



**HAL**  
open science

# Regression Predictive Model to Analyze Big Data Analytics in Supply Chain Management

Elena Puica

► **To cite this version:**

Elena Puica. Regression Predictive Model to Analyze Big Data Analytics in Supply Chain Management. 17th IFIP International Conference on Artificial Intelligence Applications and Innovations (AIAI), Jun 2021, Hersonissos, Crete, Greece. pp.94-101, 10.1007/978-3-030-79150-6\_8. hal-03287687

**HAL Id: hal-03287687**

**<https://inria.hal.science/hal-03287687>**

Submitted on 15 Jul 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Regression Predictive Model to analyze Big Data Analytics in Supply Chain Management

Elena Puica

<sup>1</sup> Economic Informatics Doctoral School, The Bucharest University of Economic Studies,  
Bucharest, Romania  
elena.puica@gmail.com

**Abstract.** The research problem that is the interest in this thesis is to understand the Big Data Analytics (BDA) potential in achieving a much better Supply Chain Management (SCM). Based on this premise, it was conducted a Regression Predictive Model to comprehend the usage of Big Data Analytics in SCM and to have insights of the requirements for the potential applications of BDA. In this study were analyzed the main sources of BDA utilized in present by Supply Chain professionals and it was provided future suggestions. The findings of the study suggest that BDA may bring operational and strategic benefit to SCM, and the application of BDA may have positive implication for industry sector.

**Keywords:** Big Data Analytics, Supply Chain, Supply Chain Management, Regression Predictive Model, Big Data in Supply Chain Management

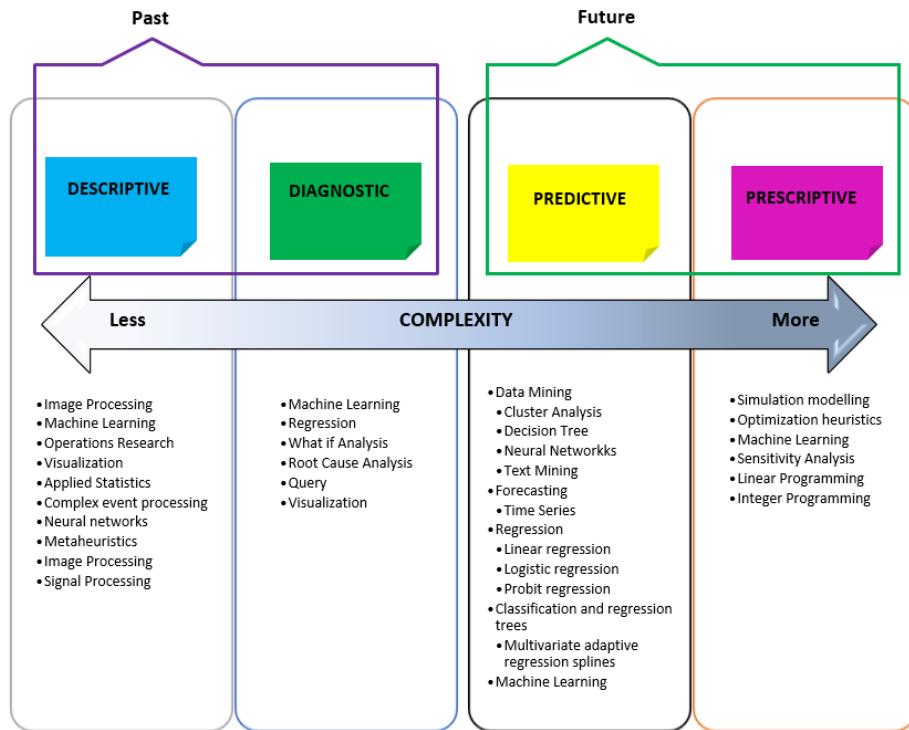
## 1 Introduction

The present study aims to identify the expectations, the requirements, and implications from the application of Big Data Analytics (BDA) in Supply Chain Management (SCM). Growing network complexity, global competition, and increasing product diversity, while customer expectation remains as high as ever [1] has directed the SCM in the direction of development. The research contributes to the existing literature by a regression predictive model to discover the most common used sources of BDA in SCM. This way, the research adds to the literature the implication of BDA in SCM by applying a regression predictive model to discover what is the principal usage of BDA and to understand what are the less ways that BDA is used in SCM. The research can lead to inspire the SCM professionals to take proactive actions and to discover opportunities in bettering their processes, which can lead to a high positive impact in the industry sector.

The remainder of this paper is organized as follows: Section two, reviews relevant BDA literature and the studies related to the practices and techniques of BDA used. Section three is concentrated to provide a better understanding of the implication of BDA in SCM. Section four presents the model applied to analyze the use of BDA in SCM and presents the results of regression predictive model analyses. The last section is composed of the conclusion of the study and provides direction for future research.

## 2 Big Data Analytics

Data is expressed in different types and formats, and the access to data is also different, these facts all point the issue in one direction: the ability to search, aggregate, visualize and cross-reference large data sets in a reasonable time and when BDA is linked to SCM, new challenges arise. BDA, generally, supports SCM in innovation, productivity, and competition [2] and has been defined as the technique that is deployed to uncover hidden patterns and bring insight into interesting relations in understanding contexts by examining, processing, discovering, and exhibiting the result [3]. The four main types of analytics techniques will be discussed: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics.



**Fig.1.** Types and techniques of Big Data Analytics

Predictive and prescriptive analytics play a vital role in helping SCM to make effective decisions about the strategic direction. BDA is driving the SCM for development, and it uses models, technologies, and tools to help SCM make performance analysis that is fast, efficient, and effective. (see Fig.1.)

Descriptive analytics summarize the raw data and is a data analysis used to describe past situations to make trends, patterns, and anomalies visible [4]. While descriptive analytics describes what has happened, diagnostic analytics attempts to get at the root

cause of some anomaly or occurrence. Predictive analysis examines real-time and historical data, using mathematical algorithms and programming to discover interpretation and prediction patterns within the data, in other words, it makes predictions in the form of probabilities for the future events [5]. Predictive analytics uses advanced statistics, machine learning techniques and data-driven algorithms to generate models and fall into two major categories: regression techniques and machine learning techniques. Prescriptive analysis mainly is determining and evaluating many complex objectives and alternative decisions. It mainly uses data and mathematical algorithms to achieve the purpose of improving business performance and includes multi-criteria decision making, optimization, and simulation. BDA will make the organization's system generates a large amount of information [6] and identifying problems and opportunities in existing processes and functions, can be achieved by using BDA, which generates competitive advantage for the SCM [7].

### **3 Big Data Analytics in Supply Chain Management**

Digital technologies can lead to a better understand for a more complex SCM, which can continuously monitor the physical environment and generate large amounts of data at an unprecedented rate. These technologies are generating large amounts of data, called Big Data [8], which gradually becomes an important information technology regarding SCM decisions [9]. BDA should not only be used to help SCM to make strategic decisions in procurement, supply chain network design, and product design and development, but it should also be applied to all stages of the entire SCM [10], in order to industry sector to benefit from it. Predictive analytics can have the ability to predict consumer behaviors, prevent fraud, mitigate risk, identify new customers, and improve operations, being able to identify customer's spending behaviors and cross-sell efficiently or sell additional products to their customers, enhance customer satisfaction and customer loyalty, identify the most effective marketing campaign and communication channels, identify fraudulent payment transactions, flag potential fraudulent claims and pay legitimate insurance claims immediately and predict when machinery will fail. SCM professionals are overwhelmed by massive amounts of data, on one side, opens new ways to generate, organize, and analyze data, and on the other side, is pushing SCM to adopt and improve BDA functions to enhance SCM processes and ultimately improve performance[12].

Predictive analytics can help SCM mitigate their risk, maximize their profits, optimize their operations, and gain a competitive advantage. SCM use predictive analytics to solve complex business problems for different sectors (construction, manufacturing, retail, transportation, telecommunication and utilities), in order to discover new opportunities, and to gain competitive advantages. The data utilized in SCM is extracted from different sources, these sources are:

- a. Analyse big data from geolocation of portable devices;
- b. Analyse big data from smart devices or sensors;
- c. Analyse big data generated from social media;
- d. Analyse big data internally from any data source.

Applications of BDA captures customer demand, micro-segmentation and predicts consumers' behavior, is minimizing transportation costs, in the process of helping the products to be transported through the supply chain. In operations applications BDA can optimize labour, track attendance, and reduce costs while ensuring service. Sourcing applications, use BDA to optimize procurement channel selection and integrate suppliers into their operations, these data sources include expenses, supplier performance assessments, and internal or external negotiations [13].



**Fig. 2.** Percentage of enterprises that use Big Data

Source: own processing based on <https://ec.europa.eu/eurostat>

BDA technology in the industry sectors was used by 15% of enterprises in the construction industry, 22.83% in manufacturing, 23.5% in retail, 36% in the telecommunications industry, 37% in the transport industry and 24% in the utilities industry for the category that analyse Big Data from the geolocation of portable devices. For category that used Big Data from social media, 7% was in the construction, 12.17% in manufacturing, 13% in retail, 26% in telecommunications, 13% in the transportation and 10% in utilities. The enterprises that used Big Data internally from their own or external data sources had a percentage of 7% in construction, 16% in manufacturing, 9.5% in retail, 19% in telecommunications, 15% in transportation and 11% in utilities. The companies that used Big Data from their own smart devices or sensors had a percentage of 2% in construction, 14.08% in manufacturing, 7% in retail, 20% in telecommunications, 10% in transportation and 9% in utilities. (see Fig. 2.)

## 4 Implementation of Regression Predictive Model with SAP Analytics Cloud

Regression analysis examines the degree of relationship that exists between a set of input variables and a target variable. The relationships between the target variable and the input variables are associative only and any cause-effect is merely subjective. To analyze the usage of BDA in SCM for different industry sectors (construction, manufacturing, retail, transportation, telecommunication, and utilities) it was used the application of SAP Analytic Cloud (SAC), which is an analytic software provided by SAP. SAC is a platform independent and allows to discover, analyze, plan, and predict data. SAC offers connection to a variety of data sources to create models and develop reports with charts, including Geo Maps, and tables.

#### **4.1 Identification of the business problem**

The regression predictive model it is applied in this study to discover what is exactly the usage of BDA among SCM for different industry sectors (construction, manufacturing, retail, transportation, telecommunication, and utilities).

#### **4.2 Definition of the hypotheses**

The purpose of the hypotheses is to narrow down the business problem and make predictions about the relationships between two or more data variables. By applying the regression predictive model, the following objectives are intended to be met, the first one is to find out what is the principal usage of BDA in SCM, the second is to understand what are the less applied sources of usage of BDA in SCM, the third, is to know from where to start, in order to take action to discover opportunities for those sources where BDA is used less and the last one is to encourage SCM professionals to take proactive actions to use BDA in SCM. The target variable is the “percentage of Big Data usage”, which represents the event to be predicted.

The following hypotheses were analyzed:

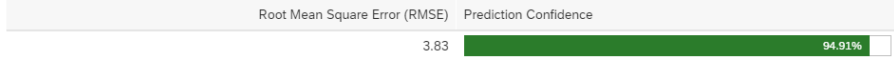
- H1 - Analyze big data from geolocation of portable devices.
- H2 - Analyze big data from smart devices or sensors.
- H3 - Analyze big data generated from social media.
- H4 - Analyze big data internally from any data source.

#### **4.3 Collecting the data**

Data collected is classified as structured (spreadsheets) and it was extracted from the Eurostat database. Eurostat is the statistical office of the European Union, is a world-leading database, widely known for its extensive, reliable content and high-quality statistics and data on Europe.[14]

#### **4.4 Data analysis, development of the predictive model and the determination of the best-fit model**

To analyze the quality of the model more indicators were considered. The Prediction Confidence, which measures the robustness of the predictive model or its ability to reproduce the same detection on new data and has the role to measure if the predictive model can do the predictions with the same reliability when new cases arrive. The Prediction Confidence should be as close as possible to 100%. The quality of the regression model is measured by The Root Mean Square Error (RMSE). It is a statistical indicator which measures the average of the square difference between values predicted by the predictive model and actual values of the target for all cases of the validation dataset. The smaller is this difference, the better the quality of the predictive model is.



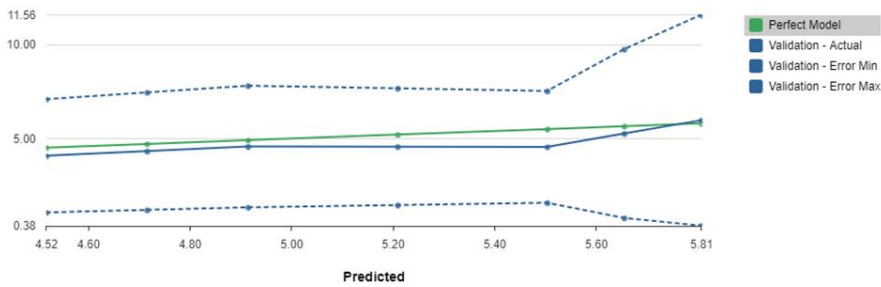
**Fig. 4.** Performance Indicators

The model that was trained for this study has the result of RMSE equal with 3.83 and the prediction confidence indicator is 94,91%. (see Fig. 4.)

Data Partition	Minimum	Maximum	Mean	Standard Deviation
Training	0	23	5.2	4.95
Validation	0	22	4.75	3.85

**Fig. 5.** Target Statistics

The mean of the target is around 5,2 for the estimation results and 4,75 for the validation dataset. The standard deviation is around 4,95 for the estimated results and 3,85 for the validated results (see Fig. 5.). The predictive model has a good quality, its robustness is around 95%, and thus is also very good.

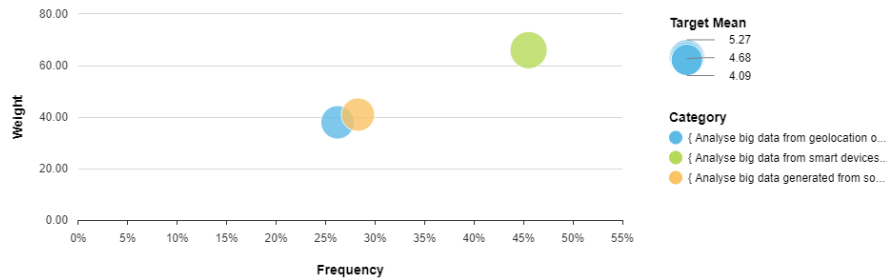


**Fig. 6.** Predicted vs. actuals data.

In Fig. 6. is described the predicted versus actual data, the green curve represents the perfect model, which shows no error and predicts exactly the correct opportunity value. The blue curve is the predictive model determined by SAC Smart Predict. The dotted-blue curves are the error min and error max on the validation dataset. If the green and blue curves do not match at all, this mean that the quality and the robustness of the predictive model are quite poor. If these two curves match closely, the predictive model is good and can be trusted to predict the value of the unknown target. Last case is when the two curves match a lot except on few segments. This means that the predictive model is good but can be improved.

**4.5 Utilize the model, referred to as scoring.**

Following the choice of the best variables to build a qualitative and robust model, a predictive regression model was determined.



**Fig. 7.** Usage of Big Data Analysis in SCM

In Fig. 7. Is the representation of the validated data, the category that has the lowest value of the target mean is represented in the graphic by orange dot, the variable „Analyze Big Data from social media”. The lowest values of frequency are represented by the blue dot, the variable “Analyze Big Data from geolocation of portable devices”. The highest value of target mean is for the “Analyze Big Data from devices or sensors” and “Analyze Big Data internally” which are the most frequent variables determined in the model. (see Fig.7.)

Influencer	Category	Frequency	Target Mean	Target Standard Deviation
Big Data Usage	{ Analyse big data from geolocation of portable devices }	26.21%	4.58	3.22
Big Data Usage	{ Analyse big data from smart devices or sensors, Analyse big data internally from any data source }	45.52%	5.27	4.53
Big Data Usage	{ Analyse big data generated from social media }	28.28%	4.09	3

**Fig. 8.** Explanation of category validated statistics.

In Fig. 8. is the representation of statistics explanation of variable validated. For example, the category that contains the variables “Analyze Big Data from devices or sensors” and “Analyze Big Data internally from any data sources” have the value 45,52%, which indicates that these are the most common ways BDA is used in SCM. The second used is the variable „Analyze Big Data generated from social media”, which represents 28,28% and last source of BDA used in SCM is the variable “Analyze Big Data from geolocation of portable devices”, which is 26,21%. (see Fig. 8.)

## Conclusion

The advancements in the sphere of artificial intelligence and machine learning have revolutionized the world of computation. This article initiates a possibility of development of new models of regression predictive analytics and add additional highlights to literature review of the existing models. The Regression predictive model applied in this study is providing insights for the objective proposed, which is to find out what is the principal usage of BDA in SCM for different industry sectors (construction, manufacturing, retail, transportation, telecommunication, and utilities), to understand what are the less used sources of BDA in SCM, to know from where to start, in order



to act and to discover opportunities for the less used BDA ways of usage and lastly, to encourage the SCM professionals to take proactive actions to use BDA in SCM. According to the results of this study, the most popular ways or used sources of BDA in SCM is to analyze Big Data from devices or sensors and to analyze Big Data internally from any data sources. Analysis of Big Data from geolocation of portable devices is the least used source in SCM.

For SCM to have a positive impact and change is to invest in portable and geolocation devices, because the data that is generated from these devices is valuable and voluminous, which could help the SCM to make improved decisions and have better control over their process. In this paper, it is presented ongoing research work about BDA solutions for increasing the SCM visibility. The future prospect is that several SCM enterprises from different industries will develop Big Data ecosystems for achieving new business models and offering new services to customers and will lead to even a more increased complexity of SCM.

## References

1. Dawe, Peter, Alicia Pittman, and Elfrun von Koeller, "Segmentation in the Consumer Supply Chain: One Size Does Not Fit All.", Technical Report. The Boston Consulting Group. (2015).
2. Dong J, Yang C., Business value of big data analytics: A systems-theoretic approach and empirical test. In: Information & Management. (2018).  
<https://doi.org/10.1016/j.im.2018.11.001>
3. Wang H et al., Randomly attracted firefly algorithm with neighborhood search and dynamic parameter adjustment mechanism. Journal of Soft Computing.; 21(18):5325-5339. (2017)
4. Ansari, Z. and Kant, R., A State-Of-Art Literature Review Reflecting 15 Years of Focus on Sustainable Supply Chain Management. Journal of Cleaner Production, [e-journal] pp. 2524-2543. (2016). DOI:<http://dx.doi.org/10.1016/j.jclepro.2016.11.023>
5. Arunachalam, D., Kumar, N. and Kawalek, J., Understanding Big Data Analytics Capabilities In Supply Chain Management: Unravelling The Issues, Challenges And Implications For Practice. Transportation Research Part E, [e-journal] pp.416-436.v (2017). DOI: <http://dx.doi.org/10.1016/j.tre.2017.04.001>
6. Barbosa, M., Vicente, A., Ladeira, M. and Oliveira, M., Managing supply chain resources with Big Data Analytics: a systematic review. International Journal of Logistics Research and Applications, [e-journal] 21(3), pp.177-200. (2018). DOI: <https://doi.org/10.1080/13675567.2017.1369501>
7. Smart Village Technology. Modeling and Optimization in Science and Technologies. Cham: Springer. vol 17, (2020).
8. Ahearn, M., Armbruster, W. and Young, R., Big Data's potential to improve food supply environment sustainability and food safety. International Food Agribusiness Management Review. (e-journal) 19, pp. 177 – 172 (2016). DOI:<http://dx.doi.org/10.22004/ag.econ.240704>
9. Bronson, K. and Knezevic, I., Big Data in food and agriculture. Big Data & Society, [e-journal] 3(1), p.205395171664817. (2016). DOI: <https://doi.org/10.1177/2053951716648174>
10. Hazen, B.T., Skipper, J.B., Ezell, J.D. and Boone, C.A., Big data and predictive analytics for supply chain sustainability: A theory-driven research agenda. Computers & Industrial Engineering, 101, pp.592-598. (2016).

11. Hazen, B.T., Boone, C.A., Ezell, J.D. and Jones-Farmer, L.A., 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. *International Journal of Production Economics*, [e-journal] 154, pp.72-80. (2014). DOI:<https://doi.org/10.1016/j.ijpe.2014.04.018>.
12. Addo-Tenkorang, R. and Helo, P., Big Data Applications In Operations/Supply-Chain Management: A Literature Review. *Computers & Industrial Engineering*, [e-journal] pp.528-543. (2016). DOI: <http://dx.doi.org/10.1016/j.cie.2016.09.023>
13. Eurostat, (accessed at 24.04.2021), <https://ec.europa.eu/eurostat>