International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319-6734, ISSN (Print): 2319-6726
www.ijesi.org ||Volume 11 Issue 10 October 2022 || PP 23-28

Introduction Methodology of Big Data and Machine Learning According To Rami Architecture: An Empirical Study

Anderson Brito de Lima
Federal Institute of Education, Science and Technology of Amazonas
Manaus Industrial District Campus

Hiparco Leitão Vieira
Federal Institute of Education, Science and Technology of Amazonas
Manaus Industrial District Campus

Nivaldo Rodrigues e Silva
Federal Institute of Education, Science and Technology of Amazonas
Manaus Industrial District Campus

Maria Cassiana Andrade Braga
Federal Institute of Education, Science and Technology of Amazonas
Manaus Industrial District Campus

Samirames da Silva Fleury
Federal Institute of Education, Science and Technology of Amazonas
Manaus Industrial District Campus

Roberto Alcides de Lima Prazeres
Federal Institute of Education, Science and Technology of Amazonas
Manaus Industrial District Campus

Daniel Nascimento-e-Silva
Federal Institute of Education, Science and Technology of Amazonas
Manaus Industrial District Campus

ABSTRACT
With the ascendancy of technology in the scope of data collection, treatment and analysis, there is an opportunity to build a method that envisions a path for introducing big data and machine learning in the electronics industry. In this sense, this study proposesintroducing big data and machine learning in the electronics industry based on the Rami architecture guidelines. Its purpose is to apply and demonstrate the analysis results to meet the industrial demands regarding the prediction of the number of defects from the analysis of a database. The method used was the scientific-technological one, composed of eight stages, subdivided into four of a scientific nature and four of a technological nature. The scientific questions generate the answers that are handled in the methodological steps. The results indicated the possibility of introducing machine learning and big data in six stages, starting with data collection, then data cleaning, data mining, content analysis, data visualization, and ending with data integration. The conclusion shows that these steps are in line with the ins and outs of the Rami 4.0 Architecture, which interconnects data collection with learning from countless other spheres of production processes.

Keywords: *Big data; Machine learning; Rami 4.0; Scientific-technological method; Technological methodology.*

Date of Submission: 10-10-2022 Date of Acceptance: 23-10-2022

I. INTRODUCTION

With the ascendancy of technology in the scope of data collection, treatment and analysis, there is an opportunity to build a method that envisions a path for introducing big data and machine learning in the electronics industry. It is a fact that the sector must update itself to remain competitive in capitalism; however, seeing a simple way to do this is not always so evident. The study was carried out based on the reality of a Brazilian company located in Amazon that is on the ascendancy in this market and has just acquired a factory with all the human capital of a multinational. Its physical structure and employees are the same, but all the systematics and technology of an international do not come in the package. In this context, it is not easy to acquire competitiveness in the short term.

Unpredictability cannot be predicted, or there are not enough resources. The unpredictability of defects occurs in different ways. They can be generated by a poorly adjusted machine or an inattentive operator to the poor quality of electronic components, which can unexpectedly increase the number of defects. The defect of the electronics industrial process is the rejection of the unfinished or finished product due to failure in the automatic assembly process. In turn, the amount of non-conforming product occurrence can be recorded daily and analyzed over a given period.

To get an idea, imagine a printed circuit board where a few hundred electronic components are mounted. If an accident occurs somewhere in your production process, in which one of these components ends up being damaged by hitting a loose part of a belt, the functional tests of the board cannot detect it. The defect goes through several processes, and only in the final operation would it be seen. In a shallow analysis, all we could do is eliminate all possible causes but never state the actual root cause or in which process this defect occurred. It is one of several cases we could cite here, but that would make this job very dull, which is beside the point.

The RAMI architecture from big data and machine learning can help to avoid these defects in several ways. However, the area addressed in this work is the one that demands quantitative data in a structured way to make predictions with machine learning algorithms to add as an indicator in decision-making.

In this sense, this study proposesintroducing big data and machine learning in the electronics industry based on the Rami architecture guidelines. This introduction aims to apply and demonstrate the analysis results in the form of predictions to meet the industrial demands regarding the prediction of the number of defects from the analysis of a database. RAMI, in turn, is architected in three dimensions: life-cycle and value stream, hierarchy levels, and layers. The scientific-technological method proposed by Nascimento-e-Silva (2020; 2021a; 2021b. 2021c) in its two natures, scientific and technological, was used for elaborating this methodology.

A significant benefit that encompasses the power of big data and machine learning is prediction through structured quantitative data. Structured data is data that is organized in a logical structure. In general, they are categorized as quantitative. This way, management can decide on the allocation of labor where it needs it. Prediction is a proposal that can be made through machine learning algorithms. With this, among other things, we can, for example, hire human capital for a specific demand in a future scenario based on machine learning algorithms.

II. BIG DATA AND MACHINE LEARNING

Big data and machine learning are two tools with different forms of application in practically all the intricacies of production processes. Contemporarily, its application extends to many other fields of associated human life, which are beyond the concern of this study. What has been seen, for example, is that knowledge and new scientific discoveries about these two phenomena are constantly increasing and improving every day. Here, however, the intention is only to demonstrate their conceptual scopes, which represent, in practice, the limits of scientific knowledge about each of them, all elaborated from scientific studies published in the year 2022.

2.1 Big Data

Data are values that are attributed to something. These values can be extracted for measurement or measurement purposes. The data can be studied in a multidisciplinary way, since they can be extracted from different branches of science. Data from people all over the world can be examples typified as big data. According to science, big data can be defined as data, as shown by studies by Srividya and Tripathy (2022), Allaymoun et al. (2023), Sharma, Agarwal, and Arya (2022), and Redieki and Widvarto (2022), as shown in the definitions presented in table 1. What differentiates big data from all other types of data is that big data is extensive, very many, and extremely large, as shown by the data in Table 1.

Table 1. What is big data?

References	What is big data?
Srividya& Tripathy (2022)	Big data can be defined as data with a large volume whose growth is exponential in nature.
Allaymoun et al. (2023)	Big Data can be defined as large amounts of widely diverse data that are generated, processed, and collected at a high rate.
Hongmei (2022)	Big Data can be defined as vast sets of unstructured data coming in various forms gathered from different sources at such a rapid rate that it is far beyond the processing power of a traditional server.
Sharma, Agarwal& Arya (2022)	Big Data can be defined as huge sets of unstructured data coming in various forms gathered from different sources at such a rapid rate that it is far beyond the processing power of a traditional server.
Koman et al. (2022)	Big Data can be defined as a huge amount of data generated by various human activities
Redjeki& Widyarto (2022).	Big Data can be defined as datasets that are so large that traditional technologies and tools for extracting them are no longer usable within a reasonable timeframe and are ineffective in terms of cost.

Source: data collected by the authors.

Big data is also cataloged by science as a set of data, as seen in the studies by Hongmei (2022) and Koman et al. (2022). A huge is a series of elements with at least one characteristic in common; that is, they have at least one thing in common with all the other components. It is the case that Monday and Tuesday are part of the set of days of the week. What they have in common is that they are days of the week. In the case of big data, the dataset gathers properties of the object of study. The dataset can be housed and organized to facilitate interpretation and analysis, as distributed in columns and rows. One way to visualize an example of a dataset is when they are arranged in an Excel spreadsheet. The difference between a common dataset and big data is that big data is an extensive set. For example, a standard setting would be the marital status, age, income, and other characteristics of 100 people who took a vocational training course.A big data, for example, would be composed of each measure of electricity that enters per second in each residence of a metropolis of one million inhabitants in a horizon of ten years of measurement, which leads to astronomical numbers in the billions, trillions, and even quadrillions of data.

2.2 Machine Learning

Science presents at least five ways of looking at machine learning or machine learning. Machine learning can be defined as a process, as shown in the study by Abou-Zamzam (2022) in table 2. A process denotes the continuous sequencing of activity, with a well-determined beginning, middle, and end. In addition to presenting a sequential order, the processes mustproduce a product or service when the last step is completed. Machine learning is a process because it can also be highlighted as a continuous action; it is maintained with a certain regularity to end with learning something. The algorithms must be programmed to perform activitiesclearly and adequately to reach the intended result.

Another way of looking at machine learning is as a set of algorithms, as in the study by Muthiah et al. (2022). This set of algorithms is a step-by-step, logical sequence of instructions so that the computer can adequately execute each of them and result in a partial or total result. A result is partial when it is used to fuel an even larger one. For example, the age calculation might yield the average age as a partial result. In contrast, a global result might be the instruction to use the age average to allow or not allow a user access to certain content in an application.

Machine learning can be defined as a category of artificial intelligence (Baig,Syed&Mohammad, 2022). Categorizing is classifying and organizing things into groups with the same classes. For example, when we organize our bedroom, we put clothes in one place, shoes in another, and bedding materials in a third space. In practice, when we do this, we categorize these materials into clothing, shoes, and bedding. Machine learning is a category of artificial intelligence because there are other ways of applying artificial intelligence than machine learning does. The main characteristic of machine learning as a category is the responsibility for "training" and modeling machine learning through mathematical resources and procedures.

Table 2. What is machine learning?

References	Whatis machine learning?
Abou-Zamzam (2022)	Machine learning can be defined as the process by which a computer is able to improve its own performance (as in image file analysis) by continually incorporating new data into an existing statistical model.
Muthiah et al. (2022)	Machine learning can be defined as a set of algorithms or programs that work best with "experience."
Baig, Syed& Mohammad (2022)	Machine learning can be defined as a category of Artificial Intelligence, where the notion of mathematical modeling of data is adopted to train the machine learning classifier.
Bhati, Chugh& Bhati (2022)	Machine learning can be defined as the ability to apply some techniques to raw data to make sense of it and gain some knowledge.
Alférez et al. (2022)	Machine learning can be defined as the ability of computers to recognize patterns without being explicitly programmed.
Odabași et al. (2022)	Machine learning can be defined as the practice of using algorithms to learn from data, extract hidden knowledge, or make predictions in a related field, as ML algorithms can transform a dataset into a model.

Source: data collected by the authors.

Table 2 shows that another way of looking at machine learning found in scientific databases is as an application capability, as shown in the study by Bhati, Chugh and Bhati (2022). More specifically, machine learning focuses on applying different ways of transforming raw, seemingly chaotic data into elaborate schemas in which a particular order appears and becomes perceptible. Millions of data from other variables apply these techniques, using a computer and mathematical programming, in a way that human skills and minds would not be able to do. And all this is done at a speed that impresses even those who design the learning schemes for these machines.

The fifth way in the literature is to view machine learning as a practice (Odabaşi et al., 2022). As strange as this may seem, machine learning is practically impossible without preparation because it is through each repetition that mistakes and successes are generated, and learning is continuously improved. What is practiced in machine learning is the algorithm that guides and determines the intricacies of learning, not its content. The machine is, therefore, free to practice learning within a given context. This context is limited, but the content, what will be known, is free. The context is the universe of healthy plants and a diseased plant's characteristics. It is the case, for example, of detecting a diseased plant among thousands of others. Machines are free to learn and determine whether or not a plant is sick, which is the content of learning.

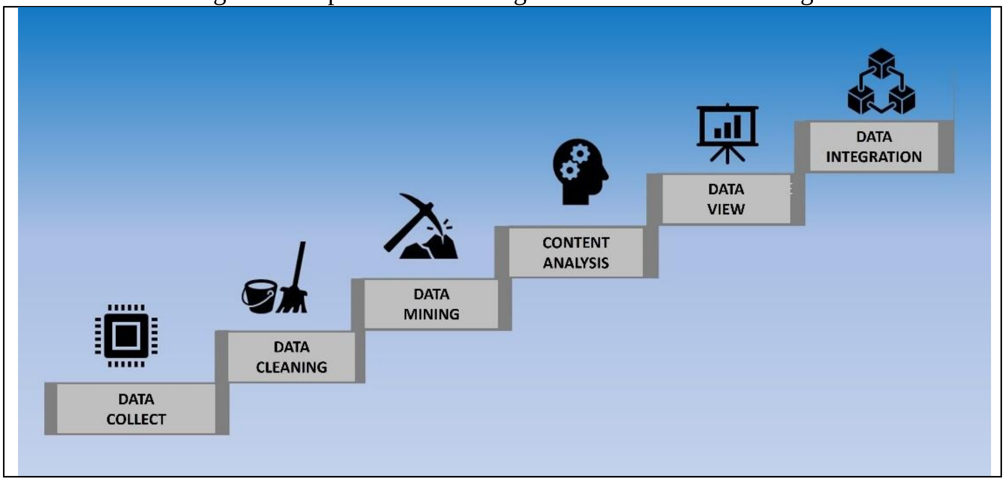
III. STEPS INTRODUCING BIG DATA AND MACHINE LEARNING

The literature review identified the main stages of introducing big data and machine learning in manufacturing production processes. The intention of building these procedures was to guide the practice of these tools in organizational decisions. For this reason, each step was described briefly so that the entire methodology could be seen in a panoramic way.

First Step: Data Collection

The first stage consists of data collection, as can be inferred from the studies by Chinchanikar and Shaikh (2022), Nti et al. (2022), and Gandomi, Chen and Abualigah (2022). Of course, the first step is to develop a strategy that makes it possible to collect the information. For this, it is interesting to define the objective of the action previously. If the intention is to gain knowledge about the behavior of the process, the collection is carried out automatically through sensors and actuators. An example of what can be done in this step is the automatic and real-time capture of the operating parameters of a solder paste reflow oven. This equipment needs its temperature to remain constant at each process step. For this, it would be necessary to install sensors at each stage to collect this data, as show Figure 1.

Figure 1. Steps to introduce big data and machine learning



Source: Prepared by the authors.

Second Step: Data Cleaning

The second step is data cleaning, as can be deduced from the studies by Sharma et al. (2022) and Chahal, Gulia and Gill (2022). This phase is one of the most crucial for the successful use of Big Data. The objective here is to identify and eliminate anomalies that could compromise the efficiency of the process. Data cleaning can be done to highlight the problem we want to solve, segregating one variable or more that, a priori, is suspected of relevance to the issue in question. Cleaning is performed by thorough inspection of the collected information. Using statistical methods, it is possible to recognize deviations and determine their relevance to the analysis as a whole. Thus, null, duplicate or contradictory values are removed from the equation, it is crucial to ensure the legitimacy and effectiveness of the collection, but it can serve other purposes. At this stage, the evaluation of what was collected ends up generating suggestions to improve future activities and enrich the database.

ThirdStep: Data Mining

The third step is data mining, based on studies by Sánchez-Gutiérrez and González-Pérez (2022), Brackenridge et al. (2022), and Hohmann (2022). The name of this step is quite intuitive about what it proposes. Simply put, it is sifting through massive amounts of data to extract consistent patterns over the superhuman amount of information generated daily by Big Data. That is, this characteristic makes it unfeasible for mining to be carried out without the aid of specific programs. To this end, we use learning algorithms that combine concepts of artificial intelligence and machine learning. Once found, the patterns go through a validation process and can finally be considered valuable information. Automating this process also prevents errors caused by human interference. Our search for behavior patterns can corrupt the essence of the step, which is to look at the data in general. Thus, issues that would go unnoticed by us end up being identified by purely statistical analysis.

Fourth Step 4: Content Analysis

The fourth step is content analysis, based on the studies by Emu et al. (2022) and Xing et al. (2022).It is one of the Big Data steps that most depend on a clearly defined strategy. Efficiently analyzing the content of the collected, filtered, and validated information will generate beneficial insights for the enterprise. Therefore, we must emphasize what we discussed in the first topic: knowing your goal. If the idea is to clarify the economic scenario in which the company is inserted, a descriptive analysis is necessary, providing real-time performance data. To anticipate trends and possible future designs, you should strive to develop predictive analytics. Likewise, it is possible to take a prescriptive approach. It aims to clarify the results of actions already taken and, thus, provide insights that help optimize the strategy. Finally, we have the diagnostic analysis, which seeks to contextualize possible failures during some process.

Step 5: Data Visualization

The fifth step is data visualization, based on studies by Bikakis et al. (2022), Yoo, Leung and Nasridinov (2022) and Hohmann (2022). Information Visualization Ensuring an intuitive visualization of data is essential to the enterprise's success. After all, the information gathered was not created during the process. They were already there; they just hadn't been panned and exposed. At this stage, the challenge is to make access even more accessible for everyone involved in the operation.

For this, graphic adaptations are used to eliminate noise and factors that may divert the focus of the person responsible for the analysis. Charts, infographics, spreadsheets, and tables are helpful resources to facilitate understanding. As each individual can count on a greater inclination to one of the methods, it is essential to know the characteristics of the professionals involved.

Step 6: Data Integration

The sixth step is data integration, defined based on studies by Mirzaei, Aslani and Schneider (2022) and Jahangir et al. (2022). It is necessary to understand the importance of having sectors aligned in their objectives, which, therefore, can have certain integrated functions. Thus, the Big Data steps must generate collaboration between all parties. An excellent way to promote this integration is to adopt an internal communication system that facilitates the vertical exchange of information. Big Data steps can vary according to the characteristics of the process and the objectives being pursued. However, points such as establishing goals, mining and validating information, and integrating data are essential in any approach. Following these steps, the way is opened for the generation of insights that enable more effective decisions for decision-making. The Rami architecture presents a method that simplifies and exemplifies the introduction of industry 4.0 in processes in a way that covers the entire production chain.

IV. CONCLUSION

This study showed a methodology for introducing big data and machine learning in line with the ins and outs of the Rami 4.0 architecture. The methodology consists of six steps: data collection, data cleaning, data mining, content analysis, data visualization, and data integration. The logic between these two tools is that data are collected so that they can feed a machine-learning scheme that allows them to perform activities based on the knowledge and learning they have acquired. Furthermore, the experiences of applying the generated knowledge provide back the learning scheme and provoke new and elevated operating procedures in a continuous process.In turn, the Rami architecture interconnects data collection and learning with countless other spheres of production processes, both internal to the organization and external to it.

ACKNOWLEDGMENTS

We greatly appreciate the financial, managerial, and institutional support of our partners: Envision Indústria de Produtos Eletrônicos Ltda., Centro Internacional de Tecnologia de Software do Amazonas(CITS AMAZONAS), Fundação de Apoio ao Ensino, Pesquisa, Extensão e Interiorização do IFAM (FAEPI), Pró-Reitoria de Extensão do IFAM (PROEX-IFAM), Polo de Inovação do IFAM, Suframa, and CAPDA.

REFERENCES

[1]. Abou-Zamzam, A. M. (2022). Learning from machine learning. Journal of Vascular Surgery, 75(1), 286-272. https://doi.org/10.1016/j.jvs.2021.07.119.

[2]. Alférez, G. H. et al. (2022). Automated machine learning pipeline for geochemical analysis. Earth Science Informatics, 1-16. https://doi.org/10.1007/s12145-022-00821-8.

[3]. Allaymoun, M. H. et al. (2023). Big data analysis and data visualization to help make a decision-Islamic banks case study. In International Conference on Business and Technology. Cham: Springer.

[4]. Baig, Z., Syed, N., & Mohammad, N. (2022). Securing the smart city airspace: drone cyber attack detection through machine learning. Future Internet, 14(7), 1-19. https://doi.org/10.3390/fi14070205.

[5]. Bhati, N. S., Chugh, G., & Bhati, B. S. (2022). Federated machine learning with data mining in healthcare. In Yadav, S. P. et al. (Eds.). Federated learning for IoT applications. Cham: Springer.

[6]. Bikakis, N. et al. (2022). Special issue on machine learning approaches in big data visualization. IEEE Computer Graphics and Applications, 42(3), 39-40. https://doi.org/10.1109/MCG.2022.3167749.

[7]. Brackenridge, R. E. et al. (2022). Improving subsurface characterisation with 'big data'mining and machine learning. Energies, 15(3), 1-23. https://doi.org/10.3390/en15031070.

[8]. Chahal, A., Gulia, P., & Gill, N. S. (2022). Different analytical frameworks and big data models for the internet of things. Indonesian Journal of Electrical Engineering And Computer Science, 25(2). https://doi.org/10.11591/ijeecs.v25.i2.pp1159-1166.

[9]. Chinchanikar, S., & Shaikh, A. A. (2022). A review on machine learning, big data analytics, and design for additive manufacturing for aerospace applications. Journal of Materials Engineering and Performance, (31), 6112-6130. https://doi.org/10.1007/s11665-022-07125-4.

[10]. Emu, M. et al. (2022). Fatality prediction for motor vehicle collisions: mining big data using deep learning and ensemble methods. IEEE Open Journal of Intelligent Transportation Systems, 3, 199-209. https://doi.org/10.1109/OJITS.2022.3160404.

[11]. Gandomi, A. H., Chen, F., & Abualigah, L. (2022). Machine learning technologies for big data analytics. Electronics, 11(3), 421. https://doi.org/10.3390/electronics11030421.

[12]. Hohmann, E. (2022). Editorial commentary: big data and machine learning in medicine. Arthroscopy: The Journal of Arthroscopic & Related Surgery, 38(3), 848-849.

[13]. Hongmei, X. (2022). Role of big data in the accounting profession. International Journal of Management and Education in Human Development, 2(1), 324-326.

[14]. Jahangir, S. et al. (2022). Data integration for big data analytics to identify the gaps in rheumatoid arthritis genomics in a post-GWAS era. In 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE). IEEE 2022 (pp. 956-964.)

[15]. Koman, G. et al. (2022). Business-making supported via the application of big data to achieve economic sustainability. Entrepreneurship and Sustainability Issues, 9(4), 336-358. http://doi.org/10.9770/jesi.2022.9.4(18).

[16]. Mirzaei, A., Aslani, P., & Schneider, C. R. (2022). Healthcare data integration using machine learning: A case study evaluation with health information-seeking behavior databases. Research in Social and Administrative Pharmacy, in press. https://doi.org/10.1016/j.sapharm.2022.08.001.

[17]. Muthiah, C. T. et al. (2022). Application of machine learning in fused deposition modeling: A review. In AIP Conference Proceedings. AIP Publishing LLC. https://doi.org/10.1063/5.0074144.

[18]. Nascimento-e-Silva, D. (2020). Manual do método científico-tecnológico: edição sintética. Florianópolis: DNS Editor.

[19]. Nascimento-e-Silva, D. (2021a). Handbook of the scientific-technological method: synthetic edition. Manaus: DNS Editor.

[20]. Nascimento-e-Silva, D. (2021b). O método científico-tecnológico: fundamentos. Manaus: DNS Editor.

[21]. Nascimento-e-Silva, D. (202c). O método científico-tecnológico: questões de pesquisa. Manaus: DNS Editor.

[22]. Nti, I. K. et al. (2022). A mini-review of machine learning in big data analytics: Applications, challenges, and prospects. Big Data Mining and Analytics, 5(2), 81-97. https://doi.org/10.26599/BDMA.2021.9020028.

[23]. Odabaşi, Ç. et al. (2022). Investigation of the factors affecting reverse osmosis membrane performance using machine-learning techniques. Computers & Chemical Engineering, 159(3), 107669. https://doi.org/10.1016/j.compchemeng.2022.107669.

[24]. Redjeki, S., & Widyarto, S. (2022). Big data analytics for prediction using sentiment analysis approach. Journal of Theoretical and Applied Information Technology, 100(13), 4987-5000.

[25]. Sánchez-Gutiérrez, M. E., & González-Pérez, P. P. (2022). Modeling and simulation of cell signaling networks for subsequent analytics processes using big data and machine learning. Bioinformatics and Biology Insights, 16, 1-13. https://doi.org/10.1177/11779322221091739.

[26]. Sharma, R. et al. (2022). Detecting Simpson's paradox: A machine learning perspective. In Strauss, C. et al. (Eds.). International Conference on Database and Expert Systems Applications. Cham: Springer.

[27]. Sharma, R., Agarwal, P., & Arya, A. (2022). Natural language processing and big data: a strapping combination. In: Sharma, R., & Sharma, D. (Eds.). New trends and applications in internet of things (IoT) and big data analytics. Cham: Springer.

[28]. Srividya, V., & Tripathy, B. K. Role of big data in supply chain management. In Perumal, K., Chowdhary, C. L., & Chella, L. (Eds.). Innovative supply chain management via digitalization and artificial intelligence. Singapore: Springer

[29]. Xing, Y. et al. (2022). Research on opinion polarization by big data analytics capabilities in online social networks. Technology in Society, 68(2), 1-12. https://doi.org/10.1016/j.techsoc.2022.101902.

[30]. Yoo, K.-H., Leung, C. K., & Nasridinov, A. (2022). Big data analysis and visualization: challenges and solutions. Applied Sciences, 12(16), 1-5. https://doi.org/10.3390/ app12168248.