



Data Driven to Improve Root Cause Analysis: A Bibliometric Analysis

Juan Pablo Carrión-Salazar^a, Emilia Ahumada-Tapia^a, Carolina Rojas-Córdova^a, Rosley Anholon^b, Izabela Simon Rampasso^a

^a Departamento de Ingeniería Industrial, Universidad Católica del Norte, Antofagasta, Chile

^b School of Mechanical Engineering, University of Campinas, Campinas Brazil

Abstract. Data Root Cause Analysis (RCA) allows organizations to improve their product and services, increasing their competitiveness. This process aims to identify the root cause of the failures or non-conformities, in this aspect, RCA plays a key role in continuous improvement process by correcting the failures from the root and minimizing its recurrence [3]. The exponential growth in computer's processing capacity has allowed applying and improving data driven techniques such as Machine Learning, Data Science, Data Mining, Big Data, among others related; in the process of identifying root causes.

There are significant benefits to applying Data-driven techniques in the RCA process [3]. Data-driven approaches allows increasing the available information for decision-making, improving the efficiency and effectiveness of the continuous improvement process and reducing the dependency of this process to the knowledge and opinions of the experts [4].

Despite the benefits of Data Driven RCA, the research in the field lacks a deeper understanding of the tendencies, challenges and evolution of itself. For that reason, we propose a comprehensive bibliometric analysis to accomplish this goal. Regarding the methodological procedures of the study, a search was made in the Scopus database to understand the literature on Data Driven RCA development using the software SciMat, considering the guidelines of Cobo et al. [2] and Cobo et al. [1]. Strategical diagrams are graphical tools, used to plot themes according to two measures, density and centrality. Through these diagrams, themes can be classified as "Motor", "Highly developed and isolated", "Emerging or declining", and "Basic and transversal" [1].

The search in Scopus was made using the string =("big data" OR "artificial intelligence" OR "machine learning" OR "pattern mining" OR "knowledge discovery" OR "data analysis" OR "data analytics" OR "data science" OR "data-driven" OR "data driven" OR "data mining") AND ("root cause analysis" OR "failure cause" OR "fault cause"). Through this search, 335 documents were found. From these documents, 206 articles were selected since they fit within the focus of the study. The range of articles' periods were from 2002 until 2022. To perform the analysis in SciMat, three sub-periods were established: 2002-2013 (27 articles), 2014-

2018 (65 articles) and 2019-2022 (114 articles). As motor themes, the first period presented “Root Cause Analysis”, with high centrality value but low density value, indicating that the term was important but low developed. In the cluster network of “Root Cause Analysis”, the terms statistical analysis, data analysis and data mining are highlighted. In the second period, the motor themes were “Anomaly Detection”, “Data Management”, and “Big Data”. According to the centrality analysis, data management was identified as the most relevant. In its cluster analysis, the machine learning was evidenced. “Data Mining” is a motor theme that can be evidenced in the last period, considering its combination of density and centrality values. Regarding its cluster analysis, manufacturing and performance analysis are outstanding topics.

Our presented bibliometric analysis provides great information about the research field of Data-Driven RCA. Through the analysis presented, it is possible to verify a large expansion of the research field. The analysis of Data Driven RCA shows an evolution to quality and process center developments, and the production of more complex systems that allow improving organizational efficiency and fast solutions to failures and non-conformity states. In addition, there are many articles that develop these systems in the context of Industry 4.0, Internet of Things, Cloud Computing and Big Data. Showing that its application may take advantage of the company’s processes’ data generated and its technological capabilities. Further research is needed in this field that develop data driven RCA but with a focus in the operational interaction and impacts between the data driven method and the process. Besides, the production of research focusing on understanding the faults or non-conformity states are needed, intending to build better frameworks for the development of Data-Driven RCA. The research is developed in different application contexts, the specific characteristic of different industries could provide better general understanding of the application of Data-driven RCA.

Keywords: Data-Driven, Root Cause Analysis, Bibliometric, Science Mapping, SciMat

References

1. Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*. 62 (7), 1382–1402 (2011)
2. Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., Herrera, F.: SciMAT: A new science mapping analysis software tool. *Journal of the American Society for Information Science and Technology*. 63 (8), 1609–1630 (2012)
3. Ma, Q., Li, H., Thorstenson, A.: A big data-driven root cause analysis system: Application of Machine Learning in quality problem solving. *Comput Ind Eng*. 160 107580 (2021)
4. Mueller, T., Greipel, J., Weber, T., Schmitt, R.H.: Automated Root Cause Analysis of Non-Conformities with Machine Learning Algorithms. *Journal of Machine Engineering*. 18 (4), 60–72 (2018)