International Journal of Innovative Research in Science and Engineering
Vol. No.3, Issue 03, March 2017
www.ijirse.com
ISSN (O) 2454-9665
ISSN (P) 2454-0663

COMPARING CONVENTIONAL DATA MINING ALGORITHMS WITH HADOOP BASED MAP-REDUCE ALGORITHM CONSIDERING ELECTIONS PERSPECTIVE

Amandeep Kaur[1], Dr. Gagandeep Jagdev[2]

[1]*Research Scholar, Ph.D(Comp. Appl.), Guru Kashi University, Talwandi Sabo (PB)*

[2]*Dept. of Computer Science, Punjabi University Guru Kashi College, Damdama Sahib (PB)*

ABSTRACT

In simple term, big data can be defined as any data which challenges the currently existing techniques for handling it. Big data presents a grand challenge for database and data analytics research. The central theme of the research paper is based on the concept of mining huge data concerned with elections in Indian sub-continent. It highlights the use of big data concepts in such a manner that it would help the political parties in canvassing and targeting voters and at the same time it would help electorates to appoint an efficient representative from their respective constituency. This paper includes discussion about the beginning of the era of big data in the Indian politics which has been utilized to maximum potential in 16th Lok Sabha national elections by BJP (Bhartiya Janta Party) to gain total majority. Gone are the days when politicians fought elections on the basis of religion, caste, and ethnicity. With the advent of big data, elections are now been fought on the basis of statistics, data, and facts. The paper also details the working of Hadoop framework responsible for handling Big Data. In addition to this, we would also discuss the logical challenges faced by big data in handling election process in India.

Keywords: Big Data, Data Mining, Electoral, Hadoop framework, Map Reduce,

I. INTRODUCTION

Internet is the major source which has resulted in the tsunami of data in the past few years. Big data is too big, it moves too fast, and doesn't fit the structures of our existing database architectures. It is like an ocean of data in which we people swim in every day with an effort to come on the surface, but every day the level of data increases tremendously. Gone are the days when memory was used to be measured in Gigabytes or Terabytes or Petabytes, today it is measured in exabytes, zettabytes or yottabytes. With Big Data solutions, organizations can dive into all data and gain valuable insights that were previously unimaginable. The term "big data" can be pretty nebulous, in the same way that the term "cloud" covers diverse technologies. Utilizing big data requires transforming information infrastructure into a more flexible, distributed, and open environment.

Big data promises deeper insights that data scientists are highly involved in exploring this data in such a manner that organizations are benefited to its best with total customer satisfaction. Big data analytics is one of the great new frontiers of IT. Emerging technologies such as the Hadoop framework and MapReduce offer new and exciting ways to process and transform big data—defined as complex, unstructured, or large amounts of data—into meaningful insights, but also require IT to deploy infrastructure differently to support the distributed processing requirements and real-time demands of big data analytics [3, 4].

II. BIG DATA AND MATTERS OF CONCERN

Data Volume – It refers to the enormous amount of data that is been created each second, each minute and each hour of the day from various sources like social media, web, sensors, different sectors like banking, health sector, retail sector and many more. 571 websites are created in a single minute. Total of 625000 GB of data is transferred from one end to another in single internet minute, may be terms of mails, pictures, posts etc.

Data Velocity – Data is being created at such high velocity that companies are finding it difficult to cope up with such high speed. They have to establish their infrastructure in such a manner that it is capable of handling such generated data Social media, E-Commerce has rapidly increased the speed and richness of data used for different business transactions.

Data Variety - All the data being generated is totally diverse consisting of raw, structured, semi structured and even unstructured data which is difficult to be handled by the existing traditional analytic systems. Mismatched data formats and data structures represent significant challenges that can lead to analytic collapse.

Data Value –There is a huge gap in between the business leaders and the IT professionals. The main concern of business leaders is to just add value to their business and to maximize their profit. On the other hand, IT leaders deal with technicalities of the storage and processing.

Data Complexity – The biggest complexity faced while running big data using relational databases is that they require parallel software running on hundreds of servers and data scientists have to match and transform data across systems coming from various sources.

Data Veracity - Veracity refers to the preciseness of data or how much faith one can have on data. The data on internet is not always accurate or precise. For example, if some male pretends himself as a female on his facebook profile, there is no authenticity check in such cases. Similarly, twitter makes use of abbreviations and hash tags, but big data enables us to work with even this type of imprecise data [1, 6].

III. COMPONENTS OF DATA MINING

With an enormous amount of data stored in databases and data warehouses, it is increasingly important to develop powerful tools for analysis of such data and mining interesting knowledge from it. Data mining is a process of inferring knowledge from such huge data. Data Mining has three major components

- Clustering or Classification
- Association Rules
- Sequence Analysis.

ISSN (O) 2454-9665
ISSN (P) 2454-0663

By simple definition, in classification/clustering one analyzes a set of data and generate a set of grouping rules which can be used to classify future data. For example, one may classify diseases and provide the symptoms which describe each class or subclass. This has much in common with traditional work in statistics and machine learning. However, there are important new issues which arise because of the sheer size of the data. One of the important problem in data mining is the Classification-rule learning which involves finding rules that partition given data into predefined classes. In the data mining domain where millions of records and a large number of attributes are involved, the execution time of existing algorithms can become prohibitive, particularly in interactive applications.

An association rule is a rule which implies certain association relationships among a set of objects in a database. In this process, one discovers a set of association rules at multiple levels of abstraction from the relevant set(s) of data in a database. For example, one may discover a set of symptoms often occurring together with certain kinds of diseases and further study the reasons behind them. Since finding interesting association rules in databases may disclose some useful patterns for decision support, selective marketing, financial forecast, medical diagnosis, and many other applications, it has attracted a lot of attention in recent data mining research. Mining association rules may require iterative scanning of large transaction or relational databases which is quite costly in processing. Therefore, efficient mining of association rules in transaction and/or relational databases has been studied substantially.

In *sequential Analysis*, one seeks to discover patterns that occur in sequence. This deals with data that appear in separate transactions (as opposed to data that appear in the same transaction in the case of association). For e.g.: If a shopper buys item A in the first week of the month, then he buys item B in the second week etc.

IV. DIFFERENT ALGORITHMS CONCERNED WITH DATA MINING

Conventional Data Mining Algorithm

Some of the popular conventional algorithms involved in data mining process are mentioned as under.

- C4.5
- k-means
- Support vector machines
- Apriori

A. C4.5 - C4.5 constructs a classifier in the form of a decision tree. In order to do this, C4.5 is given a set of data representing things that are already classified. A classifier is a tool in data mining that takes a bunch of data representing things we want to classify and attempts to predict which class the new data belongs to. Arguably, the best-selling point of decision trees is their ease of interpretation and explanation. They are also quite fast, quite popular and the output is human readable.

B. k-means - k-means creates *k* groups from a set of objects so that the members of a group are more similar. It's a popular cluster analysis technique for exploring a dataset. Cluster analysis is a family of algorithms designed to form groups such that the group members are more similar versus non-group members. Clusters and groups are synonymous in the world of cluster analysis.

ISSN (O) 2454-9665
ISSN (P) 2454-0663

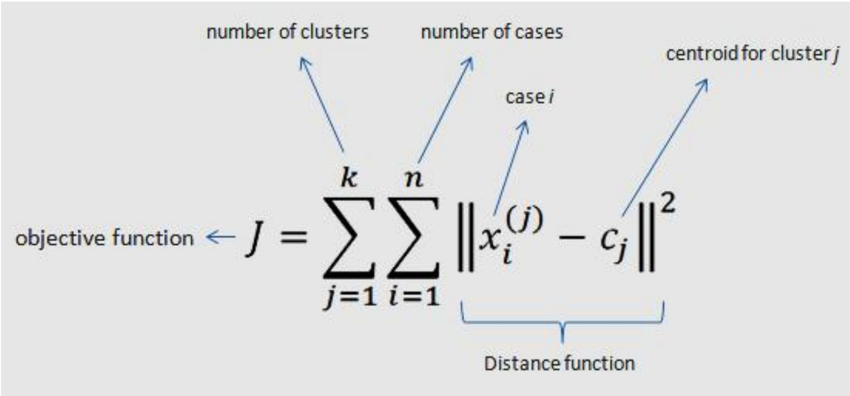

Fig. 1 Formula for K-Means clustering

Algorithm for k-means

1. Clusters the data into *k* groups where *k* is predefined.
2. Select *k* points at random as cluster centers.
3. Assign objects to their closest cluster center according to the *Euclidean distance* function.
4. Calculate the centroid or mean of all objects in each cluster.
5. Repeat steps 2, 3 and 4 until the same points are assigned to each cluster in consecutive rounds.

C. Support Vector Machine (SVM) - Support vector machine (SVM) learns a hyperplane to classify data into 2 classes. At a high-level, SVM performs a similar task like C4.5 except SVM doesn't use decision trees at all. A hyperplane is a function like the equation for a line, y = mx + b. In fact, for a simple classification task with just 2 features, the hyperplane can be a line. SVM can perform a trick to project your data into higher dimensions. Once projected into higher dimension, SVM figures out the best hyperplane which separates your data into the 2 classes. For example - Consider a bunch of red and blue balls on a table. If the balls aren't too mixed together, you could take a stick and without moving the balls, separate them with the stick. When a new ball is added on the table, by knowing which side of the stick the ball is on, you can predict its color. The balls represent data points, and the red and blue color represent 2 classes. The stick represents the simplest hyperplane which is a line.

D. K - Nearest Neighbor (kNN) - KNN can be used for both classification and regression predictive problems. However, it is more widely used in classification problems in the industry. Consider the example mentioned as under to place KNN in the scale. Fig. 2 shows the spread of red circles (RC) and green squares (GS):

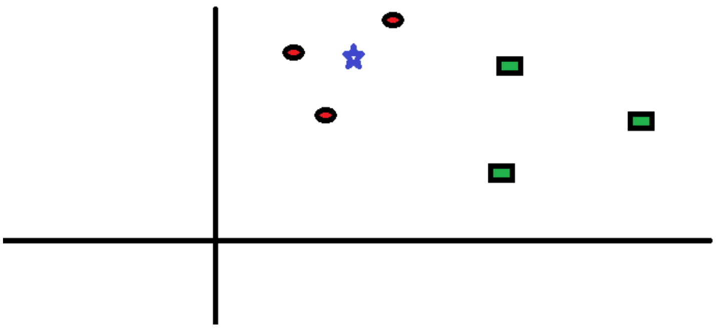

Fig. 2 Figure depicts the spread of red circles and green squares

We intend to find out the class of the blue star (BS). BS can either be RC or GS and nothing else. The “K” is KNN algorithm is the nearest neighbors we wish to take vote from. Let’s say K = 3. Hence, we will now make a circle with BS as center just as big as to enclose only three data points on the plane. Refer to Fig. 3 below.

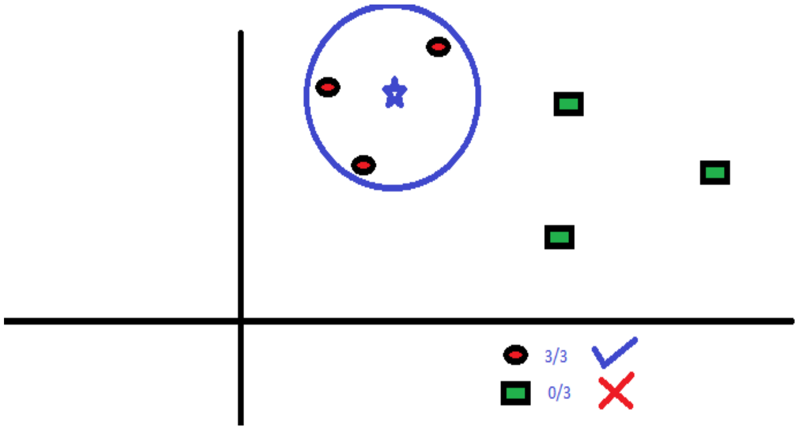

Fig. 3 Figure depicts three similar data points enclosed on the plane.

The three closest points to BS is all RC. Hence, with good confidence level we can say that the BS should belong to the class RC. Here, the choice became very obvious as all three votes from the closest neighbor went to RC. The choice of the parameter K is very crucial in this algorithm. Next, we will understand what are the factors to be considered to conclude the best K.

E. Apriori Algorithm - It is a classic algorithm used in data mining for learning association rules. It is nowhere as complex as it sounds, on the contrary it is very simple. Apriori is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database. Apriori uses a "bottom up" approach, where frequent subsets are extended one item at a time and groups of candidates are tested against the data. The algorithm terminates when no further successful extensions are found.

Map-Reduce Algorithm for Big Data

Map-Reduce - Hadoop is one such java framework which is capable of handling large data sets in a distributed computing environment. Apache Software Foundation is the authority sponsoring Hadoop. Applications run on systems constructed of thousands of nodes involved in processing thousands of Terabytes of data. Because of distributed nature, Hadoop facilitates very fast data transfer and system continues to operate even if any node failure occurs. This notion minimizes the risk of devastating system failure even when multiple nodes become non-operational. The creator of this technology was Google. It was developed by them during their early days to index all valuable textural and structured information. The primary motive behind all this was to provide meaningful results to the users. Hadoop finds its application in several sectors which comprises of retail, sports, medical science, business, education and of course now in elections.

Map-Reduce algorithm works behind Hadoop and can be written in any language [5, 12].

The flowchart depicting the working of Map-Reduce technology is shown in Fig. 4.

International Journal of Innovative Research in Science and Engineering
Vol. No.3, Issue 03, March 2017
www.ijirse.com
ISSN (O) 2454-9665
ISSN (P) 2454-0663

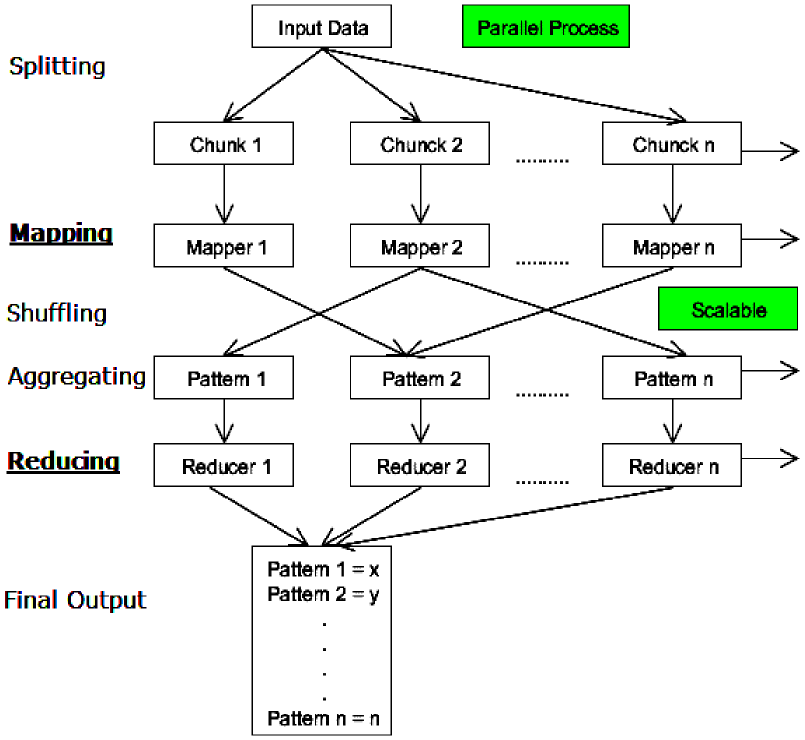

Fig. 4. Flowchart showing working of Map-Reduce technology

Algorithm for Map-Reduce

- The incoming data can be alienated into n number of modules which depends upon the amount of input data and processing power of the individual unit.
- All these fragmented modules are then passed over to mapper function where these modules undergo simultaneous parallel processing.
- Thereafter, shuffling is conducted in order to gather similar looking patterns.
- Finally, reducer function is called which is responsible for getting the ultimate output in a reduced form.
- Moreover, this technique is scalable and depending upon increase in the data to be processed, the processing units can be further extended.

V. MAP-REDUCE VS. APRIORI ALGORITHM

The research work is based on mining a huge database constructed and to use best available platform or algorithm or technology for this purpose. Among many prominent data mining algorithms already existing, Apriori algorithm is regarded as the best whenever there's a need to perform mapping or mining database in combinations. So, the target is to compare Apriori algorithm with the Map-Reduce algorithm related to Hadoop platform. For this, a small excel file is taken as input comprising of few numbers. The snapshot of the input file is shown in Fig. 5.

ISSN (O) 2454-9665
ISSN (P) 2454-0663

A1	A2	A3	A4	A5	Sum
1	5	2	0	0	8
2	3	0	1	0	6
3	4	0	0	0	7
2	1	3	0	0	6
1	2	3	0	0	6

Fig. 5 Input data file

A source code is constructed in C++ for implementing working of Apriori algorithm and the above-mentioned data in Fig. 1 is given as input.

On executing the source code of Apriori algorithm, the following output was obtained.

1 2 3 3

The result shows that 1, 2 and 3 appears 3 times together in the input file and it took 17.967033 seconds. The snapshots shown in Fig. 6 shows time consumed as the output obtained on implementing Apriori algorithm.

Fig. 6 Output obtained after executing Apriori algorithm source code

When the same input was given to Hadoop Framework, the result was obtained in only 1.2 Seconds. The comparative graph of Apriori and Hadoop is shown below in Fig. 7.

International Journal of Innovative Research in Science and Engineering
Vol. No.3, Issue 03, March 2017
www.ijirse.com
ISSN (O) 2454-9665
ISSN (P) 2454-0663

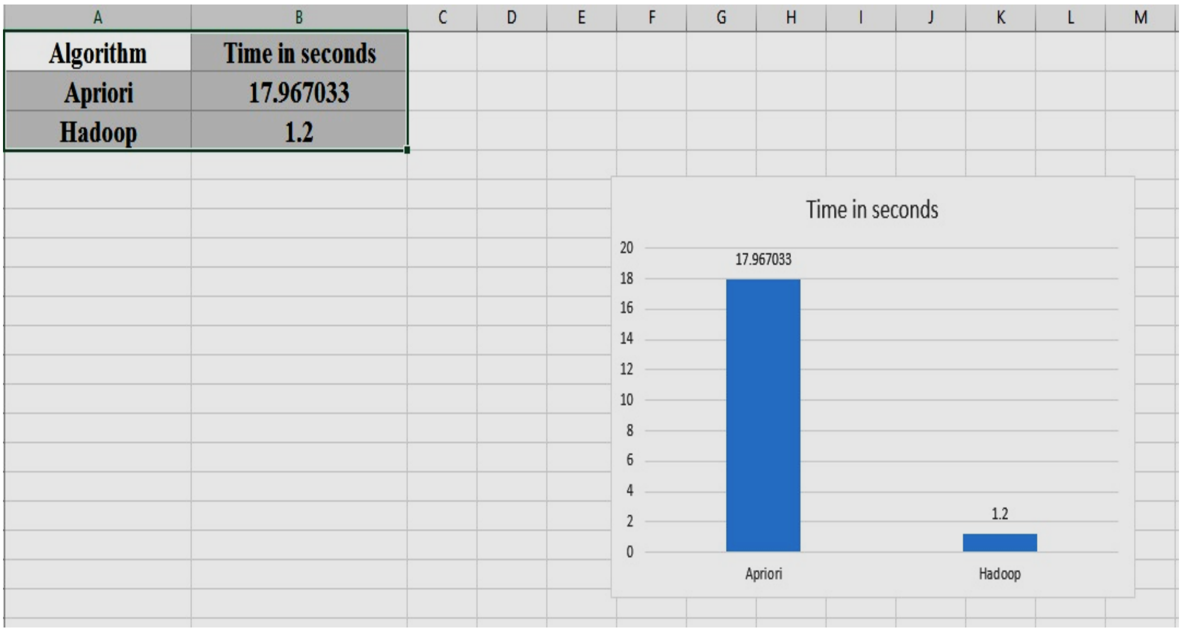

Fig. 7. Comparison of Apriori and Map-Reduce algorithm in terms of time in seconds.

This result shows that Map-Reduce algorithm is much more speedy and efficient in mining as compared to Apriori algorithm. So, because of this I opted for Apache Hadoop framework.

VI. FIVE PHASES OF BIG DATA

Big data processing involves five different phases [2, 6, 7].

Data Acquisition and Recording – Big data definitely have some source of origin. It is not created from a vacuum. Different scientific experiments being carried out in the world today produces petabytes of data per day. Much of this data is of no use and has to be filtered out. We need research in the science of data reduction that can intelligently process this raw data to a size that its users can handle while not missing the needle in the haystack. The second challenge encountered is related to automatically generating right metadata to illustrate what data is recorded, how it is recorded and measured.

Information Extraction and Cleaning – It is mention able here that information collected is not in an analysis ready format. The data in this format cannot be effectively analyzed. An information extraction process should be applied to such data to pull out the required information from the sources under consideration and present it in a structured format suitable for analysis. This is really a big challenge. This data may include images and videos and such extraction is highly application dependent.

Data Integration, Aggregation, and Representation – It is not enough to merely collect record and throw the data into a repository. If we have large data sets in repository, then it will be almost impossible for the user to find the desired data when required. But with sufficient metadata, there is some hope but still challenges persist due to differences in experimental details and in data record structure. Data challenging is much more than simply locating, identifying, understanding and citing data. All this process needs to occur in a complete automated manner for an effective large scale analysis. Suitable database design is most important. There are many different ways in which data can be stored. Certain designs will be better than others for certain purposes

ISSN (O) 2454-9665
ISSN (P) 2454-0663

and possibly may carry drawbacks for other purposes. Therefore, it can be concluded that database design is an art and needs to be carefully executed by trained professionals.

Query Processing, Data Modeling, and Analysis Methods for querying and mining – There is no doubt in the fact that big data is diverse, imprecise and unstructured. Even then big data is of much value as compared to small individual observations as general statistics obtained from large sample are more precise. When it comes to mining, it requires clean and efficiently accessible data. Provision should be there for declarative query and mining interfaces. Efficient mining algorithms and computing environments is another important requirement.

Interpretation – The analysis of big data remains of no value if users are not able to understand the analysis concept. Decision maker is provided with the result of analysis and is expected to interpret these results. This interpretation requires efforts. It involves deeply examining all the assumptions made and retracing the analysis. There are several sources of errors like system may carry bugs and conclusions may be based on error prone data. No responsible user will yield authority to computer system for all this. Instead one will try to understand and verify the results produced by computer system. All this should be made easy by computer system and this is a big challenge with big data due to its complexity.

VII. BIG DATA CONTRIBUTION IN ELECTIONS

Modern campaigns develop databases of detailed information about citizens to inform electoral strategy and to guide tactical efforts. Despite sensational reports about the value of individual consumer data, the most valuable information campaigns acquire comes from the behaviors and direct responses provided by citizens themselves. Campaign data analysts develop models using this information to produce individual-level predictions about citizens' likelihoods of performing certain political behaviors, of supporting candidates and issues, and of changing their support conditional on being targeted with specific campaign interventions. The use of these predictive scores has increased dramatically since 2004, and their use could yield sizable gains to campaigns that harness them. At the same time, their widespread use effectively creates a coordination game with incomplete information between allied organizations. As such, organizations would benefit from partitioning the electorate to not duplicate efforts, but legal and political constraints preclude that possibility.

Elections in India have always comprised issues based on caste, religion, sentiments, traditional wisdom, opinion polls and rallies. But 2014 Lok Sabha elections witnessed the use of technology to its very best by political parties. All this idea was actually borrowed by the way Barack Obama contested his election sin America and raise to power in 2008 and 2012.

In an unprecedented digital effort in India, Google and other social platforms executed an aggressive digital informational campaign in an attempt to engage digitally literate citizens to engage in the electoral process. Google India launched Elections Hub where citizens were able to research their political candidates, political parties, election platforms, and voting information in their districts. They launched a Counting Results site on Election Day that offered a glimpse into the live counting of votes, as well. The search data trends revealed that Narendra Modi consistently topped the search trends when compared to other candidates.

For conducting 2014 Lok Sabha elections, 543 Parliamentary constituencies and 4120 assembly constituencies were set up. All over India total of 9 lakhs 30 thousand polling booths were set up for conducting fair elections. Voter rolls were prepared in 12 different languages and total of 9 lakh pdf files which amounted to 2.5 crore pages were deciphered. The real challenge was extraction of voter info from these 2.5 crore PDF pages and transliteration of the same into English to fuse with other sources. Technology was a big hurdle.

Campaigns also use data to construct predictive models to make targeting campaign communications more efficient and to support broader campaign strategies. These predictive models result in three categories of "predictive scores" for each citizen in the voter database: behavior scores, support scores, and responsiveness scores.

Behavior scores use past behavior and demographic information to calculate explicit probabilities that citizens will engage in particular forms of political activity. The primary outcomes campaigns are concerned with include voter turnout and donations, but other outcomes such as volunteering and rally attendance are also of interest.

Support scores predict the political preferences of citizens. In the ideal world of campaign advisers, campaigns would contact all citizens and ask them about their candidate and issue preferences. However, in the real world of budget constraints, campaigns contact a subset of citizens and use their responses as data to develop models that predict the preferences of the rest of the citizens who are registered to vote. These support scores typically range from 0 – 100 and generally are interpreted to mean "if you sample 100 citizens with a score of X, X percent would prefer the candidate/issue". A support score of "0" means that no one in a sample of 100 citizens would support the candidate/issue, "100" means that everyone in the sample would support the candidate/issue, and "50" means that half of the sample would support the candidate/issue. Support scores only predict the preferences at the aggregate-level, not the individual-level. That is, people with support scores of 50 are not necessarily undecided or ambivalent about the candidate/issue and, in fact, may have strong preferences. But when citizens have support scores of 50, it means that it is difficult to predict their political preferences.

Responsiveness scores predict how citizens will respond to campaign outreach. While there are theoretical rationales as to who might be most responsive to blandishments to vote, and attempts at persuasion, in general, predicting which individuals will be most and least responsive to particular direct communications in a given electoral context is difficult. Campaigns can use fully randomized field experiments to measure the response to a campaign tactic. The results of these experiments can then be analyzed to detect and model heterogeneous treatment effects (i.e., predictive scores) that guide targeting decisions. Some of the results of these experiments can only be used to inform decisions in future elections (e.g., the results of most voter turnout experiments necessarily come after Election Day), but others can be conducted during the election cycle to improve efficiency in real time. For example, the lessons from experiments evaluating the efficacy of treatments aimed at increasing observable behaviors like donations and volunteering can be put to immediate use. Similarly, the persuasiveness of campaign communications can be gauged through randomized experiments that measure voter preferences through post-treatment polling of the treatment and control groups.

The last time election process was carried out in India, people saw the largest democracy in the world pull in almost 600 million of its residents into the voting process which ushered in the new government. These residents are diverse in every way possible which includes beliefs, sentiments, faith, language and motivations. The selections made by people are also based on a multitude of factors which may be direct or indirect. Among direct factors falls policies for that region, local polarizations, past records and indirect includes factors like geography, television penetration, mobile penetration, financial stability, climate, readership, media, etc. A large part of the voting population (about 30 percent) still feel confused about who to vote for and are guided by social, familial or political influencers. And also, there are people who just do not vote – either because they forget, or don't care, or live in regions which are really hard to reach. For a political party to win, they need a combination of new voters, influenced voters and those that form the core groups who always vote for them. Political parties construct different strategies to target each category. Imagine this, what if a political party is able to identify those people who are most likely to vote? Why would they then spend a fortune to do general canvassing with the people who are not likely to vote? Shouldn't they rather analyze social circles and find out key influencers and mavens and try to convince them to side with them. They could spend all their effort to combine the "want to's" with the "have to's". How can they do this?

An individual can be judged by examining his or her interest in reading the type of newspaper or magazine. If you are in the west and you subscribe to "Samna" or if you are in the north and subscribe to "Ajit", it does display your appetite towards a certain political party. Similarly, if you are a woman and subscribe to newspaper or magazines related to politics or national and international affairs, you may be more likely to vote than a person reading good Housekeeping magazines. Political parties could then analyze the views you express on social media as well as identify your social circles. If they find you influential enough, they could reach out to you over a variety of channels.

Further imagine what if the parties could find out what a voter really wants to listen to and then target them with specific key words and messages. We watch the television daily and if one smart party tells us something really nice about themselves during an advertisement break, we are bound to listen to it. Once we see the message we could tweet something good or bad about it and if the party finds public response against them, the party can change its messaging. Further to this, they could try finding the right spokesperson (like a movie star or a sportsperson, etc.) which may be selected from local movies, media and teams, who is best received by a target segment and dynamically keep changing them based on the kind of chatter it generates. Voters can also draw out comparative scorecards of various party candidates and how they have done in various constituencies based on news feed, policy decisions, project completions, etc., all automatically updated based on inputs from various channels. Such score cards may help people decide objectively on which candidate is actually better.

It has been analyzed that about 160 million of those who are not sure about who to vote could be targeted through mobile phones and about a 100 million through television. These people are waiting to hear the right message to make that choice of which party to vote for and may be the right message is hidden somewhere waiting to be uncovered. So, it can be concluded that big data analytics could act as a key to reveal the winning mantra which could get a political party their major win [8].

VIII. CONCLUSION

It can be concluded that big data is all set to play a major role in any national elections to be conducted in future. Political parties have to hire smart data scientists in order to decide about the strategies to be followed. Political parties have to concentrate on technological aspects much more than other factors. Appropriate utilization of Big Data guarantees big win of the political parties. The use of Big Data by BJP during Lok Sabha elections 2014 and later by Nitish Kumar in Bihar in 2015 have already proved the significance of using Big Data in elections. Since Big data is an emerging technology and is at its youth, so it needs to attract organizations and youth with diverse new skill sets. The skills should extend from technical to research, analytical, interpretive and creative ones.

REFERENCES

[1] Laney, Doug. 2012. “3D Data Management: Controlling Data Volume, Velocity and Variety.”
[2] Information Week. 2012. “Big Data Widens Analytic Talent Gap.” Information Week April.
[3] Heudecker, Nick. 2013. “Hype Cycle for Big Data.” Gartner G00252431
[4] Edala, Seshu. 2012. “Big Data Analytics: Not Just for Big Business Anymore.” Forbes.
[5] Dean, Jeffery, and Ghemawat Sanjay. 2004. “MapReduce: Simplified Data Processing on Large Clusters.” Google.
[6] Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013). Big Data: Issues and Challenges Moving Forward. International Conference on System Sciences (pp. 995-1004). Hawaii:IEEE Computer Soceity.
[7] Katal, A., Wazid, M., & Goudar, R. H. (2013). Big Data: Issues, Challenges, Tools and Good Practices. IEEE, 404-409.
[8] Gagandeep Jagdev et. al.,” Scrutinizing Elections Strategies by Political Parties via Mining Big Data for Ensuring Big Win in Indian Subcontinent”, 4th Edition of International Conference on Wireless Networks and Embedded Systems.
[9] http://hadoop.apache.org/docs/r1.2.1/ hdfs_design.html
[10] Dr. Gagandeep Jagdev et. al., “Big Data proposes an innovative concept for contesting elections in Indian subcontinent”, IJSTA Volume 1, Issue 3, pp. 23-28, 2015, ISSN No. 2454-1532.
[11] Dr. Gagandeep Jagdev et. al., “Excavating Big Data associated to Indian Elections Scenario via Apache Hadoop” in International Journal of Advanced Research in Computer Science (IJARCS), ISSN 0976 – 5697.
[12] Gagandeep Jagdev et. al., “BIG DATA ALLIANCE WITH MapR TECHNOLOGY TO GOVERN INDIAN ELECTON PANORAMA”, in Grenze International Journal by Grenze Scienctific Society, Grenze ID:02.IETET, 2016.5.10.
[13] Gagandeep Jagdev et. al., “ANALYZING AND SCRIPTING INDIAN ELECTION STATEGIES USING BIG DATA VIA APACHE HADOOP FRAMEWORK”, at IEEE International Conference WECON-2016 at Chitkara University, Rajpura (PB) on October 14-15, 2016 and has been accepted for publication in IEEE Explore.

About the Author

Dr. Gagandeep Jagdev (Ph.D, M.Phil, M.C.A.) is working in the capacity of System Analyst (Teaching) at Punjabi University Guru Kashi College, Damdama Sahib, Bathinda (PB). He has more than 100 research publications to his credit, published in international journals, international and national conferences. He is a member of editorial board of many International Journals. His total teaching experience is of more than 10 years and field of research includes Big Data, Biometrics, RFID, Sensors and VANETs.