Implementation of Big Data Analytics in Automotive Industry

Sukhpreet Singh[1] Dr. Vijay Bhardwaj[2]
[1]Research Scholar, Ph.D. (Comp. Appl.), Guru Kashi University, Talwandi Sabo, Punjab, India.
[2]Assistant Professor, UCCA, Guru Kashi University, Talwandi Sabo, Punjab, India.

Abstract

Market landscape has undergone dramatic change because of globalization, shifting marketing conditions, cost pressure, increased competition, and volatility. Transforming the operation of businesses has been possible because of the astonishing speed at which technology has witnessed the change. The automotive industry is on the edge of a revolution. The increased customer expectations, changing ownership, self-driving vehicles and much more have led to the transformation of automobiles, applications, and services from artificial intelligence, sensors, RFID to big data analysis. Large automobiles industries have been emphasizing the collection of data to gain insight into customer's expectations, preferences and budget alongside competitor's policies. Statistical methods can be applied to historical data which has been gathered from various authentic sources and can be used to identify the impact of fixed and variable marketing investments and support automakers to come up with a more effective, precise, and efficient approach to target customers. Proper analysis of supply chain data can disclose the weak links in the chain enabling to adopt timely countermeasures to minimize the adverse effects. In order to fully gain benefit from analytics, the collaboration of a detailed set of capabilities responsible for intersecting and integrating with multiple functions and teams across the business is required. Analytics and information management present a noteworthy chance for automakers to use quantitative techniques to support the planning of interferences across the customer lifecycle including but not limited to understanding the probable value of different customer segments; using that knowledge to tactically target new customers whilst maintaining the loyalty of existing customers, and improving customer experience to drive retention. The effective role played by big data analysis in the automobile industry has also been expanded in the research paper. The research paper discusses the scope and challenges of big data. The paper also elaborates on the working technology behind the concept of big data. The paper illustrates the working of MapReduce technology which executes in the back end and is responsible for performing data mining.

Keywords: Automobile industry, Big Data, MapReduce, Mining.

I. INTRODUCTION

Automakers are having to respond by offering 24/7 connectivity which requires broad growth in access channels and high dependence on social media and internet intended for communication and research. Furthermore, the role of the traditional dealer is being questioned and innovative new sales processes are being trailed by some automakers keen to exploit opportunities to differentiate and complement existing retail experiences on offer. There is a multitude of data available to support automakers in understanding their customers but the magnitude and complexity of this data limit their ability to collect, analyze and act on it. Automakers do not yet have full control over all 'insight' from external sources such as social media through to internal data and it is often spread in silos across different areas of the organization [1, 2]. To succeed automakers will need to fully understand customer needs and behaviors in order to develop a single view of the customer and thereby build compelling, differentiated offers throughout the sale and ownership cycle that are relevant in today's digital environment. For example, a powerful combination of online retailing and physical stores placed in high footfall areas, to build brand awareness, attract new target customer segments and provide a transformational retailing experience for shopping anytime, anywhere. Finally, customer retention needs to be placed high on the agenda

and with that comes the increasing importance of understanding and influencing customer experience across the lifecycle and journey, not just during the purchase phase [3, 4, 5].

The growth in touchpoints and information available on customers is increasing allowing automakers and dealers to focus on specific groups of customers with targeted messages and offers. This, together with the increasing volume and frequency of data available to track customer behavior, offers an opportunity for a more precise approach to configuring the optimal marketing and incentives mix for a targeted group of customers [6]. However, the variety of data sources makes it difficult to collate and interpret the available information to understand the impact of different offers in a timely manner. Those automakers that are able to effectively manage and target their fixed and variable marketing spend to improve engagement and the appeal of their offers, have a fantastic opportunity to improve customer engagement and drive sales through more targeted, informed and controlled marketing interventions [7, 8].

The amount of data available to automakers can be daunting and they need to find a way of collating and analyzing it so data-driven decisions can be made. Marketing analytics has the potential to significantly improve the decision making of automakers and returns they can deliver, by collating and analyzing marketing information and customer behaviors in a consistent and seamless environment. Having done this, automakers can then begin to leverage historical data to gain some insight into which levers are best applied in which situations [9, 10, 11].

Compared to other industries such as retail and banking, automakers are playing catch up when it comes to using analytics to gain insight into their customers and into effective management of marketing budgets. If done right however analytics has the potential to inform automakers of the impact of incentives on a specific model, in a specific geography for a specific customer type. Combining this with suitable dashboards and planning tools will enable OEMs to better plan and make data-driven decisions on the allocation of finite marketing spend [12]. With global spend on incentives and advertising in the industry, there is a huge opportunity for automakers. As two key elements of demand generation, if they can get their advertising and incentives right there is the opportunity to drive both increased sales performance and profitability [13, 14].

II. SOURCES OF BIG DATA

There are several sources available in today's market where data relevant to different fields and sectors can be gathered. Few of such sources are mentioned as under [15, 16].

Social Networks:

Twitter and Facebook

Blogs and comments

Pictures: Instagram, Flickr, Picasa, etc.

Videos: YouTube

Internet searches

Mobile data content (text messages)

User-generated maps

E-Mail

Traditional Business Systems:

Commercial transactions

Banking/stock records

E-commerce

Credit cards

Medical records

Internet of Things

Sensors: traffic, weather, mobile phone location, etc.

Security, surveillance videos, and images

Satellite images

Data from computer systems (logs, weblogs, etc.)

III. SCOPE OF BIG DATA ANALYTICS

Marketing Mix Analytics quantifies the contribution of marketing activities on sales by evaluating sales, marketing, and macro trends over a period of time and different geographies using statistical modeling methods. Over time this can be used to power scenario-based marketing planning; portfolio budget allocation; performance forecasting; and media optimization. It is therefore important to set expectations and align any solution to business maturity, needs and available data [17, 18]. Depending on the maturity of the automaker, an initial step will likely involve creating a performance dashboard that's able to display a range of data sources in an easily digestible format and allows better choices to be made on marketing spend focus. Once that's embedded the statistical modeling and scenario-based planning can be considered as the organization develops its analytics capability [19].

The advances witnessed in supply chain analytics have shown a dramatic shift away from reactive management models. Automakers make use of emerging capabilities to equip themselves with the ability to persistently sense and respond as the industry witness changes. Further advanced supply chain analytics can help automakers to analyze enormous databases via proven mathematical and analytical techniques, like linear and non-linear optimization, mathematical techniques, and regression analysis. Essentially, the ability to blend discrete data sources and use powerful big data tools to help drive actionable insights has improved significantly in recent years [20]. An example of this in its simplest form is highlighted through the use of product configuration and other web interactions which allow automakers to get early identification of new emerging trends such as a move to a certain option like an automatic gearbox or a particular color allowing automakers to forecast demand to a greater level of granularity. All these techniques and tools enable automakers to recognize patterns and correlations that were not visible in the past and were missed out. This provides automakers to analyze the business and supply chain with enhanced and more precise vision.

IV. TECHNOLOGICAL ASPECT

Big data comes into play where conventional database tools find themselves incompetent to handle large amounts of data in terms of volume, variety, veracity, value, and velocity. Today every sector, medical science, sports, retail, business, education, elections, and automotive industry, which generates huge data makes use of a framework like Hadoop. Hadoop is a Java-based programming framework and runs applications on large clusters of commodity hardware which is responsible for generating huge data. Hadoop is by far the most reliable, scalable, and flexible framework available for handling big data. HDFS in Hadoop stores a huge amount of data. Hadoop makes use of the MapReduce algorithm which operates in three stages; mapping, shuffling and reducing, alongside several other tools embedded within it like Pig, Hive, Zookeeper, HCatalog, Sqoop, Avro, Chukwa and many more, each responsible for performing a definite task.

A proper big data analysis delivers the opportunity to expose hidden markets, finds customer demands and cost reduction opportunities. It can be helpful in improving efficiency in telecommunication and medical treatments, and big data analytics is also helpful in social media campaigns and related digital marketing. Fig. 1 depicts the Hadoop framework along with its different tools where each tool performs a specific task.

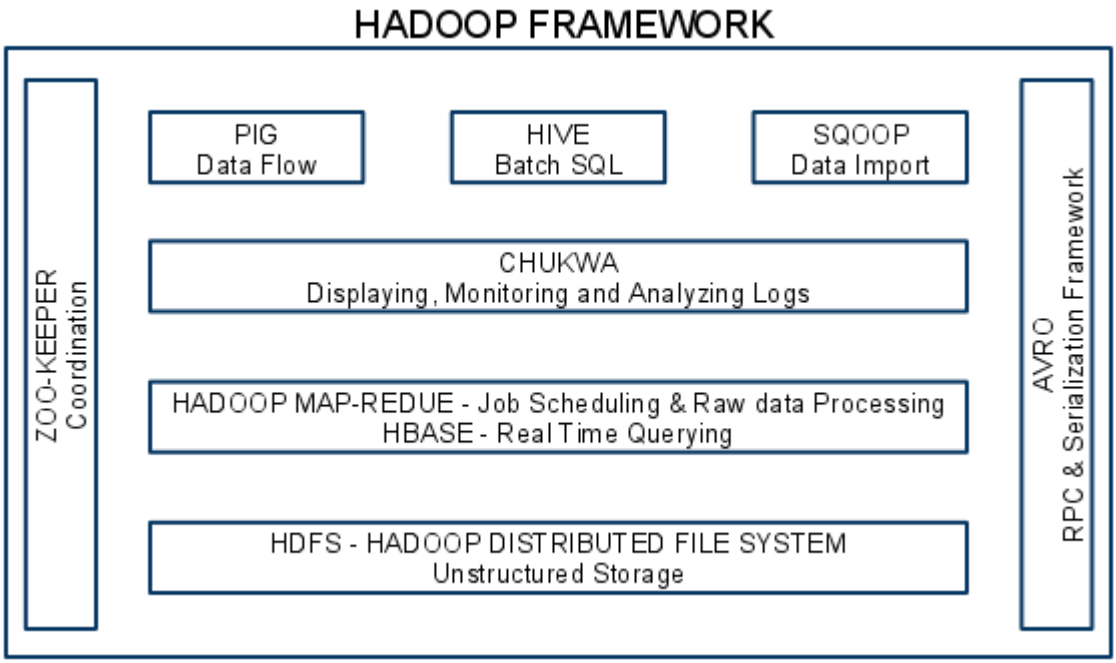

Fig. 1: Figure depicts the Hadoop framework with its tools

MapReduce algorithm works in three steps as illustrated in Fig. 2 mentioned below.

1. Map phase – The map phase is the first phase of the procedure. This phase is concentrated on conducting mapping. The mapping is initiated via making <key, value> pairs.

   This step performs the following two sub-steps:

   - Splitting – The function of the splitting step is to take data set as input from the source and divide it into smaller sub-datasets.
   - Mapping – These smaller sub-datasets obtained from splitting are given as input to the mapping phase to perform required action or computation on each sub-data set.

2. Shuffle phase- Shuffling also comprises of two steps mentioned as under.

   - Merging - Merging is performed on the basis of key-value pairs that have the same keys.
   - Sorting – The output obtained from the merging step acts as input for sorting. Sorting is intended to sort all key-value pairs by using Keys.

3. Reduce phase – Reduce phase is the last step and the output obtained in this final step is similar to the first step output. However, there is a huge difference between the final step <Key, Value> pairs and first step <Key, Value> pairs.

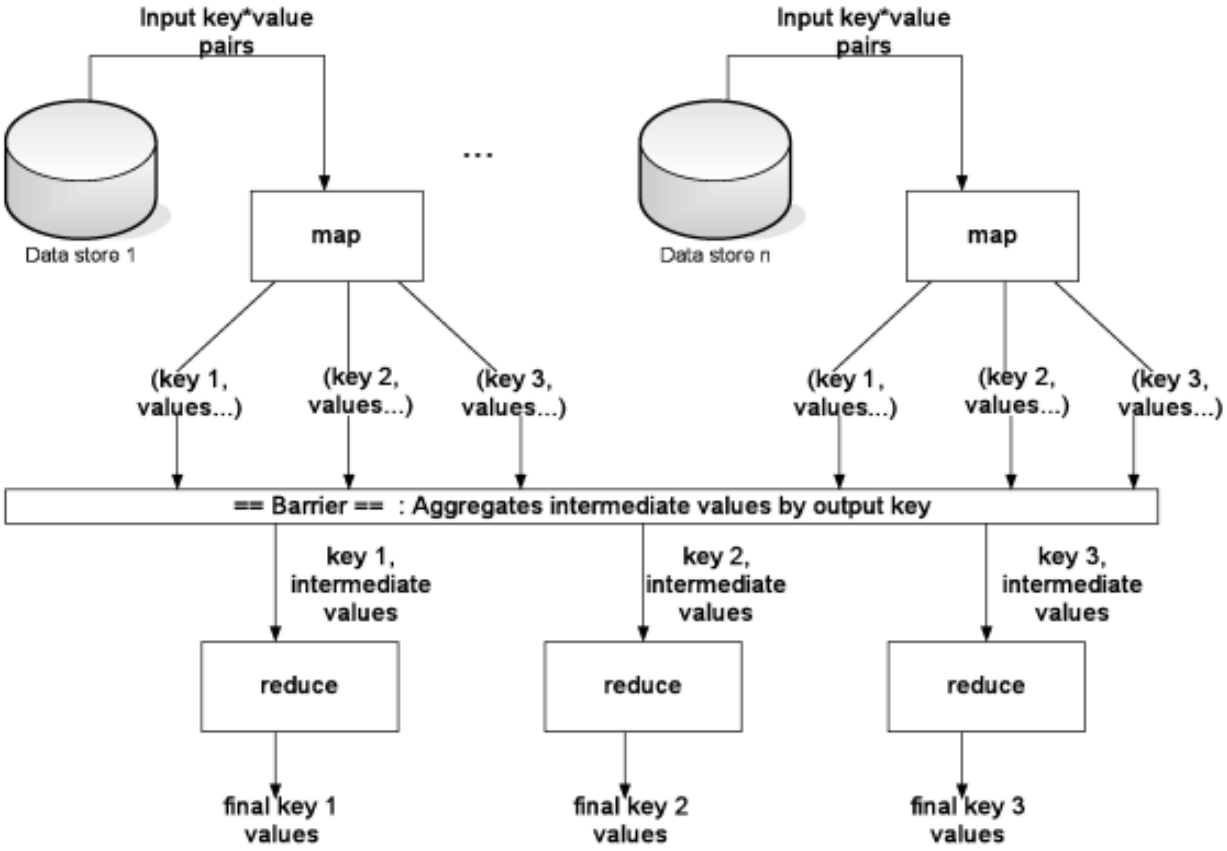

Fig. 2: Figure shows the working concept of the MapReduce algorithm

V. CONTRIBUTION AND IMPLEMENTATION

Firstly, a database is constructed which would include several attributes which describe the features of the cars under study. This database would be broadly divided into six segments described below.

Naming - This segment contains general information about the car.

Price - This segment contains information related to the price of the car.

Body - This segment contains information related to dimensions, wheelbase, and seating capacity of the car.

Engine - This segment includes information relevant to the power and displacement of car engine.

Fuel - This segment contains information related to the fuel type used in the car.

Tyres - This segment includes information related to dimensions of the tyres of the car.

The details of the attributes included in each section are mentioned below.

Naming

Make - Name of company manufacturing the car.

Model - Model of the car

Version	-	Version of the car, like 1.4 MultiAir 140Hp Automatic.
Status	-	Defines whether car manufacturing have been discontinued or is in production.
Car_class	-	Defines whether the car is smaal car, mini car, compact car, compact SUV etc.
Body_style	-	Defines whether car is Hatchback, sedan, SUV, MPV
Price		
Price_production_car	-	Provides price of car in production
Price_discontinued_car	-	Price of discontinued car
Body		
Length	-	Length of car in mm.
Width	-	Width of car in mm.
Height	-	Height of car in mm.
Wheelbase	-	Wheelbase of car in mm.
Ground_clearance	-	Value of ground clearance of car in mm
Bootspace	-	Value of boot space in litres
Number_of_doors	-	Number of doors in the car
Seating capacity	-	Number of seats within the car Engine
Number_of seating_rows	-	Value of seating rows in car
Engine		
Displacement Maxpower (bhp)	-	Displacement value of the car
Max_power (rpm)	-	Maximum power of the car
Max_torque (rpm)	-	Maximum torque value of the car
Transmission_type	-	Automatic or manual transmission
Number_of_gears	-	Number of gears in car
Drivertrain	-	FWD (Front Wheel Drive) or RWD (Rear Wheel Drive)
Fuel		
Fuel_type	-	Petrol (Gasoline) or Diesel
Mileage	-	Average of the vehicle

Fuel_Tank_Capacity	-	Value of capacity of fuel tank in litres
Tyres		
Tyre_Front	-	Dimension of front tyres
Tyre_Rear	-	Dimension of rear tyres

A glimpse of the constructed database is shown in Fig. 3 and Fig. 4.

Naming						
ID	Make	Model	Version	Status as 1 Nov 2017	Car class (personal opinion)	Body style
708	Chevrolet	Aveo [2006-2009]	1.4	Discontinued	Small car	Sedan
709	Chevrolet	Aveo [2006-2009]	LS 1.4	Discontinued	Small car	Sedan
1373	Chevrolet	Aveo [2006-2009]	LS 1.4 Ltd	Discontinued	Small car	Sedan
710	Chevrolet	Aveo [2006-2009]	LT 1.6	Discontinued	Small car	Sedan
711	Chevrolet	Aveo [2006-2009]	LT 1.6 Opt	Discontinued	Small car	Sedan
1810	Chevrolet	Aveo [2009-2012]	CNG 1.4	Discontinued	Small car	Sedan
1554	Chevrolet	Aveo [2009-2012]	LS 1.4	Discontinued	Small car	Sedan
1555	Chevrolet	Aveo [2009-2012]	LT 1.4	Discontinued	Small car	Sedan
1556	Chevrolet	Aveo [2009-2012]	LT 1.4 ABS	Discontinued	Small car	Sedan
1557	Chevrolet	Aveo [2009-2012]	LT 1.6 ABS	Discontinued	Small car	Sedan
1097	Chevrolet	Aveo U-VA [2006-2012]	1.2	Discontinued	Small car	Hatchback
1098	Chevrolet	Aveo U-VA [2006-2012]	LS 1.2	Discontinued	Small car	Hatchback
1548	Chevrolet	Aveo U-VA [2006-2012]	LS 1.2 Techno	Discontinued	Small car	Hatchback
1402	Chevrolet	Aveo U-VA [2006-2012]	LT 1.2	Discontinued	Small car	Hatchback
1100	Chevrolet	Aveo U-VA [2006-2012]	LT 1.2 ABS	Discontinued	Small car	Hatchback

1099	Chevrolet	Aveo U-VA [2006-2012]	LT 1.2 ABS & Airbag	Discontinued	Small car	Hatchback
4536	Chevrolet	Beat	LS Diesel	Production	Mini car	Hatchback

Fig. 3: Figure shows the glimpse of a Naming segment of the constructed database

Body								
Length (mm)	Width (mm)	Height (mm)	Wheelbase (mm)	Ground clearance (mm)	Bootspace (litres)	No of doors	Seating capacity	No of seating rows
4310	1710	1505	2480			4	5	
4310	1710	1505	2480			4	5	
4310	1710	1505	2480			4	5	
4310	1710	1505	2480			4	5	
4310	1710	1505	2480			4	5	
4310	1710	1505	2480	181	400	4	5	2
4310	1710	1505	2480	181	400	4	5	2
4310	1710	1505	2480	181	400	4	5	2
4310	1710	1505	2480	181	400	4	5	2
4310	1710	1505	2480			4	5	
3880	1670	1495	2480			5	5	
3880	1670	1495	2480	188	220	5	5	2
3880	1670	1495	2480			5	5	
3880	1670	1495	2480	188	220	5	5	2
3880	1670	1495	2480			5	5	
3880	1670	1495	2480			5	5	
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	165		5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	165		5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2
3640	1595	1520	2375	175	170	5	5	2
3640	1595	1520	2375	165	170	5	5	2

Fig. 4: Figure shows the glimpse of the Body segment of the constructed database

VI. CONCLUSION

The good news for automakers is that proactive and predictive quality analytics offers a solution.
Mature analytics systems can today process large quantities of data and offer various analysis methods, thus identifying potential faults in advance and provide opportunities to develop suitable counteractive actions through early signal detection as well as managing the data better. Analytics can allow you to see the unseen and discover insights in all areas of your business. We are able to draw upon a wide range of in-depth industry expertise and can help you make smart decisions that drive strategy, provide you with operational improvement and competitive differentiation.

REFERENCES

1. Gagandeep Jagdev et al., ―A Comparative study of Conventional Data Mining Algorithms against Map-Reduce Algorithm‖, in International Journal of Advance Research in Science and Engineering (IJARSE), ISSN (O) – 2319-8354, ISSN (P) – 2319-8346,Volume – 06, Issue – 05, May 2017
2. L. Gräning, B. Sendhoff: Shape Mining: A Holistic Data Mining Approach to Engineering Design. Advanced Engineering Informatics 28(2), 166-185, 2014.
3. Jagdev, G. (2018). Augmenting Revenue Growth in Retail Segment via Data Mining. International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), 5(3), pp.1-8. http://dx.doi.org/ 10.20431/ 2349-4859.0503001.
4. Davis J, Edgar T, Porter J, Bernaden J, Sarli M (2012) Smart manufacturing, manufacturing intelligence and demand-dynamic performance. Comput Chem Eng 47:145–156.
5. Gagandeep Jagdev et al., "Comparing Conventional Data Mining Algorithms with Hadoop based Map-Reduce Algorithm considering elections perspective*", International Journal of Innovative Research in Science and Engineering (IJIRSE),* ISSN: 2454-9665 (O), ISSN: 2455-0663(P), Volume – 3, Issue – 3, March 2017.
6. Hazen BT, Boone CA, Ezell JD, Jones-Farmer LA (2014) Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications.Int J Prod Econ 154:72–80.
7. Gagandeep Jagdevet al., "Association of Big Data with Map-Reduce Technology Augments for Economic Growth in Retail", *International Journal of Engineering Technology Science and Research (IJETSR),* ISSN: 2394 - 3386, Volume 4, Issue 2, February 2017.
8. Fosso Wamba S, Akter S, Edwards A, Chopin G, Gnanzou D. How big data can make big impact: Findings from a systematic review and a longitudinal case study. Int J Prod Econ. 2015;165:1–13.
9. Dr. Gagandeep Jagdev et.al (2017). ―Big Data in Retail Sector – An Evolution that Turned to a Revolution‖, International Journal of Research Studies in Computer Science and Engineering (IJRSCSE), ISSN: 2349-4840 (P), ISSN: 2349-4859(O), Volume 4, Issue 4, pp.43-52, DOI: http://dx.doi.org/10.2 0431/2349-4859.0404006.
10. McKinsey,"Big data: The next frontier for innovation, competition, and productivity,"2011.
11. Dr. Gagandeep Jagdev et al., ―A Study of Clustering and Classification Techniques involved in Data Mining‖, in International Journal of Advanced Technology in Engineering and Science (IJATES), ISSN – 2348-7550,Volume – 05, Issue – 05, May 2017.
12. Vera-baquero A, Colomo-palacios R, Molloy O (2014)"Towards a process to guide Big data based decision support systems for business processes,". In: Conference on ENTERprise information systems towards, vol 00., p 2212.
13. Gagandeep Jagdevet al., "Analyzing Maneuver of Hadoop Framework and MapR Algorithm Proficient in supervising Big Data", *International Journal of Advanced Technology in Engineering and Science (IJATES),* ISSN – 2348-7550,Volume – 05, Issue – 05, May 2017.

14. Wieringa R, Maiden N, Mead N, Rolland C (2006) Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requir Eng 11:102–107.

15. Gagandeep Jagdev et al., "Implementation and Applications of Big Data in Health Care Industry", *International Journal of Scientific and Technical Advancements (IJSTA)*, ISSN: 2454-1532, Volume 1, Issue 3, pp: 29 – 34.

16. Delen D, Demirkan H (2013) Data, information and analytics as services. Decis Support Syst 55(1):359–363.

17. Gagandeep Jagdev et al., "Big Data proposes an innovative concept for contesting elections in Indian subcontinent", *International Journal of Scientific and Technical Advancements (IJSTA),*Volume 1, Issue 3, pp. 23-28, 2015, ISSN No. 2454-1532.

18. Gagandeep Jagdev et al., ―Association of Big Data with Map-Reduce Technology Augments for Economic Growth in Retail‖, in International Journal of Engineering Technology Science and Research (IJETSR), ISSN: 2394 - 3386, Volume 4, Issue 2, February 2017

19. Gagandeep Jagdev et al., "Analyzing and Scripting Indian Election Strategies using Big Data via Apache Hadoop Framework", *IEEE Xplore*, DOI: 10.1109/WECON.2016.7993431, INSPEC Accession Number: 17061464, 27 July 2017

20. Gagandeep Jagdev et al., "Excavating Big Data associated to Indian election scenario via Apache Hadoop",*International Journal of Advanced Research in Computer Science*. 2016 November, 7 (6), pp. 117-123.

21. Gagandeep Jagdev et al., "A Comparative study of conventional data mining algorithms against Map-Reduce algorithm",*International Journal of Advanced Research in Science and Engineering.* 2017 May. 6 (5), pp. 325-335.