ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 12, December 2018

Big Data Analytics in the Marketing of Automotives

Vishnu T B[1], Dr. Hemalatha[2]

PhD Scholar, Department of Computer Science & Applications, Karpagam Academy of Higher Education, Coimbatore[1]

Assistant Professor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore[2]

Abstract: It is already true that Big Data has drawn huge attention from researchers in information sciences, policy and decision makers in governments and enterprises. However, there are so much potential and highly useful values hidden in the huge volume of data. A new scientific paradigm is born as Data Intensive Scientific Discovery (DISD), also known as Big Data problems. Whether industrial 4.0 nor Internet industry, for today's industrial manufacturing enterprises, it should be to make full use of information and communication technology to deal with the arrival of smart and effective large data, combining products, machinery and human resources into together, according to the unexpected speed about the mode of sales product, it can change the manufacturing enterprises to process innovation and reform. The objective is to take the automobile manufacturing industry as an example, based on sale car large data analysis, using big data technology. Recent technology innovations, many of which are based on the capture and analysis of big data, are transforming the automotive industry in a pace deemed inconceivable just a short time ago. At the heart of this transformation is the new role of the car itself, and the increasingly sophisticated abilities that "intelligent cars" possess to communicate with individuals, enterprises, and devices around them. Company leaders in the automotive industry clearly recognize that by embracing the concept of big data, they can access a mass of opportunities for differentiation, growth, and innovation that revolutionize the very core of existing business models. In order to unlock this potential, the key challenge is to develop and implement a big data strategy, which is tailored to the capture, analysis, and interpretation of the ever increasing quantities of structured and unstructured data which will be received from drivers, vehicles, and other devices. Only those companies which incorporate a big data strategy in their transformation agendas will be abled to reap the rewards offered by the zeta byte revolution. As the objective suggests, the purpose of this research is to enhance the automotive industry by applying new strategy which analyze the Big Data. The tools such as Hadoop and Map Reduce algorithm can be applied for achieving the objective.

Keywords: Big Data, Hadoop, Map Reduce

I. INTRODUCTION

Big Data has been one of the current and future research frontiers. It is right to say that Big Data will revolutionize many fields, including business, the scientific research, public administration, and so on. For the definition of the Big Data, there are various different explanations from 3Vs to 4Vs. Doug Laney used volume, velocity and variety, known as 3Vs, to characterize the concept of Big Data. The term volume is the size of the data set, velocity indicates the speed of data in and out, and variety describes the range of data types and sources. Sometimes, people extend another V according to their special requirements. The fourth V can be value, variability, or virtual. More commonly, Big Data is a collection of very huge data sets with a great diversity of types so that it becomes difficult to process by using state-of-the-art data processing approaches or traditional data processing platforms. In 2012, Gartner retrieved and gave a more detailed definition as: "Big Data are high-volume, high-velocity, and/or high-variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization". Big Data has changed the way that we adopt in doing businesses, managements and researches. Data-intensive science especially in data-intensive computing is coming into the world that aims to provide the tools that we need to handle the Big Data problems. In the field of commerce and business, According to estimates, the volume of business data worldwide, across almost companies, doubles every 1.2 years. Taking retail industry as an example, we try to give a brief demonstration for the functionalities of Big Data in commercial activities. There are around 267 million transactions per day in Wal-Mart's 6000 stores worldwide. For seeking for higher competitiveness in retail, Wal-Mart recently collaborated with Hewlett Packard to establish a data warehouse which has a capability to store 4 peta bytes.

Tools used for Big Data Analysis:

- NoSQL DatabasesMongoDB, CouchDB, Cassandra, Redis, BigTable, Hbase, Hypertable, Voldemort, Riak, ZooKeeper

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 12, December 2018

- MapReduce Hadoop, Hive, Pig, Cascading, Cascalog, mrjob, Caffeine, S4, MapR, Acunu, Flume, Kafka, Azkaban, Oozie, Greenplum
- Storage S3, Hadoop Distributed File System
- Servers EC2, Google App Engine, Elastic, Beanstalk, Heroku

Here, the research deals with the use of Big Data in the Automotive Industry. "Intelligent Car" concept and "AutoMAT" (which is European project has implemented the provision of accessing of general data for the entire vehicle) helps the automotive companies to perform more in the industry. The survey report produces the brief of various studies based on the topic. Based on the previous studies, the proposed system adds something more for the betterment of the automotive manufacturers, dealers and customers.

II. EXISTING SYSTEM (REVIEWS)

1: Mr. Takashi, Consultant of Automotive and Transportation at FROST & SULLIVAN presents his findings for the future automotive industry with the use of Big Data analytics.
The paper presents an analysis of the Connected Car concept. According to the analysis, Big Data analytics is expected to add value to business models involving autonomous vehicles, which are expected to generate more than 750MB of data per second. He have presented various factors that affect the manufacturing of new automobiles in the automotive industry such as "Lead of Digital Market", "Internet Aggregators", "Predictive maintenance", "Product Performance Analysis", "User and Dealer satisfaction" and "Advanced Mobility services for the vehicle" etc.
The factor "Lead of Digital Market" (The Digital Retailing experience) creates opportunities up to 80% from Big Data. An annual savings of $700 to $800 per vehicle is expected with the use of Big Data. The Data Analytics will enable newer visualization techniques and create market opportunities. Big Data will bring about differentiation based on varied metrics such as brand awareness, digital engagement of customers, Response time, and vehicle configurability satisfaction. In future, the industry will witness the foray of integrated service providers with expertise in providing end-to-end services across the automotive value chain, i.e., from consulting to implementation.

2: Dr. M Hanumanthappa, Dept of Computer Science, Banglore University, presented a paper in ACEEE Int. J on Information Technology on the topic "Predicting the Future of Car Manufacturing Industry using Data Mining techniques".

Data Mining tasks can be categorized into either Prediction or Description. Descriptive Mining techniques are Clustering, Association Rule Mining and Sequential Pattern Mining. The Predictive Mining techniques are Classification, Regression and Deviation Detection. Predictive modeling is a process used in Predictive Analysis which is a combination of Statistics, Machine Learning, Database techniques, Pattern recognition and Optimization techniques, which is concerned with future probabilities and trends. Linear Regression is the simplest technique used to model continuous valued functions.
For prediction linear regression can be used to fit a predictive model to an observed data set of Y and X values. After developing such a model, if an additional value of X is then given without its accompanying value of Y, the fitted model can be used to make a prediction of the value of Y. Here we are trying to predict the number of cars to be manufactured by a car manufacturing company by considering the following numeric parameters. The response variable, Y = number of cars manufactured and Predictor variable, X = year. Now when an additional value for X that is a particular year is given it is possible to find the possible value for Y, the number of cars to be manufactured, using the following equations.

$$Y = b + wX$$

Where 'b' is the regression coefficients specifying the 'Y' intercept 'w' is the regression coefficients specifying the slope of the line. Now if we consider the regression coefficients as weights, then

$$Y = w0 + w1X$$

Let D be a training set consisting of values of predictor variable, X for some population and their associated values for response variable, Y. The training set contains |D| data points of the form (x1, y1), (x2, y2), ……., (x|D|, y|D|)
The regression coefficients can be estimated with the method of least squares, were

$$w_1 = \frac{\sum_{i=1}^{|D|}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{|D|}(x_i - \bar{x})^2}$$

and $w_0 = \bar{y} - w_1\bar{x}$ where ¯x is the mean value of x1, x2 ,……., x|D|, and is the mean value of y1,y2 ,……..,y|D
This paper introduces the application of data mining technology in the car manufacturing unit and obtains an analysis result from small data, may expand the sample capacity in the practical application, to obtain more accurate conclusion.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 12, December 2018

Some of the modifications that can be made to simple linear regression are rather than one predictor variable, are more predictor variables can be used.

3: Mr. Qui Zhang, Ningbo University, China published in International Congress of Information and Communication Technology (ICICT-2017) published a paper in "Car Sales Analysis based on the application of Big Data".
He presented the paper in Big Data analysis which is used in automobile industry based on the sale of cars' analysis, using Data Mining technology, through the Java program "web crawler" for Data Collection. According to the paper, the sale of cars is divided into 3 layer architecture. They are Technical Layer, Strategic Layer and Decision layer. The concept of Big Data came into existence in the industry by the year of 2015. The traditional approach to manufacturing such as Process Design, Production Operations need to be changed with the industry under the premises of Big Data challenge. The concept was illustrated by Data mining by the author. Data Mining does not have a proper definition. It is the discovery of information or knowledge from a large data. The web crawler was the java program for the author to collect the data for this analysis. According to the author the following describes the Data model for the new strategy to be applied.

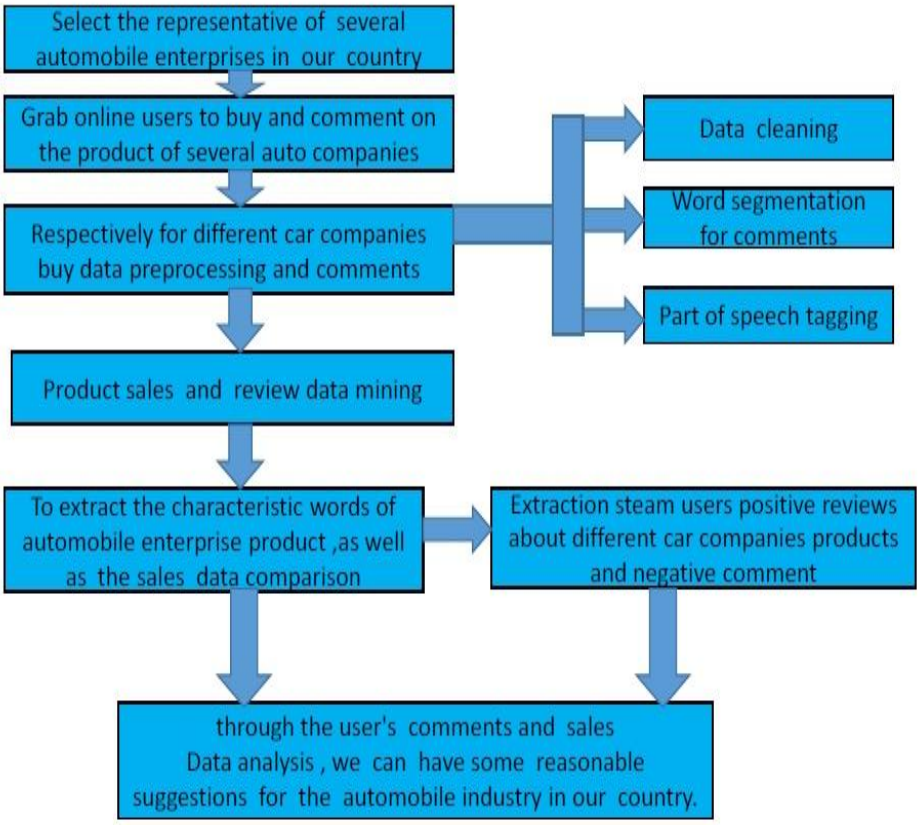


Fig: Car Sales Data Model

4: Dr. Poonam Chauhan, Aditya Mahajan and Dhiraj Lohare, Assistant Professor, K J Somaiya Institute of Management Studies and Research, Mumbai presented a paper in National Journal of Multidisciplinary Research and Development about "The Role of Big Data in retail Customer-Centric Marketing". The retail industry has experienced some significant challenges and opportunities in recent years. One such challenge is proliferation of digital media that assists makes customers test and compare products in one store and then buy them elsewhere, this phenomenon is popularly known as 'showrooming'. Maintaining customer acquisition, retention and loyalty is another demanding task. As against this there is incredible opportunity due to penetration and usage of mobile and internet technology. The use of big data analytics and reporting can generate insights that will improve the profitability for retailers.
The purpose of the paper is to provide overview of big data analytics in various stages of retail process - tracking emerging popular products, foretelling of sales and future demand through predictive simulation, explore its application in Price optimization, targeted promotions and overall enhancement in customer experience and suggest future developments in marketing, consumer profiling and predictive analytics for retail business. The authors collected data from various journals, books and through interviews. From the brief summary of the interview with Mr. Naimesh Tungare, Assistant VP, TATA Group, according to the strategy of TATA Group, there are minimum four people dedicated for Big Data Analytics for one particular store. Continuous data is collected from online and offline mediums

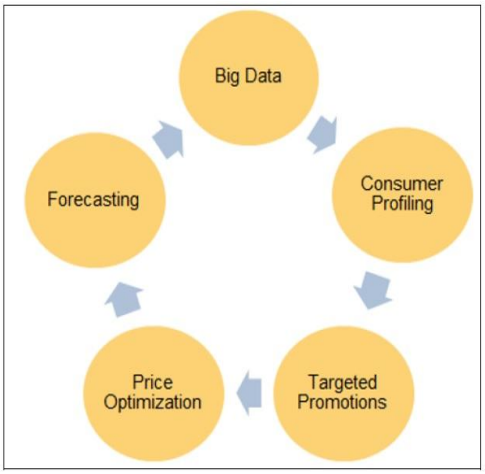
ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 12, December 2018

like surveys, reviews, comments, ratings, salesperson interaction, calls etc. The inter relationship is explained from the following figure. The paper concludes that the value of Big Data analytics has to be extricated from the hype surrounding it. Though there are challenges, like the lack of clear big data strategies, security concerns and the need for workforce re-skilling, the growth potential of Big Data is extraordinary.



5: Mr. Michael Voigt (Business Transformation Chief, SAP, Germany), Mr. Christopher Bennison (Business Transformation senior Consultant, SAP, Germany) and Prof. Maik Hammerschidt (Professor of Marketing, University of Goettingen), presented paper in the journal "360- the Business Transformation Journal", No: 10, April 2015, on the topic "Big Data in the Automotive Industry" titled "Gaining Traction". Intelligent cars are on the rise and will change our perception of automobiles. The authors of this article illustrate the importance of a Big Data strategy in helping automotive companies to transform and offer consumers a completely new driving experience. The authors suggest that the automotive companies should turn into data-driven strategy (Big Data Analytics) for the upcoming new models. Their study shows that the data-driven decision makers have 5% higher productivity, 6% higher profits and up to 50% market value. An intelligent car concept should be the strategy of automotive companies. An intelligent car concept not only allows a huge amount of real-time data to be gathered directly from the vehicle, but it also allows the knowledge created by big data analysis to be published directly to the driver. Eg: BMW's Connected Drive, which currently offers unique apps for real-time traffic information, concierge services, intelligent emergency call, online entertainment, and BMW specific TeleServices. The challenge for the automotive companies is to build an IT infrastructure for the Big Data Analysis. Big Data Chain is the concept that is to be followed by the companies for the above challenge. Use of Hadoop technology along with Map Reduce algorithm supports the Big Data challenge. The authors suggest that only those companies which are able to implement Big Data strategy will be able to transform their business and master the "zeta byte" revolution.

III. PROPOSED SYSTEM

From the previous studies, it is clear that Big Data can influence more for the development of automotive industry. Every year the research is carried out which benefits the industry as well as the dealers, even customers also. The concept of Intelligent Cars was a turning point for the automotive industry. The objective of the proposed system is to us the Big Data Analytics (Prediction Analysis) to develop a new algorithm or a flowchart through which the market analysis for automobiles can be done. For the creation of this solution, the variables which are affected should be identified and use them effectively. It is proposed to have an IT wing (Big Data Chain) which can handle the Big Data Challenges.

IV. DATA COLLECTION

The Big Data analysis can be done using unstructured data. In the field of marketing, the data can be collected from the following various sources.

- Analyze purchase behavior in terms of purchase history, credit and return history available.-TRANSACTION PROFILE
- Hit/page view/Click path/duration of stay - WEB BEHAVIOR
- SMS/Geo location analysis- MOBILE ACTIVITY
- Open rates/spam complaints/bounce- EMAIL BEHAVIOR

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 7, Issue 12, December 2018

- Age/Sex/location/other info-DEMOGRAPHICS
- Followers/Posts/Influencers etc- SOCIAL MEDIA
- Subscriptions/language- PERSONALIZATION

V. TOOLS USED

1: R language: R is an open source software package to perform statistical analysis on data. R is a programming language used by data scientist statisticians and others who need to make statistical analysis of data and glean key insights from data using mechanisms, such as regression, clustering, classification, and text analysis. R provides a wide variety of statistical, machine learning (linear and nonlinear modeling, classic statistical tests, time-series analysis, classification, clustering) and graphical techniques, and is highly extensible. R has various built-in as well as extended functions for statistical, machine learning, and visualization tasks such as:

• Data extraction
• Data cleaning
• Data loading
• Data transformation
• Statistical analysis
• Predictive modeling
• Data visualization

2: Hadoop: Apache Hadoop is an open source Java framework for processing and querying vast amounts of data on large clusters of commodity hardware. Apache project, initiated and led by Yahoo! and Doug Cutting. It relies on an active community of contributors from all over the world for its success. With a significant technology investment by Yahoo!, Apache Hadoop has become an enterprise-ready cloud computing technology. It is becoming the industry de facto framework for Big Data processing. Hadoop changes the economics and the dynamics of large-scale computing. Its impact can be boiled down to four salient characteristics. Hadoop enables scalable, cost-effective, flexible, fault-tolerant solutions.

Apache Hadoop has two main features:

• HDFS (Hadoop Distributed File System): HDFS is Hadoop's own rack-aware filesystem, which is a UNIX-based data storage layer of Hadoop. HDFS is derived from concepts of Google file system. On HDFS, data files are replicated as sequences of blocks in the cluster.
•MapReduce:MapReduce is a programming model for processing large datasets distributed on a large cluster. MapReduce is the heart of Hadoop. Its programming paradigm allows performing massive data processing across thousands of servers configured with Hadoop clusters. This is derived from Google MapReduce.

CONCLUSION

To ensure being at the forefront of the industry transformation, automotive companies must devise big data strategies that go much further than just defining how the increased volume, velocity, and variety of data will be supported within the enterprise. Successful big data strategies focus far more on the subsequent use of this data for decision making purposes. Intelligent car concepts, which gather a huge amount of real-time structured data from the vehicle and publish the results of the data analysis directly to the driver, are a central element of the big data strategies of many automotive OEMs. Using the huge and unstructured data collected from various resources, finding the variables that affect the industry and by applying proper algorithms/flowcharts we can arrive at a conclusion for the strategy that is to be applied for the industry.

REFERENCES

[1]. International Congress of Information and Communication Technology (ICICT 2017), "Car Sales Analysis Based On the Application of Big Data" by Qi Zhang, Hongfei Zhan, Junhe Yu.
[2]. National Journal of Multidisciplinary Research and Development, Volume 2; Issue 3; September 2017; Page No. 484-488, "Role of big data in retail customer-centric marketing", by 1 Dr. Poonam Chauhan, 2 Aditya Mahajan, 3 Dhiraj Lohare
[3]. 360° – the Business Transformation Journal No.10|April 2015, titled "Gaining Traction, Big Data in the Automotive Industry" by Mr. Michael Voigt (Business Transformation Chief, SAP, Germany), Mr. Christopher Bennison (Business Transformation senior Consultant, SAP, Germany) and Prof. Maik Hammerschidt (Professor of Marketing, University of Goettingen).
[4]. ACEEE Int. J on Information Technology, titled "Predicting the Future of Car Manufacturing Industry using Data Mining techniques", by Dr. M Hanumanthappa, Dept of Computer Science, Banglore University.
[5]. 2015 ICT Outlook Executive Briefing, Mr. Takashi, Consultant of Automotive and Transportation at FROST & SULLIVAN