Available online at www.ijapie.org

International Journal of Advanced
Production and Industrial Engineering

IJAPIE-SI-OSCM 703 (2017) 07–11

Special Issue on

OPERATIONS
&
SUPPLY CHAIN
MANAGEMENT

Conceptual Framework for the Application of Big Data in the Automobile Industry

Shubham Fialoke*, Vibhu Dharmadhikari, Varun Bhati, Vaibhav Gupta, Saurabh Agarwal
(Delhi Technological University, Delhi, India)
**Email:* shubhamfialoke@gmail.com

Abstract : *The paper sheds light on big data analytics, the science of machine learning viz. data science and how it can be used in today's automobile industry. Data science is a new field that has emerged out of the coupling of computer science and statistics. The 'what?' and 'how?' of data sciences is explained thoroughly. After defining what Big Data and Data Science is, the job market of data science is explored. This includes the jobs data scientists undertake and the skills acquired by them viz. machine learning and natural language processing to find patterns in data and make predictions. Furthermore, the future of big data analytics in manufacturing and production is analysed, particularly in connected vehicular systems. Today, with the advancement of technology, much media attention has been focused on an area where big data and automobiles intersect: Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication, collectively referred to as Machine-to-Machine (M2M) communication. This paper focuses primarily on data science in driverless cars and the advent of Internet of Things (IoT) in automobile technology.*

Keywords: *Big Data, Automobile Industry, Intrnet of Things, Vehicle-to-Vehicle (V2V) , Vehicle-to-Infrastructure (V2I) communication,*

1. INTRODUCTION

A reference to data science, analytics, big data, or some combination thereof is imperative in almost all modern day publications, online or in the physical world. This highlights the importance and extensive use of data in today's technology driven world. Data is everywhere. In fact, the amount of digital data that exists is growing at a rapid rate, doubling every two years, and changing the way we live. An article by Forbes states that data is growing faster than ever before and by the year 2020, about 1.7 megabytes of new information will be created every second for every human being on the planet.
To have a better understanding of today's systems, it is crucial to differentiate between Data Science, Big Data, and Data Analytics, based on what it is, where it is used, the skills you need to become a professional in the field, and the salary prospects in each field.

1.1 Data Science: Dealing with unstructured and structured data, Data Science is a field that comprises of everything related to data cleansing, preparation, and analysis. Data Science is the combination of statistics, mathematics, programming, problem solving, capturing data in ingenious ways, the ability to look at things differently, and the activity of cleansing, preparing, and aligning the data.
In simple terms, it is the umbrella of techniques used when trying to extract insights and information from data.

1.2 Big Data: Big Data refers to humongous volumes of data that cannot be processed effectively with the traditional applications that exist. The processing of Big Data begins with the raw data that isn't aggregated and is most often impossible to store in the memory of a single computer.

A buzzword that is used to describe immense volumes of data, both unstructured and structured, Big Data inundates a business on a day-to-day basis. Big Data is something that can be used to analyse insights which can lead to better decision and strategic business moves.
The definition of Big Data, given by Gartner is, "Big data is high-volume, and high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation".

1.3 Data Analytics: Data Analytics is the science of examining raw data with the purpose of drawing conclusions about that information. Data Analytics involves applying an algorithmic or mechanical process to derive insights. For example, running through a number of data sets to look for meaningful correlations between each other. It is used in a number of industries to allow the organizations and companies to make better decisions as well as verify and disprove existing theories or models.

The focus of Data Analytics lies in inference, which is the process of deriving conclusions that are solely based on what the researcher already knows.

Fig. 1: Comparison between Data Science, Big Data and Data Analytics [A]

2 APPLICATIONS

2.1 *Data Science:*

2.1.1 Internet search: Search engines make use of data science algorithms to deliver best results for search queries in fraction of seconds.

2.1.2 Digital Advertisements: The entire digital marketing spectrum uses the data science algorithms - from display banners to digital billboards. This is the mean reason for digital ads getting higher CTR than traditional advertisements.

2.1.3 Recommender systems: The recommender systems not only make it easy to find relevant products from billions of products available but also adds a lot to user experience. A lot of companies use this system to promote their products and suggestions in accordance to the user's demands and relevance of information. The recommendations are based on the user's previous search results.

2.2 *Big Data:*

2.2.1 Big Data for financial services: Credit card companies, retail banks, private wealth management advisories, insurance firms, venture finds, and institutional investment banks use big data for their financial services. The common problem among them all is the massive amounts of multi structured data living in multiple disparate systems which can be solved by big data. Thus big data is used in a number of ways like:

i) Customer Analytics
ii) Compliance Analytics
iii) Fraud Analytics
iv) Operational Analytics

2.2.2 Big Data in communications: Gaining new subscribers, retaining customers, and expanding within current subscriber bases are top priorities for telecommunication service providers. The solutions to these challenges lie in the ability to combine and analyse the masses of customer generated data and machine generated data that is being created every day.

2.2.3 Big Data for Retail: Brick and Mortar or an online e-tailer, the answer to staying the game and being competitive is understanding the customer better to serve them. This requires the ability to analyse all the disparate data sources that companies deal with every day, including the weblogs, customer transaction data, social media, store branded credit card data, and loyalty program data.

2.3 *Data Analysis:*

2.3.1 Healthcare: The main challenge for hospitals with cost pressures tightens is to treat as many patients as they can efficiently, keeping in mind the improvement of quality of care. Instrument and machine data is being used increasingly to track as well as optimize patient flow, treatment, and equipment use in the hospitals. It is estimated that there will be a 1% efficiency gain that could yield more than $63 billion in the global health care savings.

2.3.2 Travel: Data analytics is able to optimize the buying experience through the mobile/ web log and the social media data analysis. Travel sights can gain insights into the customer's desires and preferences. Products can be up-sold by correlating the current sales to the subsequent browsing increase browse-to-buy conversions via customized packages and offers. Personalized travel recommendations can also be delivered by data analytics based on social media data.

2.3.3 Gaming: Data Analytics helps in collecting data to optimize and spend within as well as across games. Game companies gain insight into the dislikes, the relationships, and the likes of the users.

2.3.4 Energy Management: Most firms are using data analytics for energy management, including smart-grid management, energy optimization, energy distribution, and building automation in utility companies. The application here is centred on the controlling and monitoring of network devices, dispatch crews, and manage service outrages. Utilities are given the ability to integrate millions of data points in the network performance and lets the engineers to use the analytics to monitor the network.

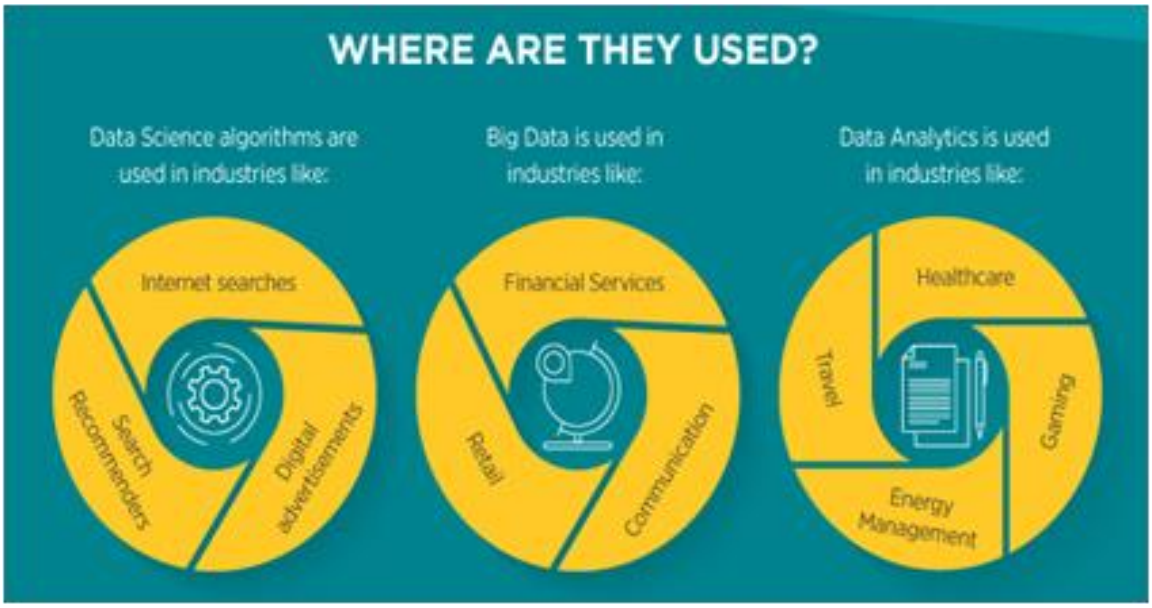

Fig. 2: Applications of Data Science, Big Data and Data Analytics [B]

3. JOB OF A DATA SCIENTIST AND THE SKILL SET REQUIRED

A data scientist does three tasks- data analysis, modelling/statistics and engineering/prototyping. The datasets that a data scientist deals with are usually too large for typical spreadsheet programs like SQL and Excel. The data is first cleaned, which means it is formatted to be more homogeneous. The next step is exploratory data analysis, which combines visualization and data sense. This involves plotting the data, building models and algorithms and finding patterns in it. Finally, the result that is obtained is such that it can be communicated easily and the implications of it can be understood. The whole process needs persistence, stats, and programming skills — skills that allow the data scientist to understand biases in the data and for debugging logging output from code.

3.1 SQL: An absolute necessity since it is something that is being used for a long time and most people are comfortable with it in the industry.

3.2 Hadoop: An open-source framework that is used to store and process big data in a distributed environment across clusters of computers using simple programming models.

3.3 UNIX: Basic UNIX command line is useful in pulling data. Raw data is stored in UNIX/Linux servers and is valuable because it allows access to the data without the need for a database.

Since the industry does not have an agreed upon definition of a data scientist, the skills required to be one aren't defined either. There are some general tools that most data scientist use. Tools for collecting/pulling data:

Tools for data analysis and presenting results:

1. R: The programming language is useful for data visualisation.

2. Tableau: Another tool for visualisation that is very easy to use.

3. SAS: A business analytics software.

4. Python: The programming language has useful packages such as matplotlib, numpy and scipy that can be used for data analysis and visualisation.

4. BIG DATA AND THE AUTOMOTIVE INDUSTRY

In the near future automakers will not only be evaluated on the safety, physical and performance characteristics of their vehicles, but also on the basis of their solutions in different dimensions like Electric, Autonomous, Connected, Mobility Services (EAC+MS), and Information. In the following two parts we will discuss the big data challenge faced by the automotive industry.

4.1 The first post provides the why and makes two points:

4.1.1 The Automakers must take initiatives in the information business and to be effectual in the same, automakers must change their viewpoint and start thinking about big data in and around the car.

4.1.2 Information that is given in the "EAC+MS" implies that players in the automotive system become serious about big data. Companies which are new to the automotive industry such as Google, Tesla, Faraday Future, etc. realize this imperative.

4.1.3 To understand the importance of the fifth dimension i.e. information, one needs to understand that offering "EAC+MS" means that, in addition to manufacturing vehicles with variations of electrification (from fully electric to hybrid-electric, to fuel cell vehicles, etc.), levels of autonomous driving and Internet connectivity, the automaker provides transportation solutions. These include the vehicle, mobility services like parking, roadside assistance, and also broader transportation services, e.g., ride-sharing, car rental. All these solutions require data and in order to offer such solutions, automakers must be there in the information business. They must be able to create information, build systems that utilize big data, and utilize third-party data, e.g., weather data to better serve the vehicle's occupants.

5. CONCEPTUAL FRAMEWORK OF THE DRIVERLESS CAR

Laser Illuminating Detection and Ranging is the technique which is extensively used in driverless cars. It is used to build a 3D map of the surroundings. Such a car can see lines on the road to know where the lane is. The cars are equipped with radar units that enable them to detect the speed of the surrounding objects.

There are three main hardwares in the driverless car model: sensors, processors, and actuators. Images and information gleaned from sensors travel through the processor, which effectively tells the car what to do via actuators—tools that allow a computer to control physical components like brakes, or steering wheels.

Data can also come from other cars. This data can also be used to better map the environments. If cars were to process the image of a new sign put in place, that data would eventually become part of the system. This is possible with the cloud access that the cars must be equipped with.

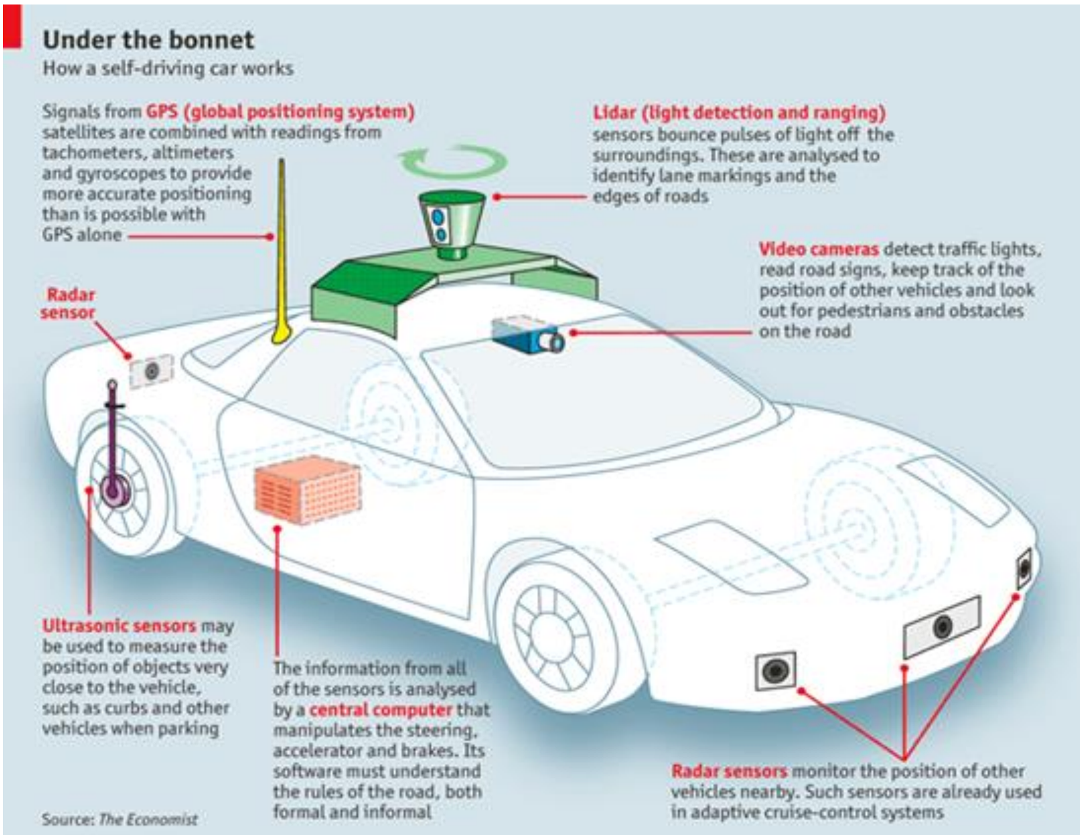

Fig. 3: Working of a self-driving car[C]

Driverless cars are the result of smart technology, learning algorithms, real time data processing and catalogued information based on previous experience. It's the incredible software that processes the responses and behaviours in real time. The more a car is driven, the more it knows. But the difficulty arises in the very specific split-second decisions that drivers have to make every day. Scientists cannot possibly program a car to recognize every object and to behave appropriately in every situation. But by absorbing more data through driving and real experience, it may learn. It may learn that items in the road will cause other drivers to swerve and behave appropriately in such situations.

By transforming driver experiences into programmable information, scientists are making driverless cars much more practical. The ability to process real time data doesn't necessarily equal the perfect car. Data science is necessary to determine which data is and isn't important.

These cars use various models like predictive and prescriptive models to deal with the incoming sensory information in a practical way.

6. DISADVANTAGES OF DATA SCIENCE IN DRIVERLESS CARS

Data science and analytics may be the up and coming technology in the future of smart cars, but it surely has some disadvantages too, as listed below.

6.1 The major disadvantage of driverless cars is that they can be hacked.

6.2 These cars generate huge amounts of data that can be cumbersome to manage.

6.3 These cars have access to cloud data for data processing and manipulations.

6.4 The advertisers can track consumers and direct stream commercials to the dashboard.

7. CONCLUSION

Nonetheless, the mixture of complexity and common sense that is driving automated cars is breath-taking. The magic isn't just in images of driverless, hovering cars, but the ways that scientists are teaching computers to learn. No, they might never be capable of split-second decisions quite like a human, but with the right data, teachers and algorithms, robots can mimic the human mind very well. The driverless car is surely a step forward to the future of smart cars which can prove to be highly efficient and convenient.

REFERENCES

[1.] Lo Frank ; What is data science? ; https://datajobs.com/what-is-datascience

[2.] Levine Daniel ; How many data scientists are there? ; https://blog.rjmetrics.com/2015/10/05/how-many-data-scientists-are-there/

[3.] Davenport Thomas, Patil DJ ; Data scientist: The sexiest job of the 21st century ; https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21stcentury/

[4.] Agarwal Ritu, Dhar Vasant ; Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research ; Information Systems Research Vol. 25, No. 3, September 2014, pp. 443–448

[5.] [A.] Future of big data analytics: machine data and connected vehicles ; http://www.simafore.com/blog/bid/207756/Future-of-big-data-analytics-machine-data-and-connected-vehicles

[6.] [C.] Augur Hannah ; How data science is driving the driverless car ; http://dataconomy.com/how-data-science-is-driving-the-driverless-car/

[7.] Simoudis Evangelos ; The automotive industry's big data challenge (Part 1) ; https://www.enterpriseirregulars.com/104636/automotive-industrys-big-data-challenge-part-1/

[8.] [B.] Schober Scott ; Cybersecurity and the future of smart cars: Are smart cars too smart for their own good? ; http://www.ibmbigdatahub.com/blog/cybersecurity-and-future-smart-cars