

## Political Sentiment Analysis Using Social Media Data: A Case Study

Jwala Sharma<sup>1</sup>, Samarjeet Borah<sup>2</sup> and Author Name<sup>3\*</sup>

<sup>1</sup>Department of Information Tech., DDE Sikkim Manipal University, Sikkim, India <sup>2</sup> Department of Computer Appl., SMIT, Sikkim Manipal University, Sikkim, India

<sup>3</sup>Author Affiliation(s) Department, University, Country

<sup>1</sup> jwalachapagai@gmail.com, <sup>2</sup>samarjeetborah@gmail.com, <sup>3</sup>E-mail

### Abstract

*Social network has provided a large platform for online communication and interaction. In the last recent years, it has gained much more popularity than traditional Television. Users around the world spend ample amount of time on social media, sharing post, tweets, videos and discussing political and social events around. This popularity has attracted numbers of spammer that are responsible for generating contents, redistributing of such contents, which in general we can say automating human activities with bot amplifier. The consequences of these kind of activities has resulted in spreading rumors, havoc in social and public life. Bots are used in social media to provoke trending topics and spread a message in social media to favor a situation or to misrepresent the counterparts. This paper discusses few political events as case study. Impact of social media on these events is investigated based on available literature. In addition to this, possibility of adoption of sentiment analysis techniques is explored in order to track the same. Further, classification of causal factor in such events, i.e. human or bot is studied.*

**Keywords**— Social Network, Social Media, Social Bots, Sentiment Analysis, Election

### 1. INTRODUCTION

Social media has been the fastest and easiest way to share any kind of information. Social media has become very popular way to communicate and interact globally, with this popularity it has also attracted many spammers that generates contents and regenerates it that is leading to misunderstanding among public. Social media is playing huge role in each of the important spectrum that is related to either the nation's economy, democracy, election, business as a medium to spread information and influence the public. Almost no one can deny or escape the use of social media in any forms to exchange information. [1] Around 3.2 billion social media users are currently active across the globe and this number is in meteoric rise and it is adequate to 42% of current population. Social media contains all kinds of content and serves as a platform for marketers and business to promote their products and services. It allows mass cultural exchange and intellectual communication. Social media is not only confined to common people, politicians are also part of it, as to promote their party propaganda and when it comes to support major political events that happen across the nation, which is the most trending and important factor for shaping the democracy of the nation. One of the most threatening activities that are carried through social media are done through terrorism, ISIS use social media to spread radicalism by swaying youth to embrace and support their cause.

#### 1.1 ORGANIZATION OF THE PAPER

In this paper, study of major political events has been discussed to highlight the impact of social media and use of social bots in influencing and manipulating public opinions. Contents of the paper are organized in the following way: Section 1 discusses the impacts of social media and use of social bots. Section 2 contains literature review of two aspects:

---

Received:

Reviewed:

Accepted:

\* Corresponding Author

- Major political events discussed by socialist
- Political sentiment analysis: Techniques and some related works

In section 3 we have done analysis about impacts of social bots and have done comparative analysis about different detection methods and highlighted some factors to be considered while working in future and section 4 contains the conclusion. and future aspects. There is not conflict of interest in this paper.

## 1.2 SOCIAL BOTS AND IMPACTS

Today's internet is not merely a human to human conversation or interaction, bots are software that are programmed to automate users' activities and mimic human behaviour and perform simple and repetitive task more and more quickly than human. In social media bots are used for similar activities such as repost or retweet, they simply inject message to social media as in favour of or against of any political figure, or any kind of market products or services, as botnets are considered to be highly persuasive in public opinions, they automated to attract people with objective of making them adopt radical ideologies of terrorism. The ideas and ideologies that are created in social media are then augmented by tens of thousands of automated accounts called bots that are used for repurposing, retweeting, and republishing the same language, consequently when such automated accounts, those of large numbers are actually being run by a single user or group, they leave an impression that majority of the people believe in the same thing and as a result in this way one user or a group is successful in manipulating and swaying public opinion, which is quite a threat to democracy. The presence of social bots has a competence of having infinite patience to debate and convince any kind of online radical or political conversation. As a result, it is responsible for creating following issues in social media:

- Influence can be delivered across the suspicious accounts that is further operated with defamatory intention,
- The political discourse can be diverged
- The third issue would be that the unverified information will be amplified

## 2. LITERATURE REVIEW

This section cites some works from literature on impact of social media in various political events. As a case study few major political events are considered for the same.

### 2.1 SOCIAL MEDIA IN INDIAN POLITICS

We have already witnessed that social media has dominated the era of television in influencing the mass. [2] About 60% of Indians spend time on different social networking sites so, it is probable that social media has been playing and will continue to play a greater role in influencing the public. Social media is playing a vital role in changing the way the public used to conceive about politics. Many political leaders are actively using social media as a platform to convey their views, agenda of any topic. And they are no less than any celebrity in having millions of fans on Twitter and supporters on Facebook as well as Google+. In 2009 Sashi Tharoor Indian politician was the only Indian to have 6000 followers in Twitter, five years down the line Tharoor became second most with 2.16 million followers after Narendra Modi having 3.89 million followers. Many politicians are now using social media as a medium of campaign and spread their political agenda. The recent activities and agendas of political party BJP led by Narendra Modi got huge social media attention.

The idea behind BJP's social media strategy was to conduct, positive campaigning message of Modi, and the ideological and religious polarization. [3] A major attribute of the 2019 election is about collating digital election and populist politics and introducing app-based election by launching (NaMo) app. The NaMo app had initiated the people to take part in government policies and to sign up as volunteers for his programs by ringing up directly more or often. Social bots were used as an amplifier for promoting the party's ideologies and agenda.

If we happen to observe the pervious general election of 2014, [4] India had approximately 100m smartphone users among 1.3 billion citizens, this has tremendously developed to more than 400m. India has become one of the biggest users of Facebook with more than 200m active users, other social media apps are also being used by large number of Indian citizens. These apps and social media are the easiest way to reach to the common people and it has been used as a medium for disseminating political propaganda. The study suggests that the Bhartiya Janata Party had used social bots for elections and it was estimated that approximately 150,000 bots were deployed as cyber-soldiers in the state Uttar Pradesh alone, to amplify the work and spread the political agenda and its ideologies.

When Prime Minister, Narendra Modi visited Tamil Nadu, the mass quite actively took part in social media to express their opinion, this traffic was analyzed by “The Atlantic Council, an American think-tank”. The analyzation depicts that pro-modi and anti-modi hashtags were initiated by Bots per second, with pro-Modi bots amplifying the tweets three times harder. All these studies and investigation draws one major conclusion that political parties that used social media to respond to their supports and promote their ideologies and employed cyber warrior to help booth-level campaign turned out to be the most successful in election, where BJP secured 285 seats out of 545 seats.

## 2.2 2016 US PRESIDENTIAL ELECTION

Social media is changing the belief of our understanding of politics and democracy. Many studies have revealed that social media is responsible for enhancing the democratic conversation and political issues. Twitter and other social media apps provide enormous space for political parties to spread their agendas. The study has shown the active presence of bots being involved in 2016 US presidential elections to manipulate political discourse public opinion on twitter. Bots were used to inject messages to social media as to favour a political figure or to misinterpret that figure. The study [5] has investigated the roles and effects of social bots in 2016 US presidential election. Twitter data prior to election and all three presidential debates were being collected as a dataset, it contained election related content and hashtags of over 20 million tweets and about 2.8 million distinct twitter users. Data collection was done by manually crafting list of hashtags and keywords that were related or used more often in 2016 US presidential election. The hashtags and keywords such as, “#donalddrump, #trump2016, #neverhillary, #trump, #hillaryclinton, #imwithher, #nevertrump, #hillary”, were listed.

BotorNot is an openly accessible solution which provides a machine learning framework that is used to extract and examines a feature to detect presence of bots in social media as an automated account that were used to populate social media with election related conversation. BotorNot was trained with the features collected that contained about a set of over thousands of social bots features and as a result the accuracy of 95% or above was achieved and further it was found out that approximately 400,000 accounts were likely to be bots, which is 15% of total population used as a study.

## 2.3 2017 FRENCH PRESIDENTIAL ELECTION

One of the major political events talks about disinformation and social bot operations of 2017 French Presidential election. [6] Disinformation in this study is referred to contents of Macronleaks. Macronleaks unveiled two forms of content: unverified nature of content usually known as rumour and second form is synchronized effort behind sharing a spreading of information. The documents that were distributed online was believed to have evidence against Macron’s tax frauds and other illegitimate activities. Therefore, spreading such rumours and effort behind disseminating it makes it a Disinformation campaign.

The research work was conducted by collecting a dataset from Twitter search API, which was collected in the year 2017 in between 27th April to 7th May, till the day of election. The dataset contained about 17 million of unique tweets that was posted by over 2068728 million users Further they also created a subset of tweets associated with macronleaks disinformation campaign with list of hashtags and keywords used in “#MacronLeaks, #MacronGate, #SortonsMacron, #Bayrougate, and #RejoignezMarine”. Using machine learning techniques and cognitive behavioural modelling, the result has been found out that accounts that were used to support 2016 US presidential election candidate Trump, had been brought back to active state to join Macronleak disinformation Campaign.

As stated in [7] Facebook faced a mounting pressure from government around the world to control fake news and propaganda, as a result of analysis and research it states in a report “information operations”. In this report it is found that amplifier accounts were responsible for spreading false news stories to influence public opinion. 70,000 Facebook accounts were suspended in France that were promoting propaganda and spams

With immense pressure from government bodies and civil society France was successful to anticipate and manage its response between the staff of Macron campaign, the government, and civil society [8]. The effort of a foreign individual to interrupt the electoral process in recent years remained ineffective. These all studies and research points out the presence of black market for reusable political disinformation bots.

All cases of the above discussion involve some form of sentiment analysis. The same has been discussed in the following section in detail.

### **3. RELATED WORKS**

#### **3.1. POLITICAL SENTIMENT ANALYSIS: TECHNIQUES**

Fake news and bias have become very prominent in the news and the severity of attacks are quite dominant as compared to the countermeasure against them. [9] The data that is generated every day is almost about 2.5 quintillion bytes of data. The data are very dynamic in nature and can be applicable for different purposes, such as in commercial application it can be used for marketing analysis, public relations, product reviews, product feedback, customer feedback. These data are often found in unstructured format and to understand these data, it is required to convert these data into structured format. Social media is one of the popular platforms for exchanging ideas, and information all around the world. When it comes to have a political debates and politicians promoting their ideas and agendas, social media plays a vital role.

Therefore, to understand the subject that is being talked about in social media and to understand the opinion of public there is a need of analysis on the social media content. Social media contains dynamic information, and this information are growing exponentially. To understand and to identify the opinion, there is a need of system that can automatically extract the attributes used to express the opinion in social media and a system that can evolve and accommodate the changes. Currently, sentiment analysis is applicable in this field for transforming unstructured data into structured information of public opinions.

Sentiment analysis is the automated process of analysing text data and classifying opinions as negative, positive or neutral. It helps to extracts attributes that are often used to express the opinions. Sentiment analysis is a field of Machine Learning, where a Learning involves the algorithm to continually evolve to improve and adjust the itself. It is also known as opinion mining, that has become one of the key tools for making sense of the data. In the field of bot detection, this technique will be helpful for analysing the text and classify the user as bot or human. In this section of literature review, we have discussed few works which focus on analysing the social media text and then determine or classify the accounts used for spreading news or any activities belong to human or are automated bot accounts. Therefore, to highlight the techniques, most recent papers are considered, that include/considered major political events as cited in this paper for classifying the accounts as bots or human.

One of the major political events analysed is about disinformation and social bot operations of 2017 French Presidential election. [6] The research work was conducted by collecting a dataset from Twitter search API, which was collected between 27th April to 7th May in year 2017 till the election day. The dataset contained about 17 million of unique tweets that was posted by over 2068728 million users. Similarly, they have collected a subset of tweets associated with MacronLeaks disinformation campaign. In this work Machine-learning framework was proposed that produced a set of over 1000 features of, content and network structures, temporal activity, user profile data, sentiment analysis. After conducting feature analysis, two important classes of features were identified :metadata and usage statistics associated with user account ,that can be used for detecting social bots and they are: The user metadata and activity features that are used in this work are “status count, followers count,

friends count, listed count, default profile, geo enabler, profile-use-background-image, verified and protected” After testing and analysing variety of algorithms, they have considered the algorithm that can deal with large dataset. Support Vector Machine and logistic Regression, decision tree, some ensemble methods (Random forest, AdaBoost), K-nearest neighbours, two-layer neural network are selected. For the rest of the work they have decided to use Logistic Regression as it yielded with more accuracy, that of 92%. Supervised leaning method has been set up that consisted of three parts: “Model training, validation and classification”. The first step in this process was training of model. Model was trained with the collected features and was labelled as two classes of users (i.e., humans and bots). A dataset provided by Cresci and collaborators, that contained over 5000 positive (bots) and negative (humans) twitter users, was used for training. Using all the labelled training data, Logistic Regression model was retrained with ten simple metadata and other activity features that was collected as a part of dataset.

Finally, the last step consisted of classification where the model that was trained was used to classify all two million users in the Twitter French election dataset. From the comprehensive analysis of the result, it has been found that the users who were engaged in MacronLeaks campaign were mostly foreign belonging twitter community then potential French voters. Another interesting finding of this work has uncovered that the accounts that was used in MacronLeaks disinformation campaign was used earlier to support then presidential candidate Trump before 2016 US election

John P. Dickerson proposed SentiBot, an architecture that used to automatically identify bots on Twitter. Using sentiment analysis techniques different features can be collected and can be further used to analyses the sentiment on those data on having different basis such as per-user basis, different topics such neighbourhood-aware semantic metric, per- topic basic, syntactic tweet metric which includes number of hashtags, number of repeated tweets and graph theoretic. [10] The features here are defined as a set of contextual variables associated with each user in a dataset. Dataset that has been used in this work is from “Indian Election Dataset”. Its dataset consisted of real-world data that was collected from 15th July 2013 to 24th March 2014 that consisted of over 550,000 twitter account’s information and 7.7 million tweets on several topics that was discussed over time of the then Indian Election 2014.

Working of SentiBot: It used twitter API to identify users and assign it to set Uo who had tweeted or shared their opinion on at least one of the topics that was being discussed during the time frame mentioned above. A Network database was formed by examining users in Uo, by examining the follower and friends’ pattern and profile of the user is the third feature collected by SetiBot where for each user in Uo profile is extracted from their twitter profile, network position, their tweets and various other features. Using all these components, a database is built.

For the training and testing of a model, a random selection of users was made which consisted of set of 897 users from the IEDS. An ensemble classifier that consisted of six different classification techniques (support vector machines, Gaussian naive Bayes, AdaBoost, gradient boosting, random forests, and extremely randomized trees) was used to categorize the user U as a bot or not bot. Total 226,434 accounts were computed, that accounts the full IDES dataset, 90% of accuracy was achieved in determining the user as bot or not bot. On independent features, the in-depth comparison has been done on bots against humans, which concludes the results as following:

- The sentiment flip-flop score for 92.5% bots and 26.5% of humans lies in the range of 0 to 0.2 (human percentage is less as compared to bots).
- Human express with strong sentiments opinions as in comparison to bots.
- One of the reliable parameters for differentiating bots and humans is frequency of tweets.
- The dissonance rank depicts that tendency of disagreeing is more often found in human than bots with entire twitter population.

Russia has been accused of using Malicious accounts for endorsing democratic discourse on social and political issues and use of bots in U.S. election campaign held in year 2016. [11] In this work a dataset was created using “Crimson Hexagon”- a social media analytic platform that provides paid DataStream access” to collect tweets. 13 million tweets and retweets related to election campaign, that was shared on twitter on the year 2016 by over a million unique users had been collected and some non-trolls or the users that did not retweet any trolls was also considered as a part of dataset as it helped in better understanding of troll’s behaviour, by having a balanced feature dataset. After

analysing and having a sufficiently large dataset containing desired metadata for bot detection, Botometer – an openly accessible solution has been used for bot detection.

Botometer provides machine learning framework that extracts and analysis over one thousand features such as user, friends, network, temporal, content and sentiment. For training the model, 1000s of such instances of social bots were used that yielded with accuracy of above 95 % of accuracy. From the study it has been analysed how political ideology was affected with engagement of content that has been posted by Russian trolls. The label propagation method has been used to label users as “Liberal or Conservative”. Bot analysers is used to analyses over 115,396 spreader account out of random sample of 200,00 account spreader.

After analysing the results, it was discovered that the conservative trolls were double as compared to liberal trolls, out of total number of trolls that includes original tweets. Another observation was that less accounts were responsible for spreading original tweets as compared to tweets and retweets of trolls. Third observation was of social bots, using Botometer 200,000 spreaders that was selected randomly, they could obtain bot score for 115,396 spreader account out of which liberal and conservative accounts were identified and bot score for each account were calculated which is depicted in Table 1.

Table 1: Bot Score of Spreader Accounts

Liberal Accounts	Conservative Accounts	Accounts Having Bot Score Above 0.5	Remarks
68,057	-	3528	5% of total
-	45,807	4896	11% of total

From all these observations and results we can say that the number of conservative spreaders is less in number as compared to liberal ones but the total number bots among conservative accounts were more in number who were responsible for writing more tweets as compared to liberal ones. [12] have analysed social media campaign during 2016 US presidential election to disclose the existence of social bots. A list of hashtags and keyword related to campaign has been manually crafted, that contained 23 terms, out of which 5 terms were of Republic party nominee Donald Trump and 4 terms for Democratic party nominee Hillary Clinton and the rest from 4 presidential debate. The dataset was created by querying the Twitter search API between 16th September and 21st October 2016. After post-processing and cleaning process, a dataset that contained 20.7 million tweets that was posted by 2.8 million unique users was set.

Botometer– an openly accessible solution has been used, which provides machine learning framework for bot detection or to determine the presence of bot in social media. Therefore, to label the account as bot or human, 50% threshold is used as it has proven to be effective. A tweeter account is labelled as bot, if it achieves a bot score of 0.5 or above 0.5 or else human is less than 0.5. Approximately 50 thousand accounts were tested, 2% of the entire population that were accountable for producing over 1.26 million tweets, about 60% of total discussion. SentiStrength was used to determine the sentiment score of each tweet in the dataset, it is a toolkit that estimates the strength of positive and negative sentiment of short text, ideally suited for social media. It reports two sentiment strength -1(not negative) to -5 (extremely negative) and 1(not positive) to 5 (extremely positive). In the related works each tweet  $t$  is assigned positive  $p+(t)$  or negative  $p-(t)$  as polarity score, that ranges from 1 to 5. Sentiment analysis based on SentiStrength shows that humans replied significantly more to other humans than to bots and on the contrary bots received replies from bots significantly more than from humans. Another interesting result depicts that bots used to support Trump in campaign generated no negative tweets against him, instead they were used to post tweets supporting them and it produced nearly two third of positive tweets in total.

David M. Beskow et. al proposed a model for detecting bot at a very granular level data. Detection, characterization and modelling the behaviour of bots, the network structure of bots and the bot masters (creator of bots) were the primary focus of the work. [13]A training data was constructed that consisted of non-random twitter screen name, that was of total:200,000 and 200,000 randomly generated 15-digit screen name. Using heuristic and tradition Machine learning models, the string was labelled as random or non-random. The work has used Scikit-learn package to explore and build the Machine Learning model for random string collected. Among Naïve Bayes, Logistic Regression, and

Support Vector Machines (SVM) with 10-fold cross-validation, it has been observed that SVM and logistic Regression are statistically same result.

For the model deployment, a dataset was constructed that consisted of a random sample from Twitter streaming API. The time frame was set for the collection of data, and it was from 23rd December 2017 to 20th June 2018. The dataset collected approximately 433 million tweets in total, that produced 7.8 million tweets from 1.7 million unique accounts. 7.8 million tweets related with 7 different languages. The analysed result depicts among the total tweets collected only 840 tweets coordinated with the locations. The location that were correlated to the languages of U.S, Japan, Middle East, Russia, Thailand. Another major observation was that, random string accounts are younger, less active, less popular then average twitter account. Median age for average twitter account is calculated as 1248 while for random string bots it is 244 days. After constant pressure from media and politics on social media companies, it was observed that in mid-December 2017 and 22nd August 2018, 247022 twitter accounts were suspended by twitter.

#### 4. ANALYSIS AND FUTURE ASPECTS

The result of case study that we have conducted in this paper shows some serious impacts of social bots. Rumours, misinformation, and further amplification of such are major fuel to any social media in these days, to manipulate the original content or message. It has been observed such rumours pollute social media's content and play a very vital role to spark public emotions.

From this study it can be observed how social bots are used to amplify the work of political parties and their agenda and get highlighted in public, or to sabotage the image of opposition party as well. (as in MacronLeaks Misinformation campaign). During political discussion social bots are used to post content mostly manipulated by bots, put comments on any sensitive matter to evoke high arousal of emotions such as anger, happiness, fear, or extremist viewpoint. (as seen in 2016 US presidential election). Our findings from second part of Literature review exhibit the use of various possible techniques to identify the accounts that demonstrates bot behaviour, so that it can be classified as bot account for further action. Machine learning is gaining huge attention in such field, to find the behavioural pattern of such bot account and differentiate them with human account, as a result there can be distinct line of differentiation between Bot accounts and human accounts.

Social media content is voluminous, it entails data of every field and applicable for various purposes, such as political data, commercial data, scientific data and so on, which is very difficult to differentiate and to understand about the exact subject there is a need to understand these voluminous data in some automated approach. Therefore, to understand the subject that is being talked about in social media and to understand the opinion of public Sentiment analysis, a field of machine learning has been used, which analyse text data and classify opinions as negative, positive or neutral. The study demonstrates that with the use of Machine learning techniques we can to some extent study the behaviour of social bots, and do the classification using various classification techniques. The important observation we have found out that in order classify the accounts as bots or humans, it is very important to set out the parameter on which basics we can differentiate bot and humans. One of the reliable parameters for differentiating bots and humans is frequency of tweets, and another is the age of the account, (random string accounts are younger, less active, less popular then average twitter account). The overall summary of the techniques and observation that has been derived for the literature review of the technical study has been depicted in Table 2.

**Table 2.** Summary of Political Sentiment Analysis: Techniques

Author	Techniques	Observation
Emilio Ferrara [6]	<ul style="list-style-type: none"> <li>Use of SVM, Logistic Regression</li> <li>Ensemble Method (Random Forest, AdaBoost)</li> </ul>	<ul style="list-style-type: none"> <li>Accuracy :92%</li> <li>Accounts used in MacronLeaks campaign was used earlier to support 2016 US election</li> </ul>

John P. Dickerson [10]	<ul style="list-style-type: none"> <li>SentiBot: for sentiment analysis</li> </ul>	<ul style="list-style-type: none"> <li>Accuracy :90%</li> <li>The sentiment flip flop score for 92.5% bots and 26.5% of humans lies in the range of 0 to 0.2 (human percentage is less as compared to bots).</li> </ul>
Adam Badawy et al. [11]	<ul style="list-style-type: none"> <li>Botometer-Machine Learning Framework</li> </ul>	<ul style="list-style-type: none"> <li>Accuracy :95%</li> <li>The total number bots among conservative accounts were more in number who were responsible for writing more tweets as compared to liberal ones.</li> </ul>
Emilio Ferrara [12]	<ul style="list-style-type: none"> <li>Botometer-Machine Learning Framework</li> <li>SentiStrength: Toolkit</li> </ul>	<ul style="list-style-type: none"> <li>Bots used to support Trump in campaign generated no negative tweets against him.</li> <li>Bots used to support Trump produced nearly two third of positive tweets in total.</li> </ul>
David M. Beskow et al. [13]	<ul style="list-style-type: none"> <li>Naïve Bayes, Logistic Regression, and Support Vector Machines (SVM)</li> </ul>	<ul style="list-style-type: none"> <li>SVM and logistic Regression are statistically same result.</li> <li>Median age for random string bots is 244 days</li> <li>Median age for average twitter account is as 1248 days</li> </ul>

#### 4.1 FACTORS TO BE CONSIDERED FOR FUTURE WORK

Majority of the classification techniques suffer false positive rate. This makes an important aspect to be noted during the feature

selection process, (as some human accounts sometimes exhibit bot behaviour).

- Increasingly popular social media data presents new challenges to feature selection. Social media data is prone to spelling mistakes, sarcasm and internet slang which adversely affect the quality of classification. Therefore, it is very important to clean the data properly in order to accurately label accounts as bots or humans.
- Parameters that clearly distinguish the behaviour of Bot accounts and human accounts.
- Selection of techniques that suffers less false positive rate and higher detection accuracy rate.

In addition to the factors that have stated above, a research methodology has been derived given in Fig. 1, that could be adopted in future for classification of Social Bots and humans, as well as for determining the sentiments of the data.

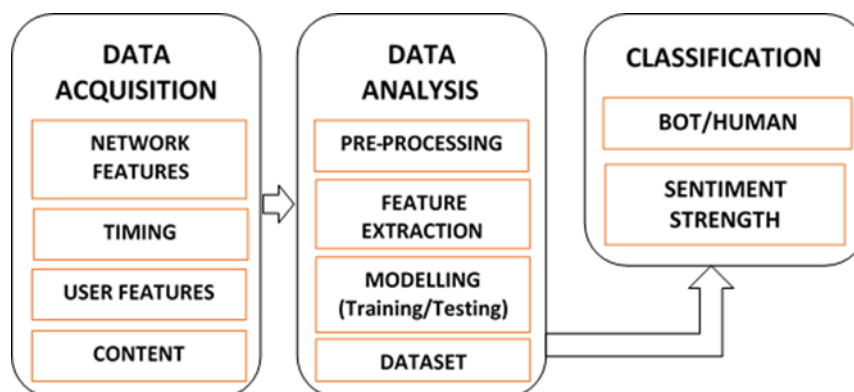


Fig. 1 Proposed Methodology for Social Bot Detection and Sentiment Analysis



## 7. CONCLUSION

Social media contains a dynamic and rich data content. It has proven the fact that it is more powerful and successful in influencing the people then compared to any other medium. Large number of political parties are active in social media and some do even have their own websites, which are often used to promote their ideas and propaganda to the public. Political parties are also employing cyber warrior to help campaign, as in the study the researcher found the use of bots that were massively amplified to swing the election and to manipulate political discourse public opinion in twitter in 2016 US presidential election. The detection and characterization of these kinds of bot army used in social media is quite challenging as predicting the political orientation with content- based methods is not easy and the sarcasm and slangs used in text can cause for poor performance of context-based methods. The dynamic nature of data, where text communication is very common and beginning, as social media flaunts heterogeneous data type, which makes very difficult to analyses the content of social media. [14] Barack Obama's implausible fake video is a clear indication that it is very much probable to make audio and video that can mimic the humans and distribution can be done by bot amplifier. Humans are already having trouble in recognizing simple text bots that behaves like a real user; and now the advancement made in technologies in area of Artificial intelligence are giving so much power and platform to humans to enhance such activities. So, the future study and research lies in even more advance methods and techniques to stop the social bots and there is a need for public also to understand and work towards generating practical responses to address such problem.

## REFERENCES

- [1] 10 Social Media Statistics You Need to Know in 2019 [Infographic], Maryam Mohsin ,7th March 2019 URL: <https://www.oberlo.com/blog/social-media-marketing-statistics> [Accessed on 5th December 2019]
- [2] Ramandeep Kaur, Impact of Social Media in India, 28th September 2015 URL: <https://www.mapsofindia.com/my-india/society/role-of-social-media-in-india>. [Accessed on 5th December 2019]
- [3] Mahapatra, Sangeeta; Plagemann, Johannes, Polarisation and Politicisation: The Social Media Strategies of Indian Political Parties, GIGA Focus Asia Number 3, March 2019, ISSN 1862-359X
- [4] India's election campaign is being fought in voters' pockets, 11th April 2019 URL: <https://www.economist.com/asia/2019/04/11/indias-election-campaign-is-being-fought-in-voters-pockets> [Accessed on 29th November 2019]
- [5] Social bots distort the 2016 U.S. Presidential Election online discussion, Alessandro Bessi and Emilio Ferrara, Volume 21, Number 11 - 7 November 2016, URL <https://firstmonday.org/article/view/7090/5653> [Accessed on 25th November 2019]
- [6] Emilio Ferrara, Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election, DOI: 10.5210/fm.v22i8.8005 June 2017.
- [7] Exclusive: Russia used Facebook to try to spy on Macron campaign – source, 27th July 2017, Joseph Menn, URL: <https://www.reuters.com/article/us-cyber-france-facebook-spies-exclusive/exclusive-russia-used-facebook-to-try-to-spy-on-macron-campaign-sources-idUSKBN1AC0EI> [Accessed on 10th November 2019]
- [8] Successfully Countering Russian Electoral Interference, Heather A. Conely 21st July 2018 URL: <https://www.csis.org/analysis/successfully-countering-russian-electoral-interference> Accessed on 12th November 2019
- [9] Deep Learning for Sentiment Analysis, Hussein Sajid July 2019, URL: <https://towardsdatascience.com/deep-learning-for-sentiment-analysis-7da8006bf6c1> Accessed on 8th November 2019

- [10] John P. Dickerson, Using Sentiment to Detect Bots on Twitter: Are Humans More Opinionated than Bots? 2014 IEEE/ACM
- [11] Adam Badawy, Aseel Addawood, Kristina Lerman, Emilio Ferrara<sup>1</sup> Characterizing the 2016 Russian IRA influence campaign, Springer, 27 June 2019
- [12] Emilio Ferrara, Bots, elections, and social media: A Brief overview, 3rd October 2019 USC Information Sciences Institute, arXiv:1910.01720 [cs.SI]
- [13] David M. Beskow , Kathleen M. Carley. It's all in a name: detecting and labeling bots by their name, doi.org/10.1007/s10588-018-09290-1, Springer, 18th December 2018
- [14] New AI research makes it easier to create fake footage of someone speaking James Vincent, 12th July URL:<https://www.theverge.com/2017/7/12/15957844/ai-fake-video-audio-speech-obama> [Accessed on 28th October 2019]