	**
Apps developed (social media)	***	*
Social unrest	*	***

Note (* star represent low, ** star represent medium, and *** star represent highest levels)

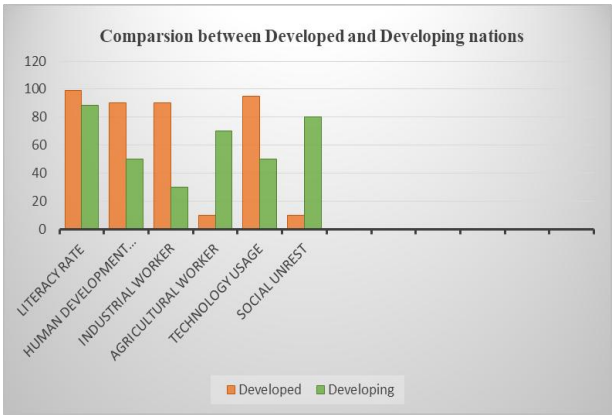

Fig1. Diagrammatically comparison of developed and developing nations.

IV CONCLUSION AND FUTURE WORK

After analyzing the different multifaceted techniques and model used for election prediction, we conclude our work by expounding that we did not find any single suitable model or classifier which can accurately predict the election results globally. Most of the researcher used social media like Twitter or Facebook for election prediction commonly known as sentimental analysis. But the main limitations of sentimental analysis is that, it can only be used in developed nations, like United States of America, Japan and South Korea etc where there is a high literacy rate, and people are frequently accessing the internet for their daily life activity. And these methods are not suitable for developing countries like India, Pakistan, China and Srilanka etc where there is less access to the internet and fewer numbers of people use social media especially twitter. For such nations, a model should be developed or propose which is merely not based upon social media but upon certain important parameters with an advanced level of computational techniques like Machine learning, big data analytics and python language. In a country like India where there is large no of social unrest, and some areas witnessed highest levels of disturbances like the state of Arunachal Pradesh, Assam, Jammu Kashmir and Manipur etc, one cannot predicts election results using only social media. To deals with these challenges, election predictions should be performed upon vital parameters like the previous record of governments, GDP, developments agenda of parties, religions, educations level, health factor etc with social media as its one component. Like in our model we will forecast the election of Jammu and Kashmir State which is based upon certain important parameters like development agenda, party type, social unrest level and religionism etc. To evaluate the model, we use Machine Learning hybrid approach which consists of classification (Support vector machine) or clustering algorithms (K Mean Clustering)) and implementation should be done using python code. The prediction model so developed will be based upon past election results with present situations. This model can also be applied to other areas which possess the similar type of geographical and cultural habitat.

.

REFERENCES

1. R. S. Erikson and C. Wlezien, “Markets vs. polls as election predictors: An historical assessment,” Elect. Stud., vol. 31, no. 3, pp. 532–539, 2012.
2. R. S. Erikson and C. Wlezien, “The economy and the presidential vote: What leading indicators reveal well in advance,” Int. J. Forecast., vol. 24, no. 2, pp. 218–226, 2008.
3. F. Pimenta, D. Obradović, and A. Dengel, “A comparative study of social media prediction potential in the 2012 U.S. Republican presidential preelections,” Proc. - 2013 IEEE 3rd Int. Conf. Cloud Green Comput. CGC 2013 2013 IEEE 3rd Int. Conf. Soc. Comput. Its Appl. SCA 2013, pp. 226–232, 2013.
4. A. Tsakalidis, S. Papadopoulos, A. I. Cristea, and Y. Kompatsiaris, “Predicting Elections for Multiple Countries Using Twitter and Polls,” IEEE Intell. Syst., vol. 30, no. 2, pp. 10–17, 2015.
5. A. Tumasjan, T. Sprenger, P. Sandner, and I. Welpe, “Predicting elections with Twitter: What 140 characters reveal about political sentiment,” Proc. Fourth Int. AAAI Conf. Weblogs Soc. Media, pp. 178–185, 2010.
6. P. Mazumder, N. A. Chowdhury, and M. Anwar-ul-azim, “Modern Approaches for Intelligent Information and Database Systems,” vol. 769, pp. 63–74, 2018.
7. J. Hendler, “Web 3.0 emerging,” Computer (Long. Beach. Calif)., vol. 42, no. 1, pp. 111–113, 2009.
8. D. a. Collier et al., “A comparison of common programming languages used in bioinformatics,” BMC Bioinformatics, vol. 9, no. 6, pp. 1–5, 2014.
9. Y. H. Ching and Y. C. Hsu, “Design-grounded assessment: A framework and a case study of Web 2.0 practices in higher education,” Australas. J. Educ. Technol., vol. 27, no. 5, pp. 781–797, 2011.
10. KaranPatel, “Incremental Journey for World Wide Web: Introduced with Web 1.0 to Recent Web 5.0 – A Survey Paper,” Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 3, no. 10, pp. 410–417, 2013.
11. M. R. Solanki and A. Dongaonkar, “A Journey of Human Comfort: Web 1.0 to Web 4.0,” Int. J. Res. Sci. Innov., vol. III, no. September 2016, pp. 75–78, 2016.
12. D. Lopez and G. Manogaran, “A survey of big data architectures and machine learning algorithms in healthcare,” Int. J. Biomed. Eng. Technol., vol. 25, no. 2/3/4, p. 182, 2017.
13. M. Chen, S. Mao, and Y. Liu, “Big data: A survey,” Mob. Networks Appl., vol. 19, no. 2, pp. 171–209, 2014.
14. A. Castiglione, M. Gribaudo, M. Iacono, and F. Palmieri, “Exploiting mean field analysis to model performances of big data architectures,” Futur. Gener. Comput. Syst., vol. 37, pp. 203–211, 2014.
15. P. Saleiro and C. Soares, “Learning from the news: Predicting entity popularity on twitter,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9897 LNCS, pp. 171–182, 2016.
16. U. Pavalanathan and J. Eisenstein, “Emoticons vs. Emojis on Twitter: A Causal Inference Approach,” no. November, 2015.
17. P. Singh, R. S. Sawhney, and K. S. Kahlon, “Predicting the outcome of spanish general elections 2016 using twitter as a tool,” Communications in Computer and Information Science, vol. 712. pp. 73–83, 2017.
18. C. B. Williams and G. J. J. Gulati, “Social networks in political campaigns: Facebook and the congressional elections of 2006 and 2008,” New Media Soc., vol. 15, no. 1, pp. 52–71, 2013.
19. Y. Kim, “The contribution of social network sites to exposure to political difference: The relationships among SNSs, online political messaging, and exposure to cross-cutting perspectives,” Comput. Human Behav, vol. 27, no. 2, pp. 971–977, 2011.
20. Y. M. Baek, “Political mobilization through social network sites: The mobilizing power of political messages received from SNS friends,” Comput. Human Behav, vol. 44, pp. 12–19, 2014.
21. G. S. Gill, “Election Result Forecasting Using Two Layer Perceptron Network,” 2008.
22. J. M. Pavía, B. Larraz, and J. M. Montero, “Election forecasts using spatiotemporal models,” J. Am. Stat. Assoc., vol. 103, no. 483, pp. 1050–1059, 2008.
23. G. R. Murray, C. Riley, and A. Scime, “Pre-election polling: Identifying likely voters using iterative expert data mining,” Public Opin. Q., vol. 73, no. 1, pp. 159–171, 2009.

24. S. E. Rigdon, S. H. Jacobson, W. K. Tam Cho, E. C. Sewell, and C. J. Rigdon, "A Bayesian prediction model for the U.S. presidential election," Am. Polit. Res., vol. 37, no. 4, pp. 700–724, 2009.
25. J. . and M. W. Kesten C. Greene and J. Scott Armstrong and Randall J. Jones, "Predicting Elections from Politicians%99 Faces," vol. 22, no. 4, 2010.
26. H. Norpoth and T. Gschwend, "The chancellor model: Forecasting German elections," Int. J. Forecast., vol. 26, no. 1, pp. 42–53, 2010.
27. M. S. Lewis-Beck and R. Nadeau, "Economic voting theory: Testing new dimensions," Elect. Stud., vol. 30, no. 2, pp. 288–294, 2011.
28. M. S. Lewis-Beck and M. Stegmaier, "Citizen forecasting: Can UK voters see the future?," Elect. Stud., vol. 30, no. 2, pp. 264–268, 2011.
29. L. Shi, N. Agarwal, A. Agrawal, R. Garg, and J. Spoelstra, "Predicting US Primary Elections with Twitter," pp. 1–8, 2012.
30. P. T. M. Kodinariya, "Visual Data Mining in Indian Election System," vol. 4, no. 07, pp. 1323–1329, 2012.
31. S. Dahlberg, "Does context matter - The impact of electoral systems, political parties and individual characteristics on voters' perceptions of party positions," Elect. Stud., vol. 32, no. 4, pp. 670–683, 2013.
32. S. E. Polykalas, G. N. Prezerakos, and A. Konidaris, "An algorithm based on Google Trends' data for future prediction. Case study: German elections," IEEE Int. Symp. Signal Process. Inf. Technol., pp. 000069–000073, 2013.
33. A. Ceron, L. Curini, and S. M. Iacus, "Using Sentiment Analysis to Monitor Electoral Campaigns: Method Matters—Evidence From the United States and Italy," Soc. Sci. Comput. Rev., vol. 33, no. 1, pp. 3–20, 2015.
34. M. Anjaria and R. M. R. Guddeti, "A novel sentiment analysis of social networks using supervised learning," Soc. Netw. Anal. Min., vol. 4, no. 1, pp. 1–15, 2014.
35. Y. Nam, Y. O. Lee, and H. W. Park, "Measuring web ecology by Facebook, Twitter, blogs and online news: 2012 general election in South Korea," Qual. Quant., vol. 49, no. 2, pp. 675–689, 2014.
36. R. Ullah and M. Irfan, "Forecasting Political Weather for Pakistan Local Government Election-Using Opinion Mining," vol. 2, no. 12, 2015.
37. Q. You, L. Cao, Y. Cong, X. Zhang, and J. Luo, "A Multifaceted Approach to Social Multimedia-Based Prediction of Elections," IEEE Trans. Multimed., vol. 17, no. 12, pp. 2271–2280, 2015.
38. V. D. Jadhav and S. N. Deshmukh, "Web Site : www.ijettcs.org Email : editor@ijettcs.org with Twitter Using Bayesian Classifier AND TRAINING," vol. 5, no. 2, pp. 2015–2017, 2016.
39. K. Ismail, "Data Mining for Social Media Analysis:Using Twitter to Predict the 2016 US Presidential Election," Int. J. Sci. Eng. Res., vol. 7, no. 10, pp. 1972–1980, 2016.
40. M. H. Wang and C. L. Lei, "Boosting election prediction accuracy by crowd wisdom on social forums," 2016 13th IEEE Annu. Consum. Commun. Netw. Conf. CCNC 2016, pp. 348–353, 2016.
41. M. Vergeer, "Adopting, Networking, and Communicating on Twitter: A Cross-National Comparative Analysis," Soc. Sci. Comput. Rev., vol. 35, no. 6, pp. 698–712, 2017.
42. R. Dassonneville, M. S. Lewis-Beck, and P. Mongrain, "Forecasting Dutch elections: An initial model from the March 2017 legislative contests," Res. Polit., vol. 4, no. 3, 2017.
43. V. K. Jain and S. Kumar, "Towards Prediction of Election Outcomes Using Social Media," Int. J. Intell. Syst. Appl., vol. 9, no. 12, pp. 20–28, 2017.
44. M. Zolghadr, S. A. A. Niaki, and S. T. A. Niaki, "Modeling and forecasting US presidential election using learning algorithms," J. Ind. Eng. Int., 2017.
45. J. Mellon and C. Prosser, "Twitter and Facebook are not representative of the general population: Political attitudes and demographics of british social media users," Res. Polit., vol. 4, no. 3, pp. 1–9, 2017.
46. Z. Xu and W. Liu, "Apriori-based prediction of the multi seat elections," 2017 2nd IEEE Int. Conf. Comput. Intell. Appl. ICCIA 2017, vol. 2017–Janua, pp. 214–218, 2017.
47. H. Ranjan and M. Reza, "Predictive Analytics on Gujrat Assembly Election 2017 Using Twitter Sentiment Analysis based on Word2Vec Model," no. December, pp. 2–4, 2017.
48. O. Anth, "Volu y 2018 REV VIEW ARTI vailable On nline at ww ational Jou urnal of Ad dvanced Re esearch in Computer Science INTRUSI ION DET TECTION N IN INTE ERNET OF O THIN NGS (IOT) The Internet of Things," pp. 504–510, 2018.
49. A. C. B. Garcia, W. Silva, and L. Correia, "The PredNews forecasting model," Proc. 19th Annu. Int. Conf. Digit. Gov. Res. Gov. Data Age - dgo '18, pp. 1–6, 2018.
50. A. Hasan, S. Moin, A. Karim, and S. Shamshirband, "Machine Learning-Based Sentiment Analysis for Twitter Accounts," Math. Comput. Appl., vol. 23, no. 1, p. 11, 2018.
51. I. El Alaoui, Y. Gahi, R. Messoussi, Y. Chaabi, A. Todoskoff, and A. Kobi, "A novel adaptable approach for sentiment analysis on big social data," J. Big Data, vol. 5, no. 1, 2018.
52. B. Sharar, "Citizens ' Perspective on the Impact of Social Media on Politics in Kuwait," 2018.
53. B. D. Drew and D. Weaver, "V Oter L Earning in the 2004 P Residential E Lection :," Spring, vol. 83, no. 1, pp. 25–42, 2006.
54. A. J. Wicaksono, Suyoto, and Pranowo, "A proposed method for predicting US presidential election by analyzing sentiment in social media," Proceeding - 2016 2nd Int. Conf. Sci. Inf. Technol. ICSITech 2016 Inf. Sci. Green Soc. Environ., pp. 276–280, 2017.
55. T. Mahmood, T. Iqbal, F. Amin, W. Lohanna, and A. Mustafa, "Mining Twitter big data to predict 2013 Pakistan election winner," 2013 16th Int. Multi Top. Conf. INMIC 2013, pp. 49–54, 2013.
56. D. Gayo-Avello, P. Metaxas, and E. Mustafaraj, "Limits of electoral predictions using twitter.," Icwsm, pp. 490–493, 2011.
57. C. Lui, P. T. Metaxas, and E. Mustafaraj, "On the Predictability of the U.S. Elections Through Search Volume Activity," Proc. IADIS Int. Conf. e-Society, vol. c, pp. 1–9, 2011.
58. J. J. Granda, "The role of bond graph modeling and simulation in mechatronics systems - An integrated software tool: CAMP-G, MATLAB-SIMULINK," Mechatronics, vol. 12, no. 9–10, pp. 1271–1295, 2002.
59. H. E. Harrison, S. P. Schaefer, and T. S. Yoo, "No Title," vol. 88, no. June 1988, pp. 20–24.
60. A. Mariano, "Pion photoproduction in an isobar consistent model," Nucl. Phys. A, vol. 790, no. 1–4, pp. 267–271, 2007.
61. H. Kosorus, J. Honigl, and J. Kung, "Using R, WEKA and RapidMiner in Time Series Analysis of Sensor Data for Structural Health Monitoring," 2011 22nd Int. Work. Database Expert Syst. Appl., pp. 306–310, 2011.
62. N. Sharma, A. Bajpai, and R. Litoriya, "Comparison the various clustering algorithms of weka tools," Int. J. Emerg. Technol. Adv. Eng., vol. 2, no. 5, pp. 73–80, 2012.
63. E. Frank, M. Hall, G. Holmes, R. Kirkby, B. Pfahringer, and H. Ian, "Data Mining and Knowledge Discovery Handbook," pp. 0–1, 2010.
64. Y. Guo, J. Rao, D. Cheng, and X. Zhou, "IShuffle: Improving hadoop performance with shuffle-on-write," IEEE Trans. Parallel Distrib. Syst., vol. 28, no. 6, pp. 1649–1662, 2017.
65. S. S. Nikam, "A Comparative Study of Classification Techniques in Data Mining Algorithms," Orient. J. Comput. Sci. Technol., vol. 8, no. 1, pp. 13–19, 2015.
66. "Decision Tree Algorithm Team Homework Assignment # 5."
67. L. Liu and M. Deng, "An evolutionary artificial neural network approach for breast cancer diagnosis," 3rd Int. Conf. Knowl. Discov. Data Mining, WKDD 2010, vol. 25, pp. 593–596, 2010.
68. A. Facchinetti et al., "Artificial neural network algorithm for online glucose prediction from continuous glucose monitoring," Diabetes Technol. Ther., vol. 12, no. 1, pp. 81–8, 2010.
69. C. Catal, U. Sevim, and B. Diri, "Practical development of an Eclipse-based software fault prediction tool using Naive Bayes algorithm," Expert Syst. Appl., vol. 38, no. 3, pp. 2347–2353, 2011.
70. T. Dalgleish et al., "[Compartive Study of data mining tools]," J. Exp. Psychol. Gen., vol. 136, no. 1, pp. 23–42, 2007.
71. M. Weis, T. Rumpf, R. Gerhards, and L. Plumer, "Comparison of different classification algorithms for weed detection from images based on shape parameters," Image Anal. Agric. Prod. Process. vol. 69, no. 2005, pp. 53–64, 2009.
72. L. Jiao, S. Member, L. Bo, and L. Wang, "Support Vector Machine," vol. 18, no. 3, pp. 685–697, 2007.
73. S. B. Kotsiantis, "Supervised Machine Learning: A Review of Classification Techniques," Informatica, vol. 31, pp. 249–268, 2007.
74. https://dictionary.cambridge.org/dictionary/english/developed-country.
75. https://en.wikipedia.org/wiki/Developing_country pp. 1–10.

AUTHORS PROFILE

Abdul Manan Koli completed his BSC from University of Jammu and MCA from University of Jammu. He is currently pursuing PhD from the department of CS&IT, MANUU, Hyderabad. He has published several conference and journal paper in national and international journals.

Dr. Muqeem Ahmed is an Assistant Professor in the Department of Computer Science & Information technology at Maulana Azad National Urdu University, Hyderabad (Telangana State). He received his Master's degree in Master of Computer Applications from Jamia Millia Islamia University, Delhi. He completed his PhD degree from Jamia Millia Islamia University, Delhi. His research interest include Semantic web Applications, Machine learning, data mining. He has many publications in journals of international repute including national journals, conferences and proceedings of the international conference. He is a member of editorial board for national and international journals and program committee member/reviewer international conferences.