Statistical techniques for big data analytics in IoT-enabled green supply chain management: a survey

BY

Wafaa A. Saleh[a], Sherine M. Abdelkader[b], Heba Rashad[a], Amal Abdelgawad[a]

a Department of Decision Support, Faculty of Computers and Informatics, Zagazig University, 44519 Zagazig

b Computers and Systems Department, Electronics Research Institute, Cairo 11843, Egypt.

Abstract:

In the manufacturing operation, intelligent Supply Chain Management systems (SCMS) can improve the quality of products, reduce cost, and accelerate the decision making process. The incorporation of environmentally sustainable processes into SCMS minimizes the overall environmental impact which is the target of Green Supply Chain Management (GSCM). The intelligence of the GSCM systems makes the business smarter. For this reason, it is always a concern to utilize cutting-edge ideas and technologies to optimize the operation of these systems. Internet of Things (IoT) is a promising Information technological (IT) concept that allows environmental objects to communicate with each other automatically and without human intervention. IoT is one of the most important IT solutions that provides intelligence and sustainability to GSCM systems. The significant feature of IoT is the huge volumes of data, called 'big data' generated by the IoT sensors, installed on the different entities of the chain. To this end, big data processing in real time is a need for decision makers to preserve their companies' competitive advantage. There are many big data analytics techniques in the literature to target this issue. Our work will focus on surveying the statistical techniques that can be used in the analysis of big data generated from the IoT sensors situated on the different parts of GSCM to improve its performance, flexibility, productivity, and optimization of its resources through the effective analysis of the large amounts of raw data involved in IoT enabled GSCM, We will also uncover the best tools that can be used for this purpose.

Keywords:
Statistical techniques, Big Data Analytics, Internet of Things, Green Supply Chain Management

1. Introduction

Supply Chain Management (SCM) is a concept that represents the centralized management of the lifecycle of products From cradle to grave, includes the management of all different activities sourcing, procurement, warehousing and logistics until the product reaches its final destination (consumer)[1]. SCM coordinate the different chain partners: suppliers, manufacturers, distributors, transporters, retailers and customers [2]. The effectivity of the SCM make the business smarter, so it is a great point of research that needs more intelligent solutions. Customers and businesses are increasingly concerned about their carbon footprint as a result of climate change and global warming, and how data can help them achieve a sustainable and Green Supply Chain [1], Green Supply Chain management (GSCM) is our focus in this survey which mandates that the environmental idea be incorporated into each and every stage of the product and service in a supply chain to address the economy's environmental problems [3]. An intelligent GSCM is a forward-thinking chain that makes use of information technology (IT) paradigms and other technological tools to boost efficiency, streamline processes, and provides higher service level.

Internet of Things (IoT) is a promising IT concept that allows environmental objects to communicate with each other automatically and without human intervention. Of course, IoT is one of the most important IT solutions that provides intelligence and sustainability to GSCM systems. The over-accelerating evolution of IoT sensors has open a new challenge for IT researchers which is the analysis of the huge amounts of data generated by these IoT entities in real time [4]. the large amount of data known as big data is being collected from a number of sources as a result of the digital revolution and the widespread availability of electronic gadgets [5]. The analysis of these collected data can provide very useful information for decision makers, but analyzing these huge amounts of data and delivering it in time and appropriate format require special analyzing techniques[5][6]. The analysis of data generated from the variant parts of The supply chain management (SCM) process can significantly improve a factory's performance, flexibility, productivity, waste reduction, resource optimization both inside and out the factory, and more environmentally friendly production methods [7].

The vast amounts of data used in the supply chain aid not only in increasing efficiency, but also in demand forecasting and know consumer preferences [8]. It has proved that IoT can enhance GSCM will allow for real-time data collection and efficiency, and real-time communication among supply chain entities [9]. However the challenge of analyzing this amount of data generated from IoT sensors requires innovative techniques to analyze the data in an effective model that facilitate the process of knowledge discovery[10]. Big Data analytics techniques can be used for this challenge, which necessitate innovative statistical and mathematical methodologies to identify useful features or patterns in the data, such as relationships, trends, and outliers that aid in taking wise business decisions. To provide the trustworthiness, validity, and consistency required for business decision making, Statistical Modeling is required to supplement Data Mining. Because graphically visualized models are easier and more efficient for sharing complicated information. Big data analysts frequently create a labelled diagram to aid in the description of the output[11].

There are many papers in the literature that survey the big data analytics techniques in SCM, but this research focus on the statistical big data analytics techniques in the analysis of IoT – enabled GSCM data. In this regard this paper makes the following contributions:

1. Discussion of the big data analytics techniques and tools with special focus on the statistical techniques is provided.
2. The overall architecture of Big Data analytics in IoT-enabled GSCM with a discussion of the implementation details is introduced.
3. A table of literature studies that relate to this paper topic is provided.

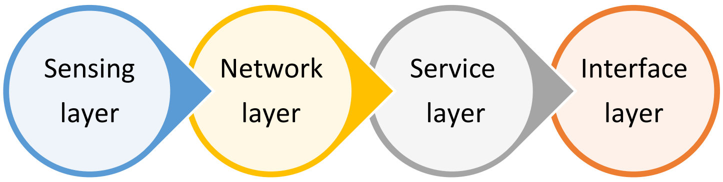
The following sections are organized as follow: section 2: IoT and big data overview, section 3: Big data analytics in IoT-enabled GSCM, section 4: Statistical techniques for analyzing the big data generated from GSCM entities, then we conclude our work in section 5 finally listed our references in section 6.

2. IoT & Big Data Overview

This section provide overview about IoT and big data technologies starting with IoT

2.1 IoT

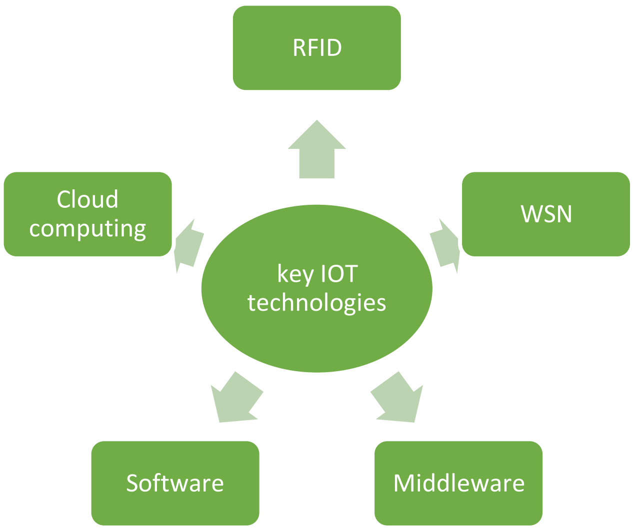
The main idea behind the Internet of Things is that we need to support computers with their own data-gathering capabilities so they can "see," "hear," and "smell" the world for themselves. Computers can now observe, identify, and understand the world without the limitations of human-entered data thanks to RFID and sensor technology[12].The two words that compose IoT: Internet and things can be analyzed to give better understanding of the concept of IoT the first word Internet refers to the internet connection needed in IoT, and the second word things refers to the things that are embedded with electronics to enable it collect data from its environment and send it on the internet for processing, these things also can receive instructions for doing some operations [13].There are four main layers that make up The IoT network [10]:



1. Sensing layer that connects various devices such as RFID tags, sensors, and actuators.

2. Networking layer that allows data to be transferred over a wired or wireless network.
3. Service layer, which uses middleware technology to connect services and applications.
4. An interface layer that shows the information to the user and allows them to interact with the system.

There are five key IoT technologies [14]:



- Radio Frequency Identification (RFID): it is a technological equipment that is placed on physical things and allows for the identification, tracking, and transmission of data.
- Wireless Sensor Networks (WSN): it is a wireless network made up of a collection of sensors that are used to sense the status of environmental objects, such as their geographical position, speed, humidity, and etc. WSN is used in many critical applications such as tele-surgery, medical diagnosis, modern agriculture, smart cities, and GSCM. The sensors in the WSN can also communicate and collaborate with RFID tags [9].
- Middleware: it is a logical layer that facilitate the communication between IoT infrastructure (like RFID tags, sensors and actuators) and application developers.
- Cloud computing: a powerful multi-access computing server accessible via the internet that allows users to share and access a collection of IT resources (hardware, software, storage, networks, and so on) on demand. Cloud computing is a main component of the IoT networks because of big data produced by IoT gadgets [9]. The produced big data must be analyzed via

powerful computing devices (cluster of connected servers such in cloud data centers) in order to enable decision makers to get timely, easy to understand, and valuable data. The real-time analysis of the IoT data collected from the supply chain entities gives the manager supply chain's updates at once, which leads to taking very efficient and profitable decisions.

- IoT application software: IoT enables creation of a wide range of applications in industry and academia. IoT applications enable reliable and robust object-to-object and individuals-to-object interactions, whereas devices and networks provide physical connectivity. Data/messages must be collected and processed timely by IoT applications on devices.

IoT has the potential to change and automate everything in our life. The smart GSCM is one of application areas of IoT. [15]**.** IoT sensors is used in the supply chain for collecting end-user data [10]. It also has a number of capabilities that can help with SCM, such as cost minimization, efficient inventory management, object tracking, resource saving, and latency minimization [7]. However, the impact of IoT on various supply chain processes is somewhat mysterious. Our goal in this review is to determine the impact of IoT on GSCM by conducting a systematic literature review.

2.2 Big Data

The big data term , can be described as the massive flood of data in the Exabyte's and beyond has broadened the scope of technological capability for storing, managing, processing, interpreting, and visualising vast amounts of data. [4] [2]. RFID and video and audio capturing sensors, tracking sensors, warehousing and transportation data, Global Positioning System (GPS), GPRS (General Packet Radio Service), Vehicle and engineering tracking systems, such as black-boxes on heavy vehicles, and container and contents tracking systems are just a few examples of big data sources [15], [16].

Big data can be categorized according to the five Vs[5], [10] as shown in figure 1 : Volume: massive amounts of data are generated every day all over the world, Velocity: the speed of collecting data , the reliability with which data is sent, the efficiency with which data is stored, the speed with which usable knowledge is discovered, and the speed with which decision-making models and algorithms are developed. Variety property of big data refers to the various types of data that must be analyzed. Veracity refers to the process of verifying data to ensure that only the best data is selected; this verification is usually carried out under the supervision of specific authorities and security levels. Value: the most important "V" from a business standpoint, the value of big data is typically derived from insight discovery and pattern recognition, which leads to more efficient operations.

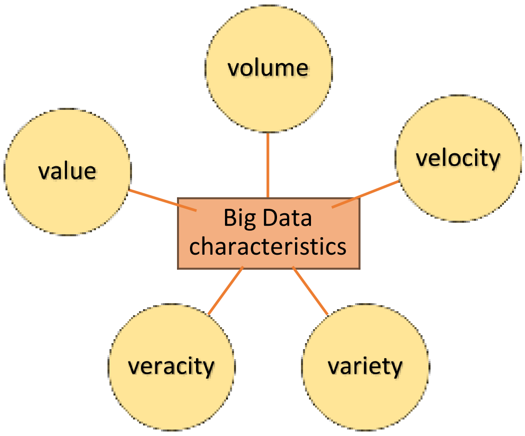


Figure 1: Big Data 5V

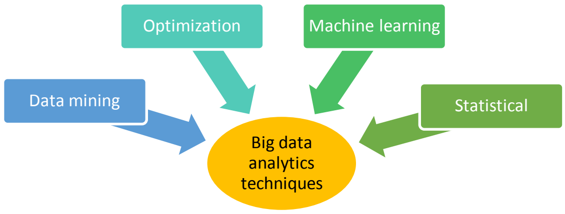
The systems that deal with big data should meet special requirements to be capable of handling big data operations efficiently [17]**:**

- **Scalability:** the system ability to manage increased demands. This ability must be increased to meet the Big Data volume. The scalability can be increased vertically by adding more processors, memory or faster hardware into a single server, or horizontally by adding more servers into a group of cooperating servers, called a cluster.
- **Distribution models:** bring many great benefits when working with Big Data. The system can store more data, handle more read or write operations per time, and provide availability even when there are network problems or a server crashes.
- **Consistency:** When the replication is used, it provides many benefits. On the other hand, it brings problems with consistency. When two users update the same data at the same time, each on a different server, they create a write-write conflict. This ability must be taken in consideration when working on big data.

3. Big Data Analytics in IoT-enabled GSCM

3.1 Big Data Analytics

Big Data Analytics (BDA) is the utilization of various analytical methodologies and the ability to rapidly process huge amounts and types of data to uncover hidden patterns and output useful insights for decision-makers [18]–[21]. It is a time-consuming task that involves everything from data acquision until visualization and presentation. As data scales reach petabyte levels, traditional data processing tools produce unsatisfactory results when working with such complex structure and volumes of datasets [22]. Traditional data analytics methods face many challenges when dealing with big data such as coding/decoding, processing, pattern detection and analysis, transfer and sharing, as well as scale and complexity issues when analyzing such large amounts of data, due to its unstructured and heterogeneous nature[1], We can create better computational models that lead to better decision-making, which can be applied to a wide range of real-world scenarios [15].



The big data analytics techniques can be categorized into four categories [23]:

- **Statistical techniques:** the basic BDA which provides a systematic framework for collecting and analyzing big data in order to draw inferences and conclusions. Correlation, regression, Delphi, and multivariate statistical analysis are examples of common statistical methods. Statistics aids in overcoming the high-speed computation requirements of BDA [6], [24].
- **Machine learning techniques:** machine learning is a branch of computer science that focus on making computers mimic human being intelligence [25]. The algorithms in this category focuses on making computers learn and think as humans. They try to extract knowledge from specific data in the same way as humans [26]. Linear regression, logistic regression, Artificial neural networks (ANN), decision tree, random forest, naïve bayes, k-nearest neighbors, k-means, and support vector machines (SVM) are examples of common machine learning techniques [27]. Machine learning is the foundation for increasing computer intelligence and is at the heart of artificial intelligence [28].
- **Data mining techniques:** The process of predicting outputs by looking for patterns, outliers, anomalies, and correlations in big data sets [29]. Text mining, web mining, and spatial mining are three common types of data mining. They all aid in the discovery of knowledge and the decision-making process [19].
- **Optimization techniques:** Analytical method for finding optimal solutions in a short amount of time by converting real-world problems into mathematical models. Genetic algorithms, whale algorithm, simulated annealing, gorilla-troops algorithm, tabu search, particle swarm algorithms, and evolutionary algorithms are examples of common optimization method [30]–[33]. For quantitative decision making, optimization aids in the discovery of optimal solutions.

According to the literature, eco-design, green innovation, and clean production are among the BDA-enabled internal environment management activities.[34]. Internal environment management has benefited from statistical techniques such as causative analysis and correlation analysis, which have assisted in revealing potential and regular operation patterns by processing large volumes of internal operation data.[35]. As a result, businesses can gain valuable forward-looking information and provide environmental solutions for internal environment management[36]. Firms can, for example, investigate energy consumption patterns and redesign production processes to save energy [35].

3.2 Big data analytics tools

Each Big Data tool focuses only on a specific field of a big data usage. For example, Kibana aims at visualization, and HDFS targets storage. Although some tools focus on multiple fields, no such tool aims at them all. Therefore, the solution of a specific big data problem may consist of a set of tools that have to be used. So the solution implementer has to have theoretical knowledge about all available tools and must be able to decide which tool fits the problem the best, and have practical experience with all of the chosen tools.

Every relevant Big Data tool can be classified into one of these components. Some tools can also be classified into more components, because of their features (transferring, resource management, storage, processing, advanced analytics, orchestration, presentation)[17]. Table 1 present the common tools that can be used for statistical big data analytics [37], [38]:

Tool	Description
Hadoop	Software framework for handling big data and clustered file systems. It uses the Map Reduce model of programming to analyze big datasets, written in java programming language [39].
R	It is free and open source programming language specialized in writing programming scripts in data science application. I is name refers to first character of its two creators' names, "Ross Ihaka" and "Robert Gentleman". R is very good and easy to write scripting language in writing statistical analytics and data visualization instructions. R is a dynamic and free multi-paradigm software environment, written in C programming language. R can be integrated with Hadoop and python in data science applications [40], [41].
python	Python is easy to learn open source programming language. Python is now called the language of data science because of its built in libraries that facilitate work with data big data. Python have many library that relate to data representation, visualization and processing such as pandas, nu mpy, scikit-learn, keras and etc [42].
Spark	Data analytics, machine learning algorithms, and fast cluster computing are all part of this open source framework. written in

	Scala, Java, Python, and R.
Storm	It is a Fault-tolerant instant computational framework that is cross-platform, parallel stream processing. storm is open-source and free, written in java programming language.
RapidMiner	Cross-platform tool that integrates data science, machine learning, and predictive analytics into one environment, provides a GUI to design and execute analytical workflows, written in Java programming language.

3.3 GSCM

Keeping our Environment safe and healthy is the responsibility of all society individuals. Recently, researchers from different fields focused their works in this topic. Therefore, it is important for any business manager to consider environmental concerns in taking business decisions. To this end, incorporating the environmental idea into each and every stage of the product and service supply chain to address the economy's environmental problems is a need. This is the reason why GSCM is taking a special focus among operations’ and other science branches’ researchers. GSCM is an approach to improving process of product manufacturing and distribution according to environmental regulations [3]. The rise in carbon gas radiation and environmental pollution by businesses operations has necessitated the realignment of supply chain operations in order to conserve scarce resources. The goal of GSCM is to eliminate or reduce energy waste throughout the supply chain.[1]. GSCM has five key elements[43]:

- **Green procurement**: this element means buying products and services that have a minimal negative impact on the environment, consisting of activities such as material reduction, reuse, and recycling during the purchasing process. It takes human beings health and environmental concerns into account when looking for high-quality goods and services at reasonable prices.
- **Green design**: it is a concept used to describe facets of design that aim to make the final product more sustainable and environmentally friendly. Green design can be used in a variety of contexts, such as car and airplane design to improve aerodynamics and reduce fuel consumption.
- **Green Operations**: relates to the incorporation of environmental management policies into product development operations in order to enhance environmental performance [44], [45].
- **Green Manufacturing and Remanufacturing:** refers to the revitalization of manufacturing processes and the establishment of ecologically responsible operations in the industry. It is, in essence, the "greening" of manufacturing, in which fewer raw materials are used, produce less pollution and waste, recycle and reuse materials, and lessen emissions in their processes.

- **Reverse logistics**: Concerned with the recycling of end user products and materials. It is a form of SCM in which goods are returned from customers to manufacturer or producer. It is required for processes such as returns and recycling after a client use the product.

The significance of GSCM is primarily due to the environment's degradation, which includes raw material resources' depleting and rising pollution levels. However, it is not only about being environmentally conscious; it is also about making sound business decisions and increasing profits [43].

3.4 The role of big data analytics in IoT-enabled GSCM

Currently, the importance of BDA has been identified by researchers in GSCM, which mandates that the environmental idea be incorporated in each and every stage of the product and service in the SC to address problems of environment. BDA assists businesses in analysing environmental data across the SC and generating valuable insights to help them improve their GSCM [46]. GSCM has recognized as a crucial concern for management staff, policymakers, and the general public at large as environmental consciousness of global warming, toxic pollutants, and chemical spills has grown [47]. Actual instant analysis of dynamic energy use and carbon emission data is possible with BDA, as well as the improvement of manufacturing procedures to save energy and reduce emissions [48]. BDA assists businesses in mining environmental data across the supply chain and generating insights to help them improve GSCM[49]. When BDA is used inefficiently to eliminate information asymmetry, it can be difficult to assess suppliers' greenness[23]. One reason for the ineffective use of BDA is a lack of knowledge about the different methodologies of BDA that can be used, and the GSCM aspects that can be improved by BDA[50]. The benefits of BDA in green product development have been proposed in recent research [51], Green supplier selection[52], green logistics[53], and customer green cooperation [54]. minimized carbon and harmful gas emissions, energy savings, improved long-term development collaboration, and increased client green sense of achievement are all expected to result from these BDA-enabled GSCM practices [23].

The selection of green suppliers is one of the most crucial aspects of green procurement[54]. BDA assists businesses in evaluating suppliers' environmental performance based on historical data [55], [52]. Collaboration with suppliers is another important aspect of green purchasing [54]. Environmental requirements and sustainability needs in addition to business information must be communicated with suppliers and sellers in order to achieve significant long-term performance. BDA eliminates information synchronization and improves information sharing with suppliers, resulting in increased environmental collaborative effort. Packaging, transportation, reverse logistics, and information sharing are all examples of customer green cooperation [47], [54], [56]. Clients' engagement in green purchasing, clean production, product recycling, and eco-design should be supported by a company's overall BDA capabilities, according to the author. As a result, businesses can minimized their carbon emission and environmental pollution. Green logistics, which is an important part of client green collaboration, refers to the application of effective logistics technology to the design and execution of logistics efforts to minimize environmental contamination and waste of business resources [57].

3.5 The architecture of Big Data analytics in IoT-enabled GSCM

BDA is ideal for processing large amounts of unstructured environmental data and generating high level insights to help with GSCM process. BDA, for example, allows for instant analysis of dynamic energy exhaustion and harmful gases emission data, as well as the evolution of manufacturing processes to save energy and reduce emissions[23], BDA assists businesses in mining environmental data across the supply chain and generating insights to help them improve GSCM.

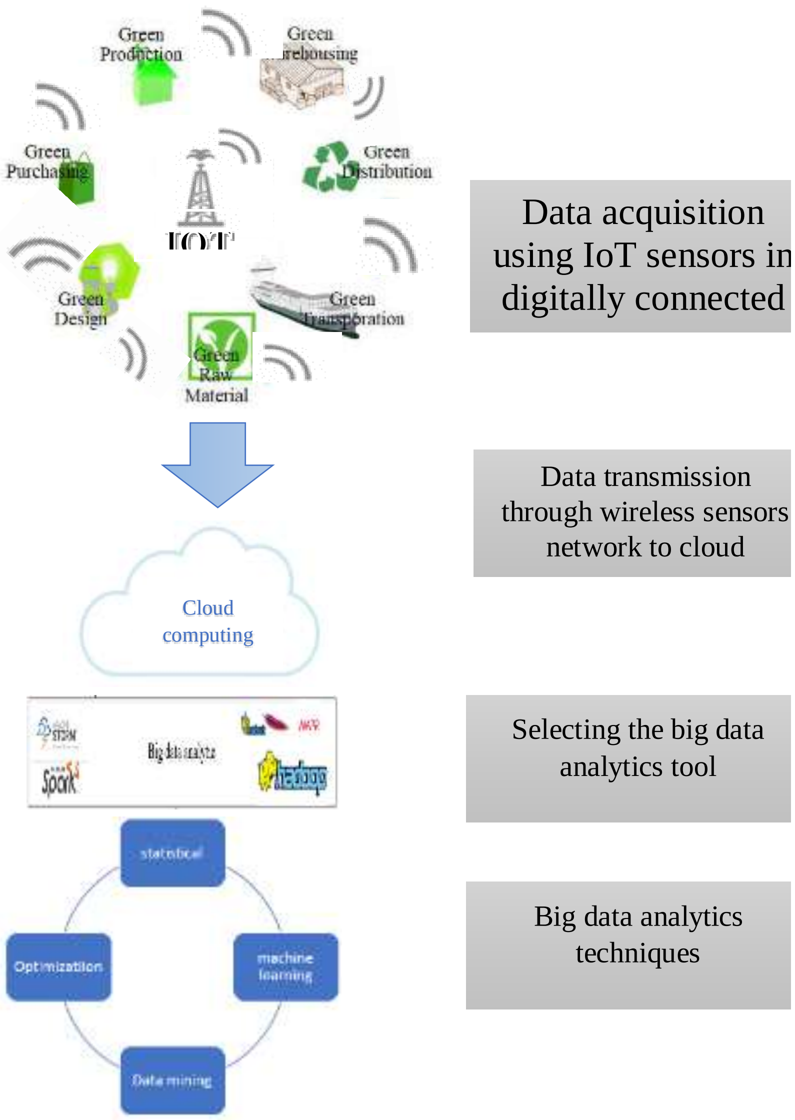


Figure 2: architectural framework of Big Data analytics in IoT-enabled GSCM

As shown in **Figure 2,** the life of big data analytics in GSCM. The IoT sensors are installed at all GSCM components to collect the data and then forward it to access points through WSN, which send it to edge computing or cloud computing for further processing.

Applications that work on IoT data collect from GSCM components run on cloud computing. These applications analyze GSCM data and generate reports, or notifications to decision makers.

Each GSCM component can be identified by RFID that is sent with data to cloud applications.

To analyze the big data collected from the GSCM components, two important things should be determined:

- The best big data analytics tool that suit the analysis.
- The type of analytics and the analytics algorithm.

So as we mentioned the analysis of big data requires systems with special analytics techniques to uncover the hidden structures in the data to extract the useful information for the decision makers[17]. The analytics techniques that can handle big data in GSCM can be divided into four areas internal environment, management, green purchasing, and customer green cooperation [19], [58].

4. Statistical techniques for analyzing the big data generated from GSCM entities.

This section discusses the recent statistical techniques that address the challenges resulted in big data systems. According to [59], [60] the statistical big data methodologies can be categorized into the following main classes:

1. **Divide and conquer:** the methodologies fall in this class follow the split and merge principle to analyze the big data challenge. The dataset is divided into N small normally computable blocks. Each block of data is passed as an input to selected statistical algorithm. The results from the different blocks are collected using appropriate strategy. The process of combining the results from the N blocks represents the challenge of the divide and conquer methods, however, the processing of the small blocks simultaneously on different cores can reduce the time complexity for the algorithm which is the big advantage of these methods [61].
2. **Online updating for stream data:** the idea behind the methods of this category is processing streamed data in real time without storing it [62].
3. **Subsampling-based methods:** The sampling methods represent efficient and general option to the big data problem. In these methods a subset of the original dataset is chosen to be used instead of the original dataset to do model estimation, prediction, and statistical inference using a carefully designed probability distribution. The design of the probability distribution for taking the sample is the most important component for these methods. The simple uniform distribution is a naive choice for the probability distribution. The Bayesian bootstraps [60] and leveraging [63] are two subsampling methods that are proposed in 2014.

Table 2 summarize the state of the art of literature in the topic of this paper. It lists the set of paper worked on utilizing statistical big data analytics techniques in the area of IoT-enabled GSCM.

Table 2: statistical techniques for analyzing the big data generated by the IoT devices installed in the GSCM entities.

Author	Journal	year	Analytics technique
El-Kassar, Abdul Nasser Singh, Sanjay Kumar [35].	Technological Forecasting and Social Change.	2019	Statistical correlation analysis.
Wang, Chenxiao, Qingpu Zhang, and Wei Zhang.[64].	Research in Transportation Business and Management.	2020	Statistical Regression analysis.
Zhang, Yingfeng, et al [65]	Journal of Cleaner Production	2017	Statistical Regression analysis
Tseng, Ming-Lang, et al [66].	Resources, Conservation and Recycling.	2019	Statistical correlation analysis.
Luthra, S., D. Garg, and A. Haleem[67].	Journal of The Institution of Engineers (India):	2014	Statistical regression analysis.

	Series C.		
Kim, Jinsoo, and Jongtae Rhee [68].	International Journal of Production Research.	2012	Statistical regression analysis.
Testa, Francesco, and Fabio Iraldo[69].	Journal of Cleaner Production.	2010	Statistical Multivariate regression analysis.
Shohan, S., et al [70].	International Journal of Sustainable Development and World Ecology.	2019	Statistical Multivariate regression analysis.

5. Conclusion:

This paper surveys the statistical techniques that can be put in use for analyzing big data generated from GSCM components by installed IoT devices. We firstly covered the concept of big data, IoT, GSCM, and Big Data Analytics then provided the framework of developing IoT applications for processing big data generated from GSCM using cloud or edge computing technologies. We also discussed the significance of big data analytics in processing GSCM data and explained the best tools that can be used for this purpose. Finally we have surveyed the most papers that worked on this subject from the literature.

6. References:

[1] A. K. Ahmed, C. B. S. Kumar, and S. Nallusamy, "Study on environmental impact through analysis of big data for sustainable and green supply chain management," vol. 8, no. 1, pp. 1245–1254, 2018.

[2] R. Y. Zhong, S. T. Newman, G. Q. Huang, and S. Lan, "Computers & Industrial Engineering Big Data for supply chain management in the service and manufacturing sectors : Challenges , opportunities , and future perspectives," *Comput. Ind. Eng.*, vol. 101, pp. 572–591, 2016, doi: 10.1016/j.cie.2016.07.013.

[3] J. Yao, H. Shi, and C. Liu, "Optimising the configuration of green supply chains under mass personalisation," *Int. J. Prod. Res.*, vol. 0, no. 0, pp. 1–19, 2020, doi: 10.1080/00207543.2020.1723814.

[4] T. Wahyuningsih, "Problems, Challenges, and Opportunities Visualization on Big Data," *J. Appl. Data Sci.*, vol. 1, no. 1, pp. 20–28, 2020, doi: 10.47738/jads.v1i1.8.

[5] F. Mannering, C. R. Bhat, V. Shankar, and M. Abdel-Aty, "Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis," *Anal. Methods Accid. Res.*, vol. 25, p. 100113, 2020, doi: 10.1016/j.amar.2020.100113.

[6] J. P. Cabrera-Sánchez and Á. F. Villarejo-Ramos, "Acceptance and use of big data techniques in services companies," *J. Retail. Consum. Serv.*, vol. 52, no. July 2019, p. 101888, 2020, doi: 10.1016/j.jretconser.2019.101888.

[7] C. Ntinas, D. Gkortzis, A. Papadopoulos, and D. Ioannidis, “Industry 4.0 sustainable supply chains: An application of an IoT enabled scrap metal management solution,” *J. Clean. Prod.*, p. 122377, 2020, doi: 10.1016/j.jclepro.2020.122377.

[8] H. Nozari, M. Fallah, H. Kazemipoor, and S. E. Najafi, “Big data analysis of IoT-based supply chain management considering FMCG industries,” *Bus. Informatics*, vol. 15, no. 1, pp. 78–96, 2021, doi: 10.17323/2587-814X.2021.1.78.96.

[9] M. Ben-daya, E. Hassini, and Z. Bahroun, “Internet of things and supply chain management : a literature review,” *Int. J. Prod. Res.*, vol. 7543, no. November, pp. 1–23, 2017, doi: 10.1080/00207543.2017.1402140.

[10] A. Aryal and B. Li, “The emerging big data analytics and IoT in supply chain management : a systematic review,” 2018, doi: 10.1108/SCM-03-2018-0149.

[11] N. R. Vajjhala and K. D. Strang, “Statistical Modeling and Visualizing Open Big Data Using a Terrorism Case Study,” 2015, doi: 10.1109/FiCloud.2015.15.

[12] A. Kevin, “That ’ Internet of Things ’ Thing,” *RFiD J.*, p. 4986, 2010.

[13] R. Badia-Melis, L. Ruiz-Garcia, J. Garcia-Hierro, and J. I. Robla Villalba, “Refrigerated fruit storage monitoring combining two different wireless sensing technologies: RFID and WSN,” *Sensors (Switzerland)*, vol. 15, no. 3, pp. 4781–4795, 2015, doi: 10.3390/s150304781.

[14] I. Lee and K. Lee, “The Internet of Things (IoT): Applications, investments, and challenges for enterprises,” *Bus. Horiz.*, vol. 58, no. 4, pp. 431–440, 2015, doi: 10.1016/j.bushor.2015.03.008.

[15] S. Verma, A. Bhatia, A. Chug, and A. P. Singh, *Recent Advancements in Multimedia Big Data Computing for IoT Applications in Precision Agriculture : Opportunities , Issues , and Challenges*. Springer Singapore.

[16] M. Marjani *et al.*, “Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges,” *IEEE Access*, vol. 5, pp. 5247–5261, 2017, doi: 10.1109/ACCESS.2017.2689040.

[17] “Tools for big data analysis,” 2018.

[18] R. D. Raut, S. K. Mangla, V. S. Narwane, M. Dora, and M. Liu, “Big Data Analytics as a mediator in Lean, Agile, Resilient, and Green (LARG) practices effects on sustainable supply chains,” *Transp. Res. Part E Logist. Transp. Rev.*, vol. 145, no. November 2020, p. 102170, 2021, doi: 10.1016/j.tre.2020.102170.

[19] T. M. Choi, S. W. Wallace, and Y. Wang, “Big Data Analytics in Operations Management,” *Prod. Oper. Manag.*, vol. 27, no. 10, pp. 1868–1883, 2018, doi: 10.1111/poms.12838.

[20] G. Palareti *et al.*, “Comparison between different D-Dimer cutoff values to assess the individual risk of recurrent venous thromboembolism: Analysis of results obtained in the DULCIS study,” *Int. J. Lab. Hematol.*, vol. 38, no. 1, pp. 42–49, 2016, doi: 10.1111/ijlh.12426.

[21] Y. Wang, L. A. Kung, S. Gupta, and S. Ozdemir, “Leveraging Big Data Analytics to Improve Quality of Care in Healthcare Organizations: A

Configurational Perspective," *Br. J. Manag.*, vol. 30, no. 2, pp. 362–388, 2019, doi: 10.1111/1467-8551.12332.
[22] A. A. Vieira, L. M. Dias, M. Y. Santos, G. A. Pereira, and J. A. Oliveira, "On the use of simulation as a Big Data semantic validator for supply chain management," *Simul. Model. Pract. Theory*, vol. 98, no. April 2019, p. 101985, 2020, doi: 10.1016/j.simpat.2019.101985.
[23] J. Liu, M. Chen, and H. Liu, "The role of big data analytics in enabling green supply chain management : a literature review," 2020.
[24] B. Roßmann, A. Canzaniello, H. von der Gracht, and E. Hartmann, "The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study," *Technol. Forecast. Soc. Change*, vol. 130, no. September 2017, pp. 135–149, 2018, doi: 10.1016/j.techfore.2017.10.005.
[25] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine Learning with Big Data: Challenges and Approaches," *IEEE Access*, vol. 5, pp. 7776–7797, 2017, doi: 10.1109/ACCESS.2017.2696365.
[26] C. Ma, H. H. Zhang, and X. Wang, "Machine learning for Big Data analytics in plants," *Trends Plant Sci.*, vol. 19, no. 12, pp. 798–808, 2014, doi: 10.1016/j.tplants.2014.08.004.
[27] H. Chen, R. H.L.Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data To Big Impact," *MIS Q.*, vol. 36, no. 4, pp. 1165–1188, 2018.
[28] N. Ali *et al.*, "Modelling supply chain information collaboration empowered with machine learning technique," *Intell. Autom. Soft Comput.*, vol. 30, no. 1, pp. 243–257, 2021, doi: 10.32604/iasc.2021.018983.
[29] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, 2011, doi: 10.1016/j.dss.2010.08.006.
[30] R. Zhao, Y. Liu, N. Zhang, and T. Huang, "SC," *J. Clean. Prod.*, 2016, doi: 10.1016/j.jclepro.2016.03.006.
[31] O. K. Erol and I. Eksin, "A new optimization method: Big Bang-Big Crunch," *Adv. Eng. Softw.*, vol. 37, no. 2, pp. 106–111, 2006, doi: 10.1016/j.advengsoft.2005.04.005.
[32] A. R. Hedar and M. Fukushima, "Tabu Search directed by direct search methods for nonlinear global optimization," *Eur. J. Oper. Res.*, vol. 170, no. 2, pp. 329–349, 2006, doi: 10.1016/j.ejor.2004.05.033.
[33] B. Abdollahzadeh, F. Soleimanian Gharehchopogh, and S. Mirjalili, "Artificial gorilla troops optimizer: A new nature-inspired metaheuristic algorithm for global optimization problems," *Int. J. Intell. Syst.*, vol. 36, no. 10, pp. 5887–5958, 2021, doi: 10.1002/int.22535.
[34] S. Y. Lee and R. D. Klassen, "Drivers and enablers that foster environmental management capabilities in small- and medium-sized suppliers in supply chains," *Prod. Oper. Manag.*, vol. 17, no. 6, pp. 573–586, 2008, doi: 10.3401/poms.1080.0063.
[35] A. N. El-Kassar and S. K. Singh, "Green innovation and organizational

performance: The influence of big data and the moderating role of management commitment and HR practices," *Technol. Forecast. Soc. Change*, vol. 144, no. December, pp. 483–498, 2019, doi: 10.1016/j.techfore.2017.12.016.
[36] J. S. Chou, N. T. Ngo, W. K. Chong, and G. E. Gibson, *Big data analytics and cloud computing for sustainable building energy efficiency*. Elsevier Ltd, 2016.
[37] J. Liu *et al.*, "A case study of MapReduce-based expressway traffic data analysis and service system," *Int. J. Internet Manuf. Serv.*, vol. 7, no. 4, pp. 278–289, 2020, doi: 10.1504/IJIMS.2020.110233.
[38] N. Wang, F. Chen, B. Yu, and Y. Qin, "Segmentation of large-scale remotely sensed images on a Spark platform: A strategy for handling massive image tiles with the MapReduce model," *ISPRS J. Photogramm. Remote Sens.*, vol. 162, no. 9, pp. 137–147, 2020, doi: 10.1016/j.isprsjprs.2020.02.012.
[39] J. P. Verma, B. Patel, and A. Patel, "Big data analysis: Recommendation system with hadoop framework," *Proc. - 2015 IEEE Int. Conf. Comput. Intell. Commun. Technol. CICT 2015*, pp. 92–97, 2015, doi: 10.1109/CICT.2015.86.
[40] Y. Zhang, C. Ordonez, and W. Cabrera, "Big Data Analytics Integrating a Parallel Columnar DBMS and the R Language," *Proc. - 2016 16th IEEE/ACM Int. Symp. Clust. Cloud, Grid Comput. CCGrid 2016*, pp. 627–630, 2016, doi: 10.1109/CCGrid.2016.94.
[41] A. Tahsin and M. Manzurul Hasan, "Big data & data science: A descriptive research on big data evolution and a proposed combined platform by integrating R and Python on Hadoop for big data analytics and visualization," *ACM Int. Conf. Proceeding Ser.*, pp. 4–5, 2020, doi: 10.1145/3377049.3377051.
[42] J. Álvarez Cid-Fuentes, P. Álvarez, R. Amela, K. Ishii, R. K. Morizawa, and R. M. Badia, "Efficient development of high performance data analytics in Python," *Futur. Gener. Comput. Syst.*, vol. 111, no. xxxx, pp. 570–581, 2020, doi: 10.1016/j.future.2019.09.051.
[43] C. S. Amemba, P. G. Nyaboke, A. Osoro, and N. Mburu, "Elements of Green Supply Chain Management," vol. 5, no. 12, pp. 51–61, 2013.
[44] C. W. Y. Wong, K. H. Lai, K. C. Shang, C. S. Lu, and T. K. P. Leung, "Green operations and the moderating role of environmental management capability of suppliers on manufacturing firm performance," *Int. J. Prod. Econ.*, vol. 140, no. 1, pp. 283–294, 2012, doi: 10.1016/j.ijpe.2011.08.031.
[45] Y. Liu, Y. Zhang, L. Batista, and K. Rong, "Green operations: What's the role of supply chain flexibility?," *Int. J. Prod. Econ.*, vol. 214, no. March, pp. 30–43, 2019, doi: 10.1016/j.ijpe.2019.03.026.
[46] A. Reyna, C. Martín, J. Chen, E. Soler, and M. Díaz, "On blockchain and its integration with IoT . Challenges and opportunities," vol. 88, no. 2018, pp. 173–190, 2020, doi: 10.1016/j.future.2018.05.046.
[47] R. Geng, S. A. Mansouri, and E. Aktas, "The relationship between green supply chain management and performance: A meta-analysis of empirical evidences in Asian emerging economies," *Int. J. Prod. Econ.*, vol. 183, pp.

245–258, 2017, doi: 10.1016/j.ijpe.2016.10.008.

[48] Y. Zuo, F. Tao, and A. Y. C. Nee, “An Internet of things and cloud-based approach for energy consumption evaluation and analysis for a product,” *Int. J. Comput. Integr. Manuf.*, vol. 31, no. 4–5, pp. 337–348, 2018, doi: 10.1080/0951192X.2017.1285429.

[49] R. Dubey, A. Gunasekaran, T. Papadopoulos, and S. J. Childe, “Green supply chain management enablers: Mixed methods research,” *Sustain. Prod. Consum.*, vol. 4, no. June, pp. 72–88, 2015, doi: 10.1016/j.spc.2015.07.001.

[50] J. Wu, S. Guo, J. Li, and D. Zeng, “Big Data Meet Green Challenges: Greening Big Data,” *IEEE Syst. J.*, vol. 10, no. 3, pp. 873–887, 2016, doi: 10.1109/JSYST.2016.2550538.

[51] R. Dubey, A. Gunasekaran, and A. Chakrabarty, “World-class sustainable manufacturing: Framework and a performance measurement system,” *Int. J. Prod. Res.*, vol. 53, no. 17, pp. 5207–5223, 2015, doi: 10.1080/00207543.2015.1012603.

[52] A. Singh, S. Kumari, H. Malekpoor, and N. Mishra, “Big data cloud computing framework for low carbon supplier selection in the beef supply chain,” *J. Clean. Prod.*, vol. 202, pp. 139–149, 2018, doi: 10.1016/j.jclepro.2018.07.236.

[53] I. S. Doolun, S. G. Ponnambalam, N. Subramanian, and G. Kanagaraj, “Data driven hybrid evolutionary analytical approach for multi objective location allocation decisions: Automotive green supply chain empirical evidence,” *Comput. Oper. Res.*, vol. 98, pp. 265–283, 2018, doi: 10.1016/j.cor.2018.01.008.

[54] R. D. Raut, S. K. Mangla, V. S. Narwane, B. B. Gardas, P. Priyadarshinee, and B. E. Narkhede, “Linking big data analytics and operational sustainability practices for sustainable business management,” *J. Clean. Prod.*, vol. 224, pp. 10–24, 2019, doi: 10.1016/j.jclepro.2019.03.181.

[55] J. J. H. Liou, Y. C. Chuang, E. K. Zavadskas, and G. H. Tzeng, “Data-driven hybrid multiple attribute decision-making model for green supplier evaluation and performance improvement,” *J. Clean. Prod.*, vol. 241, p. 118321, 2019, doi: 10.1016/j.jclepro.2019.118321.

[56] J. Liu, Y. Feng, Q. Zhu, and J. Sarkis, “Green supply chain management and the circular economy: Reviewing theory for advancement of both fields,” *Int. J. Phys. Distrib. Logist. Manag.*, vol. 48, no. 8, pp. 794–817, 2018, doi: 10.1108/IJPDLM-01-2017-0049.

[57] R. Dekker, J. Bloemhof, and I. Mallidis, “Operations Research for green logistics - An overview of aspects, issues, contributions and challenges,” *Eur. J. Oper. Res.*, vol. 219, no. 3, pp. 671–679, 2012, doi: 10.1016/j.ejor.2011.11.010.

[58] T. E. Evtodieva, D. V. Chernova, N. V. Ivanova, and J. Wirth, “The internet of things: Possibilities of application in intelligent supply chain management,” *Adv. Intell. Syst. Comput.*, vol. 908, pp. 395–403, 2020, doi: 10.1007/978-3-030-11367-4_38.

[59] C. Meng, Y. Wang, X. Zhang, A. Mandal, W. Zhong, and P. Ma, “Effective

Statistical Methods for Big Data Analytics," pp. 280–299, 2017, doi: 10.4018/978-1-5225-2498-4.ch014.
[60] C. Wang, M. H. Chen, E. Schifano, J. Wu, and J. Yan, "Statistical methods and computing for big data," *Statistics and its Interface*, vol. 9, no. 4. pp. 399–414, 2016, doi: 10.4310/SII.2016.v9.n4.a1.
[61] V. Zelenyuk, "Aggregation of inputs and outputs prior to Data Envelopment Analysis under big data," *Eur. J. Oper. Res.*, vol. 282, no. 1, pp. 172–187, 2020, doi: 10.1016/j.ejor.2019.08.007.
[62] C. Wang, M. H. Chen, J. Wu, J. Yan, Y. Zhang, and E. Schifano, "Online updating method with new variables for big data streams," *Canadian Journal of Statistics*, vol. 46, no. 1. pp. 123–146, 2018, doi: 10.1002/cjs.11330.
[63] S. Tonidandel, E. B. King, and J. M. Cortina, "Big Data Methods: Leveraging Modern Data Analytic Techniques to Build Organizational Science," *Organ. Res. Methods*, vol. 21, no. 3, pp. 525–547, 2018, doi: 10.1177/1094428116677299.
[64] C. Wang, Q. Zhang, and W. Zhang, "Corporate social responsibility, Green supply chain management and firm performance: The moderating role of big-data analytics capability," *Res. Transp. Bus. Manag.*, vol. 37, no. September, p. 100557, 2020, doi: 10.1016/j.rtbm.2020.100557.
[65] Y. Zhang, S. Ren, Y. Liu, and S. Si, "A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products," *J. Clean. Prod.*, vol. 142, pp. 626–641, 2017, doi: 10.1016/j.jclepro.2016.07.123.
[66] M. L. Tseng, M. S. Islam, N. Karia, F. A. Fauzi, and S. Afrin, "A literature review on green supply chain management: Trends and future challenges," *Resour. Conserv. Recycl.*, vol. 141, no. June 2018, pp. 145–162, 2019, doi: 10.1016/j.resconrec.2018.10.009.
[67] S. Luthra, D. Garg, and A. Haleem, "Empirical Analysis of Green Supply Chain Management Practices in Indian Automobile Industry," *J. Inst. Eng. Ser. C*, vol. 95, no. 2, pp. 119–126, 2014, doi: 10.1007/s40032-014-0112-6.
[68] J. Kim and J. Rhee, "An empirical study on the impact of critical success factors on the balanced scorecard performance in Korean Green supply chain management enterprises," *Int. J. Prod. Res.*, vol. 50, no. 9, pp. 2465–2483, 2012, doi: 10.1080/00207543.2011.581009.
[69] F. Testa and F. Iraldo, "Shadows and lights of GSCM (green supply chain management): Determinants and effects of these practices based on a multi-national study," *J. Clean. Prod.*, vol. 18, no. 10–11, pp. 953–962, 2010, doi: 10.1016/j.jclepro.2010.03.005.
[70] S. Shohan, S. M. Ali, G. Kabir, S. K. K. Ahmed, S. A. Suhi, and T. Haque, "Green supply chain management in the chemical industry: structural framework of drivers," *Int. J. Sustain. Dev. World Ecol.*, vol. 26, no. 8, pp. 752–768, 2019, doi: 10.1080/13504509.2019.1674406.

□