

A Process and Application-Based Framework for the Optimization of Manufacturing Process

Daniele Crippa*, Davide Cesani*, Fabiana Pirola**, Roberto Sala**

* *Intellimech, Bergamo, Italy* (daniele.crippa@intellimech.it, davide.cesani@intellimech.it)

** *Department of Management, Information and Production Engineering, University of Bergamo, Dalmine, Italy* (fabiana.pirola@unibg.it, roberto.sala@unibg.it)

Abstract: In the current manufacturing landscape process optimization, cost reduction, and profit maximization have become crucial. This paper aims to leverage advanced data analysis techniques such as Machine Learning (ML) to achieve these objectives, identifying issues related to operational inefficiencies, data integration challenges, and the effective application to real manufacturing problems. To do so, the paper proposes a framework to improve ML integration into manufacturing operations, highlighting the importance of customizing the strategy to suit the company needs. Through a practical implementation, this paper shows the applicability of the framework.

Copyright © 2025 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: manufacturing, machine learning, association rules, framework, optimization

1. INTRODUCTION

Thanks to the foreseen augmented decision-making possibilities and its ability to streamline analyses and tasks, manufacturing companies are more and more eager to integrate Artificial Intelligence (AI) and Machine Learning (ML) into their processes and core activities (Okuyelu et al., 2024). In particular, ML aims at developing algorithms and models allowing computers to learn and support humans in their decision-making process, thus fostering experience-based improvement of performance exploiting predictions or decisions based on data. Even though ML is seen as a major leap forward for data processing and analysis, it complements the traditional statistical methods, having therefore its own strengths and weaknesses. In particular, statistics is more interested in hypothesis testing, while ML interest is the generalization through hypothesis search (Witten et al., 2005). According to some authors, when industrial data collection and processing is scaled through suitable AI and ML techniques, companies can gain important insights, allowing to improve their productivity and reduce losses in their operations (Rai et al., 2021). This research goes in this direction, proposing an analysis on Machine Learning (ML) techniques that are aimed at allowing companies to perform better.

The paper sets its foundations in the analysis of the literature related to ML applications in manufacturing. The analysis allowed to group papers according to the process of application, the task, the data and the ML approach used. The paper intends to provide a practical insights on ML use through the consideration of the as-is situation and needs of the user, thus demonstrating how versatile and effective ML can be in different manufacturing environments.

The paper is structured as follows: Section 2 explains the methodology adopted for the work, while Section 3 provides an overview of the literature on data analysis approaches with a focus on machine learning and optimization. Section 4 describes structure of the framework developed based on the results of the literature, whose scope is to guide companies

towards optimizing their operations using ML. Section 5 presented an application of the framework in Company A, demonstrating how process inefficiencies can be identified. Section 6 concludes the paper and delineates future research.

2. RESEARCH METHODOLOGY

The methodology used to develop the work consists of:

1. *Literature review & Analysis.* In this step, literature related to the topic of ML and manufacturing was reviewed and analyzed. Only papers dealing with case studies/applications were considered as they could show how ML approaches were applied to real industrial setting clarifying their strengths and weaknesses.
2. *Summary of the findings and lesson learned.* Tables related to the process/application as well as approach/application and data types for each approach have been defined. This allowed to a better understanding of the potentialities of current ML applications in the manufacturing field.
3. *Framework proposal.* A framework guiding the selection of ML approaches and their application was proposed as a combination of the literature results and the experience of the researchers, who are working in the field since a few years. The framework was developed with the aim of aiding companies and unexperienced users in the adoption of ML in their processes and analysis
4. *Framework application.* Finally, the framework was applied to a real case study to show the application process and allow understanding its effectiveness and usefulness for the task its has been developed.

3. LITERATURE BACKGROUND

This section presents the literature background used for the development of the framework. It is composed of several subsections that first explain how the analysis has been conducted and then the main topics useful for the comprehension of the framework's structure.

3.1 Paper search

The Scopus database, due to its multidisciplinary nature and the presence of peer-reviewed articles was selected for the research. The terms “Machine Learning”, “Manufacturing” and “Case Study” were searched in the title, abstract and keywords fields using the operator AND. Due to the extensive number of results, filters related to the publication year and nature of the papers were applied. Only journal papers written between 2018 and 2024 were considered. This choice was made in order to consider only the recent and relevant literature. Additionally, only the ones in the subject area of “engineering” or “computer science” and published in English were selected, leading to a pool of 330 papers at the time of the research. Then, the title and abstract were read and additional papers were discarded due to their content not aligned with the scope of research. Eventually, the final pool of analysis consisted of 243 papers explicitly describing ML applications in the manufacturing field.

3.2 Optimization and Machine Learning

Optimization aims at the minimization or maximization of a given objective function, often under constraints (Calafiore and El Ghaoui, 2014). According to Weichert et al., (2019), it can take place for two main objectives: the improvement of product quality and the improvement of production and planning processes. ML is instead selected due to its additional capabilities to save and optimize the use of resources, reduce machining time and energy, and increase yield when traditional methods have reached their limits (Köksal et al., 2011). ML is concerned with the use of algorithms and programs aimed at ameliorating the execution of tasks and the process performance through experience (Makkar et al., 2020).

In terms of computational requirements, both optimization and ML require considerable amount of data and computing power to improve decision-making in a reliable way. As said, although they may look similar, there are important differences between them. ML is frequently used for predictive analysis as it learns from past information to find patterns and make predictions on the behavior of a system or process in the future (Carvalho et al., 2019). Thus, by using ML, companies can make long-term decisions based on data with the aim of increasing profitability (Sharma et al., 2021). On the other hand, optimization is frequently used for prescriptive purposes (Karkaria et al., 2025). In particular, it uses current data to determine what action should be taken. Given the premise of the necessity of many data to make reliable and accurate decisions, computational costs must be considered as part of the trade-off with the accuracy of the models (Wang et al., 2022). Highly accurate models might also be very expensive and, sometimes, it is better to choose a less accurate model.

ML algorithms can be roughly categorized in two big sets: Supervised learning and Unsupervised learning. Supervised learning involves labeled data and human input for tasks like labeling, feature selection, and parameter tuning, making it suitable for domains with expert knowledge (Kang et al., 2020); key techniques include Classification, Regression, and Time Series Forecasting. In contrast, unsupervised learning works with unlabeled data to discover hidden patterns or

groupings without human guidance (Makkar et al., 2020), using methods like Clustering and Dimensionality Reduction.

3.3 Deep and Reinforcement Learning

Deep Learning is a sub-discipline of ML that uses deep Neural Networks (NN). NN are among the most used approaches in ML, and consist of a connection of simple elements (i.e., units or nodes) whose functionality is inspired by the human neurons (Gurney, 2018). The nodes are connected by links that represent a weight determined based on the training data and patterns identified by the algorithm (Gurney, 2018). NN is used for statistical research and data modeling and it is a well-known alternative to “standard” nonlinear regression or cluster analysis (Cheng and Titterington, 1994). NN is frequently used for classification or forecasting problems, for instance in image, speech or textual recognition in various application fields (Gurney, 2018). Over the years, different typologies of NN have been developed (e.g., Convolutional Neural Network (CNN), Recurrent Neural Network (RNN) and others). Another important ML family is represented by Reinforcement Learning (RL), which focuses on learning processes through a trial and error approach (Makkar et al., 2020). After the definition of some rules, the algorithm explores different options and possibilities, selecting in the end the one that best satisfies the goal (Makkar et al., 2020).

4. FRAMEWORK PROPOSAL

The papers have been classified based on their application scope, which means according to the business process, application and the data used. This section presents the framework developed based on these results and the authors experience, which consists of five steps: 1) Identification of critical manufacturing process, 2) Identification of application of interest, 3) Selection of data analysis approach, 4) Verification of data type compatibility, and 5) Implementation of data analysis model. The framework is not only consisting of the sole tables presented in the following sections, as they serve as a support for the application of the framework.

4.1 Identification of critical manufacturing process

The first step allows companies to identify the process they would like to enhance according to their needs:

- *Production*. It involves the creation of goods, products, or services, considering the entire production process.
- *Testing/Quality Check*. It ensures that products meet certain standards, specifications, and requirements.
- *Maintenance*. It is based on routine checks and equipment replacement to keep machines running smoothly.
- *Logistics*. It involves the planning and coordination of the movement and storage of goods and information in the warehouse and supply chain.
- *Supply Chain*. It is the overall flow which can be traced from the purchase of raw materials to the delivery of the customer's finished product.
- *Personnel Management*. It is concerned with managing staff, including hiring, training, welfare and productivity.

4.2 Step 2: Identification of application of interest

Here, companies determine the area to improve, based on the previously selected process:

- *Flow Optimization.* It focuses on improving the efficiency and reducing idle time during manufacturing processes, including the allocation of resources, machines, shipment planning, delivery flows, and shift planning.
- *Process Analysis.* It involves the study of processes as they are, emphasizing their actual performance, inefficiencies, and improvement areas.
- *Process Forecast.* It predicts the future performance and results based on historical data and current trends.
- *Anomaly Detection.* It aims to identify unusual patterns or deviations that might point to potential problems either in the production line or final product.
- *Resource Allocation.* It involves the distribution and management of human and productive resources to maximize productivity and minimize costs.

Table 1 maps the number of papers found in the review dealing with critical manufacturing processes and their applications.

Table 1 - Papers classified according to process and application. Process: P1 (Production), P2 (Testing/Quality Check), P3 (Maintenance), P4 (Logistics), P5 (Supply chain), P6 (Personnel management). Application: A1 (Flow optimization), A2 (Process analysis), A3 (Process forecast), A4 (Anomaly detection), A5 (Resource allocation).

	A1	A2	A3	A4	A5
P1	16	59	67	28	5
P2	3	50	18	31	2
P3	0	0	7	25	0
P4	2	1	1	0	1
P5	3	1	8	2	0
P6	0	1	4	0	2

4.3 Selection of data analysis approach

Following, the most suitable data analysis approach is selected. Here, once the main application area is identified, the most suitable data analysis approaches is selected:

- *Classification.* It aims to map input features to one of the discrete output variables.
- *Regression.* It maps the input features to a numerical and continuous target.
- *Time series forecasting.* It involves predictions made based on data comprising one or more Time series.
- *Survival analysis.* It aims to analyze and model data to predict the time until the occurrence of one or more events.
- *Clustering.* It involves grouping sets of similar data, based on defined criteria and patterns.

- *Association rules.* It aim to extract correlations, frequent patterns, associations or casual structures among sets of items in transaction databases or other repositories.
- *Neural Networks.* They are connections of elements, units or nodes, whose functionality is inspired by the human neurons.
- *Reinforcement Learning.* It focuses on regimented learning processes based on a trial and error approach.
- *Optimization/Simulation.* It is the process by which the optimal solution to a problem, or optimum, is produced.

Table 2 maps the number of papers dealing with critical manufacturing processes and their major applications.

Table 2 – Papers classified according to approach and application. Approach: AP1 (Classification), AP2 (Regression), AP3 (Time series forecasting), AP4 (Survival analysis), AP5 (Clustering), AP6 (Association rules), AP7 (Neural networks), AP8 (Reinforcement Learning), AP9 (Optimization/Simulation). Application: A1 (Flow optimization), A2 (Process analysis), A3 (Process forecast), A4 (Anomaly detection), A5 (Resource allocation).

	A1	A2	A3	A4	A5
AP1	0	17	11	13	2
AP2	2	16	17	9	0
AP3	0	1	7	2	0
AP4	0	1	0	0	0
AP5	3	9	4	3	1
AP6	0	2	0	4	0
AP7	4	36	21	29	1
AP8	6	3	0	1	1
AP9	9	11	2	1	1

4.4 Step 4 & 5: Verification of data type compatibility and model implementation

Step 4 makes sure that the company’s data are aligned with the requirements for successful implementation (Table 3). If this is true, the company can move to Step 5, where the data analysis is put into practice; otherwise, it is necessary to revisit Step 3. Step 4 is crucial, because, if the data does not meet certain requirements, for the input or the output, alternative approaches must be considered as the result of the analysis would be unreliable. If the process, application and approach are coherent, the final phase begins.

Table 3 - Required data type for each approach

Approach	Data characteristics
Classification	Discrete labeled targets
Regression	Continuous targets
Time series forecasting	Ordered data sampled over time
Survival analysis	Temporal data censored
Clustering	Unlabeled data
Association rules	Transactional data
Neural networks	Big data (any)
Reinforcement learning	Descriptive data of an environment that allows its simulation
Optimum search/DES	Data must define a search space (domain/constraints)

5. FRAMEWORK APPLICATION

Company A is an Italian manufacturing company that designs, produces and distributes low-voltage electrical devices for industrial automation and energy management. It aims at studying how much machine performance is affected by human effort and errors in order to control the optimal state for productivity. In other words, one of the main aims of the company is to identify the conditions for optimal productivity of its machines, taking into consideration the man-machine interaction. To this purpose, the firm is interested in improving the production data quality and finding correlations between some variables to identify inefficiencies, particularly those related to operator performance and machine downtime. Additionally, it aims at improving KPIs, currently focused on operator productivity rather than machine performance. The authors supported the company in the application of the framework, guiding the company in identifying the needs and the data required for the application, as well as, executing the data analysis. Due to privacy concerns, it is not possible to share the original data used for the analysis.

5.1 Framework application for approach selection

In order to achieve the above-mentioned objective, Company A decided to implement the proposed framework to enhance the personnel allocation to machine by determining the state in which productivity is at its optimum level. The framework has been applied as follows:

1. Identification of critical manufacturing process – Personnel Management.
2. Identification of application of interest – Resource Allocation.
3. Selection of data analysis approach – Association Rules.
4. Verification of data type compatibility – Transactional Data.
5. Implementation of data analysis model – The main objective of the company has been divided in the following sub-objectives. Sub-Objective 1: examine machine downtime to identify the most frequent pairs of human-machine inefficiencies during the analyzed period. Sub-Objective 2: identify the most common pairs of human-machine inefficiencies considering four specific scenarios: quantitative inefficiencies, time inefficiencies, time loss due to undefined issues and setups, and impact of plant alarms on downtime.

5.2 The Apriori algorithm

Based on the need of the company, the Apriori algorithm belonging to Association Rules was selected, mainly because it is efficient, easy to execute and very simple to identify the frequent itemsets in a database, making many searches to find recurring combinations where k-itemsets are used to generate k+1-itemsets (Al-Maolegi and Arkok, 2014). The main steps for Apriori implementation are (Purdom et al., 2004):

1. Iterate for l from 1 to |I|

2. For each set J_l ensure that, for each h ∈ J_l, the set J_l – h occurs in at least k transactions. The algorithm generates each set J_l such that J_l – h occurs at least k times, for each h in J_l, and the sets that are generated are called candidates.
3. Examine the data to determine whether the set J_l occurs in at least k transactions. The algorithm examines the data to determine which set of elements occurs in at least k transactions.

A “support hyperparameter” is used to determine if a combination of items is frequent or not. In particular, if the support hyperparameter meets a minimum threshold, an itemset is considered “frequent”. By using such a proxy, it is possible to minimize the risk that the associations are incidental and do not express causality. For instance, if the support is set to 0.5 (50%), a set of elements that appear together in at least 50% of all transactions in the database is considered frequent. The formulas for the computation of the support are the following (Panjaitan et al., 2019):

$$Support(A) = \frac{Number\ of\ transactions\ containing\ A}{Total\ transactions} \quad (1)$$

$$Support(A, B) = \frac{\sum Transactions\ containing\ A\ and\ B}{\sum Transactions} \quad (2)$$

Different metrics are used to generate Association rules once frequent itemsets are identified. For this paper, confidence has been used as a metric because it measures the accuracy of a rule, which is represented by a transaction that contains A and B (Panjaitan et al., 2019):

$$Confidence = P(A|B) = \frac{\sum Transactions\ containing\ A\ and\ B}{\sum Transactions\ containing\ A} \quad (3)$$

The main objective, utilizing the company’s actual data from one month of operations, is to find out patterns that can be exploited in order to improve productivity levels.

5.3 Generation of the Association Rules for Objective 1

The Step 5 requires three subphases: data understanding, data preparation and analysis, and Apriori implementation.

Understanding the data is the foundation for all the following steps. Once the user knows what the data represents and how different data relate to each other, it is possible to move to the cleaning and filtering phase. During this stage, the data is not only cleaned and filtered, but also grouped and visualized to get it ready for the application of a ML algorithm. Two Python libraries were used for this process: Pandas and Mlxtend, which are commonly used for this.

The focus of the analysis was on process interruptions caused by alarms and undefined issues. These are logged in the dataset and are considered significant because they contribute to what the company called “interesting downtime”. Date, machine, and operator information were used to filter data. Based on this, the percentage of interesting downtime due to each human-machine pair was calculated.

To identify which combinations were most problematic, a visual analysis of idle time distribution for each machine was done (see Figure 1). A lower threshold was set to identify the

significant interesting downtime: only combinations where more than 60% of the idle time was due to interesting downtime were considered for further analysis. Finally, the results were converted into a format suitable for the Apriori algorithm using binning techniques.

Once done this, the most frequent issues that characterize the plant operations can be identified. As already explained, two main steps are required for this: a) identifying frequent itemsets and b) generating Association rules.

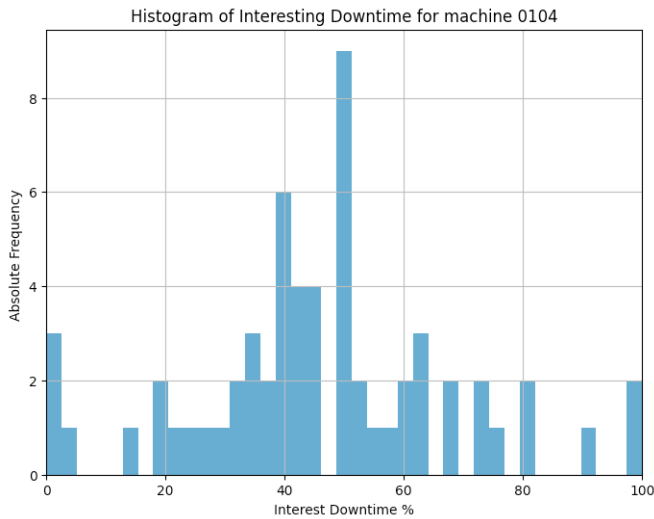


Figure 1 - Idle time distribution. Example from the dataset.

For this application, the support value is set equal to 0.05, meaning that a combination is frequent if it appears in at least 5% of all transactions, as it identifies only the most relevant pairs compared to other thresholds. Next, the Apriori algorithm is applied through the following python code:

```
apriori(df_encoded, min_support=0.05, use_colnames=True)
```

Table 4 – Interesting itemsets of Machine-Operator showing inefficiencies. Legend: Operators (Op_X). Machine: (Ma_X).

Itemset	Support
(Ma 101, Op 452)	0.053333
(Ma 104, Op 847)	0.080000
(Ma 202, Op 637)	0.080000
(Ma 202, Op 1098)	0.053333
(Ma 203, Op 1108)	0.066667

These results (Table 4) indicate that only five combinations of operator-machine show inefficiencies that occur in at least 5% of all transactions. The final step is to generate Association rules. In this case, the precision of the rules is assessed using the confidence metric with a threshold set to for confidence to 0.05 as it allowed to have a balance between the number of rules and their significance. The following python code was applied to generate the rules:

```
association_rules(frequent_itemsets, metric="confidence", min_threshold=0.05, support_only=True)
```

These rules present mutual relationships between operators and machines and represent the major general inefficiencies in the plant.

Table 5 – Association rules generated for Objective 1. Legend: Operators (Op_X). Machine: (Ma_X)

Antecedents	Consequents	Support
(Ma 101)	(Op 452)	0.053333
(Ma 104)	(Op 847)	0.080000
(Ma 202)	(Op 637)	0.080000
(Ma 202)	(Op 1098)	0.053333
(Ma 203)	(Op 1108)	0.066667

Table 4 and Table 5 identify the main human-machine combinations that caused issues in the company’s plant during one month of operations, considering, in particular, the downtime due to alarms and undefined causes compared to the total idle time of machines.

5.4 Generation of the Association Rules for Objective 2 and comparison with Objective 1

Other analyses that have been run on the dataset for Objective 2. Due to space constraints, Table 6 only reports the results. These analyses are focused on human-machine inefficiencies and were run on four scenarios: a) quantitative inefficiencies regarding the production process, b) time inefficiencies, c) time loss due to undefined issues and setups, and d) impact of plant alarms on downtime. Specific thresholds values were selected depending on the scenario. This choice was made as each inefficiency has a different impact on the production.

Table 6 – Results for Objective 2. Legend: Antecedents (Ant), Consequents (Cons), quantitative inefficiencies regarding the production process (2A), time inefficiencies (2B), time loss due to undefined issues and setups (2C), impact of plant alarms on downtime (2D).

Ant	Cons	O2A	O2B	O2C	O2D
(Ma 101)	(Op 452)	-	-	V	V
(Ma 104)	(Op 847)	-	V	-	V
(Ma 202)	(Op 1098)	V	V	V	V
(Ma 203)	(Op 666)	V	V	-	-
(Ma 107)	(Op 452)	-	-	-	V
(Ma 205)	(Op 1093)	-	-	-	V

Table 7 shows how, in certain cases, the same itemset is identified in the different analyses. This is important for two reasons: on the one hand, it allows to understand that it is necessary to conduct analyses on different perspectives to understand all the constituents of a problem and thus, to solve it. On the other, it allows to identify the reasons for each issue and thus, select appropriate countermeasures. For instance, Op_452 show inefficiencies related to undefined issues, setups and alarms. Op_104 instead, show inefficiencies related to time and problems related to alarms. It is important to clarify that this study focuses on a single one-month period and that

data should be examined over multiple periods to find if these inefficiencies are consistent or just accidental occurrences.

Table 7 - Results. Legend: Antecedents (Ant), Consequents (Cons), Objective 1 (1), Objective 2 Scenario 1 (2A), Objective 2 Scenario 2 (2B), Objective 2 Scenario 3 (2C), Objective 2 Scenario 4 (2D).

Ant	Cons	Objective				
		1	2A	2B	2C	2D
(Ma 101)	(Op 452)	V	-	-	V	V
(Ma 104)	(Op 847)	V	-	V	-	V
(Ma 202)	(Op 637)	V	-	-	-	-
(Ma 202)	(Op 1098)	V	V	V	V	V
(Ma 203)	(Op 1108)	V	-	-	-	-
(Ma 203)	(Op 666)	-	V	V	-	-
(Ma 107)	(Op 452)	-	-	-	-	V
(Ma 205)	(Op 1093)	-	-	-	-	V

6. CONCLUSIONS

This paper proposed a framework for guiding companies in the improvement of business processes through ML applications. The framework, composed of five steps, provides a structured guide for companies to select an approach that is suited for their objectives as this is the result of a literature analysis on case studies and applications in manufacturing as well as the experience of the authors on data analysis tasks. Such guide is structured in a way that it resemble a checklist that ensures that users follow all the steps and reflect on their needs and the corresponding solutions, as well as evaluating the availability of all the required data. The application case, carried out in Company A, demonstrated how it can be used by companies to optimize processes, reduce costs, and maximize profits, as well as to identify issues related to operational inefficiencies and the effective application in real-world contexts. In the case study, the Association rules approach was able to identify process inefficiencies and their cause.

This work is not free from limitations as the list of proposed approaches could be extended and further application cases could be shown to verify the strength and weaknesses of the proposed approach (e.g., including Deep Learning and Neural Networks). Additionally, the literature research was executed only on Scopus, while other useful applications might be found in papers present in other database such as IEEE or Web of Science. Future research focus on the extension of the framework and will delve deeper in the architecture of these techniques.

REFERENCES

- Al-Maolegi, M., Arkok, B., 2014. An improved Apriori algorithm for association rules. arXiv preprint arXiv:1403.3948.
- Calafiore, G.C., El Ghaoui, L., 2014. Optimization models. Cambridge university press.
- Carvalho, T.P., Soares, F.A.A.M.N., Vita, R., Francisco, R. da P., Basto, J.P., Alcalá, S.G.S., 2019. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering* 137, 106024. <https://doi.org/10.1016/j.cie.2019.106024>
- Cheng, B., Titterton, D.M., 1994. Neural networks: A review from a statistical perspective. *Statistical science* 2–30.
- Gurney, K., 2018. An introduction to neural networks. CRC press.
- Kang, Z., Catal, C., Tekinerdogan, B., 2020. Machine learning applications in production lines: A systematic literature review. *Computers & Industrial Engineering* 149, 106773.
- Karkaria, V., Tsai, Y.-K., Chen, Y.-P., Chen, W., 2025. An optimization-centric review on integrating artificial intelligence and digital twin technologies in manufacturing. *Engineering Optimization* 57, 161–207.
- Köksal, G., Batmaz, I., Testik, M.C., 2011. A review of data mining applications for quality improvement in manufacturing industry. *Expert systems with Applications* 38, 13448–13467.
- Makkar, S., Devi, G.N.R., Solanki, V.K., 2020. Applications of machine learning techniques in supply chain optimization, in: *ICICCT 2019–System Reliability, Quality Control, Safety, Maintenance and Management: Applications to Electrical, Electronics and Computer Science and Engineering*. Springer, pp. 861–869.
- Okuyelu, O.M., Doskenov, B., Aribike, R., 2024. Process Optimization in Manufacturing Industries Using Simulation Technologies. *Current Journal of Applied Science and Technology* 43, 31–41.
- Panjaitan, S., Sulindawaty, Amin, M., Lindawati, S., Watianthos, R., Sihotang, H.T., Sinaga, B., 2019. Implementation of apriori algorithm for analysis of consumer purchase patterns, in: *Journal of Physics: Conference Series*. IOP Publishing, p. 012057.
- Purdom, P.W., Van Gucht, D., Groth, D.P., 2004. Average-case performance of the apriori algorithm. *SIAM Journal on Computing* 33, 1223–1260.
- Rai, R., Tiwari, M.K., Ivanov, D., Dolgui, A., 2021. Machine learning in manufacturing and industry 4.0 applications. *International Journal of Production Research*.
- Sharma, A., Zhang, Z., Rai, R., 2021. The interpretive model of manufacturing: a theoretical framework and research agenda for machine learning in manufacturing. *International Journal of Production Research* 59, 4960–4994.
- Wang, J., Li, Y., Gao, R.X., Zhang, F., 2022. Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. *Journal of Manufacturing Systems* 63, 381–391.
- Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., Wrobel, S., 2019. A review of machine learning for the optimization of production processes. *The International Journal of Advanced Manufacturing Technology* 104, 1889–1902.
- Witten, I.H., Frank, E., Hall, M.A., Pal, C.J., Data, M., 2005. Practical machine learning tools and techniques, in: *Data Mining*. Elsevier Amsterdam, The Netherlands, pp. 403–413.