Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.10 n.45, p. 17-29. Jan/Feb, 2024

DOI: https://doi.org/10.5935/jetia.v10i45.920

RESEARCH ARTICLE OPEN ACCESS

APPLICATION OF A PRODUCTION PLANNING MODEL BASED ON LINEAR PROGRAMMING AND MACHINE LEARNING TECHNIQUES

Lucas Vianna Vaz[1], Marcelo Carneiro Gonçalves*[2], Izamara Cristina Palheta Dias[3] and Elpídio Oscar Benitez Nara[4]

[1, 2, 3, 4] Industrial Engineering, Pontifical Catholic University of Paraná - PUCPR, Curitiba, Paraná, Brazil.

[1] http://orcid.org/0009-0005-3469-0366, [2] http://orcid.org/0000-0002-4957-6057, [3] http://orcid.org/0000-0001-5413-0423, [4] http://orcid.org/0000-0002-4947-953X

Email: lucas.vaz@pucpr.edu.br, *marcelo.goncalves@unb.br, izamara.dias@pucpr.edu.br, elpidio.nara@pucpr.br

ARTICLE INFO

Article History
Received: October 03th, 2023
Revised: January 19th, 2024
Accepted: January 26th, 2024
Published: February 29th, 2024

Keywords:
Artificial Intelligence,
Machine Learning,
Operations research,
Production Planning and Control.

ABSTRACT

The absence of efficient optimization methods combined with Artificial Intelligence concepts has led to inefficiencies and high costs in the production planning of organizations. Thus, this study aims to optimize production planning in an electronic equipment company, using Linear Programming and Machine Learning to support assertive and efficient decisions. The methodological process comprises seven stages: Literature review; Collection and analysis of production data; Application of Machine Learning methods for modelling; Selection of the best model; Development and application of the Linear Programming model; Analysis of results; Validation with stakeholders. The approach resulted in optimized production planning, capable of reducing operating costs and assisting in the daily decision-making of the organization. The Machine Learning forecasting technique achieved an average error of 9%, demonstrating its accuracy in forecasting future demand. This study evidences a robust and promising approach to improve efficiency and effectiveness in production planning operations. In this context, the union between Operations Research and Machine Learning emerges as a response to existing gaps and a driving direction for continuously optimizing these crucial processes.

I. INTRODUCTION

Efficiency and effectiveness in production are crucial for the competitiveness of companies in the modern era. Production planning plays a crucial role in this process, as it must provide accurate information about the status of resources - including people, equipment, facilities, and materials - and orders, whether for purchase or production [1]. This information is essential for effective resource management, activities, and information flow, ensuring that production meets customer expectations and is completed on time and within budget [2]. A well-crafted production planning strategy can generate numerous benefits, such as cost reduction, increased productivity, inventory optimization, and greater customer satisfaction. Furthermore, it enables companies to identify and prevent problems before they occur.

In the context of technological advancement, integrating Machine Learning and Artificial Intelligence concepts has become crucial to enhance production planning further. Since the 1960s, companies have implemented advanced systems such as Material Requirements Planning (MRP) to support resource management decisions. However, more than these systems are needed to consider optimal resource allocation solutions. This is where mathematical optimization models, such as Linear Programming (LP), come into play. However, these traditional models assume that input parameters are deterministic and do not consider significant data fluctuations observed in business routines [3].

The fundamental question addressed in this article is as follows: Is it feasible to combine mathematical models from operations research with machine learning techniques to enhance the production planning of electronic equipment organizations, reduce costs, increase productivity, and Enhance customer satisfaction? This integration promises to open new perspectives and innovative solutions to companies' challenges in constantly pursuing efficiency and operational excellence.

Therefore, this article aims to optimize Production Planning in an electronic equipment company, using robust Linear Programming and Machine Learning techniques to assist in informed decision-making within the organization. To achieve the objectives, the following steps will be developed: Review the literature; Collect and analyze data; Apply machine learning methods; Select the best machine learning model; Apply a linear programming model; Analyze the model's results and validate with stakeholders.

II. THEORETICAL REFERENCE

II.1 LITERATURE REVIEW

For the literature review, this research extensively searched Scopus and Science Direct databases to identify scientific papers related to Production Planning and Control (PPC) in conjunction with machine learning techniques applied to electronic equipment data [4]. Python programming language was employed for data analysis, incorporating a range of prediction algorithms, including K-Nearest Neighbor, Random Forest, Support Vector Machines, Gradient Boosting Machine, and Linear Regression. These algorithms were initially adapted from pre-existing models and subsequently tailored to address the specific challenges presented in this study. The selection of these machine learning techniques was informed by their prevalence in recent literature, as evidenced by a comprehensive review of publications from the past five years utilizing the keywords K-Nearest Neighbor, Random Forest, Support Vector Machines, Gradient Boosting Machine, and Linear Regression. This search yielded 47 relevant scientific articles (see Annex A), with the most employed methods aligning closely with those featured in this research [4]. Notably, the application areas of these techniques varied, with the majority (49%) applied in medicine, 7% in biotechnology, 8% in computer science, and the remaining 36% spanning various other domains. Specific articles from each sector were cited as illustrative examples. Notably, no articles addressing the application of research operations and machine learning using electronic equipment data were identified, thus marking this research as an innovative contribution to the field.

II.2 MACHINE LEARNING

Machine Learning is one of the existing branches that focuses on developing algorithms and statistical models that enable computers to learn and make decisions or predictions without being explicitly programmed for specific tasks. Instead, machine learning empowers computers to learn from available data, constantly seeking to identify patterns, relationships, and valuable information through training and optimization [5].

In machine learning, it is necessary to work with a large amount of data, known as a database. This extremely large and complex data set cannot be easily managed. In other words, big data refers not only to the size of the data but also to its complexity and the need for tools and techniques to extract useful information from these datasets [6]. A large database is required to apply this technique because the model uses 70% of the data for learning and only 30% for testing to ensure a low error percentage. The more data, the more robust the model becomes and the more accurate it becomes. Machine learning is widely used in various fields, such as architecture [7], and in the quantum field with the application of natural language processing [8].

II.3 OPERATIONAL RESEARCH

According to [9], linear programming has been classified as one of the greatest scientific advancements of the 20th century. In summary, this tool has saved thousands of dollars for numerous companies and has been used in various ways and across various topics. In production, it is widely employed to provide direction and understand how to allocate limited resources to competing activities. In this context, the term "programming" does not necessarily refer to computer programming but rather to planning. Therefore, it is used to achieve optimal planning focusing on optimization.

Linear programming is a branch of mathematics that deals with optimization problems involving linear functions and constraints. The objective in these problems is to maximize or minimize a linear function, known as the objective function, subject to a set of linear constraints. To achieve this objective, it is necessary to find the values of the variables that optimize the objective function while adhering to the established constraints [9]. Examples of applications of these cases related to production planning in various organizations can be seen in [10], [11], [12], [13], and [14].

III. MATERIALS AND METHODS

This study is classified as applied research, employing quantitative and qualitative approaches. It utilizes a systematic literature review to provide a foundation for the context of robust optimization in production planning, followed by bibliometric analysis to identify trends in the application of machine learning in conjunction with linear programming. The research collected and analyzed data through interviews and spreadsheets, applied machine learning techniques, developed mathematical models, and validated results with company stakeholders.

The study can be divided into five steps as follows:

1. Systematic Review and Bibliometric Analysis: Based on [4], a systematic literature review was conducted to explore robust optimization methods in engineering, focusing on parameter uncertainty. Using the Scopus database and the Bibliometrix package, Bibliometric analysis revealed the main directions and challenges in using machine learning associated with linear programming in production planning.

2. Data Collection and Organization: Interviews were conducted to collect data, focusing on relevant products. The data were organized in Excel spreadsheets.

3. Application of Machine Learning: Using Python and Jupyter Notebook, various machine learning methods such as Random Forest (RF), k-nearest Neighbors (KNN), Support Vector Machine (SVM), Gradient Boosting Machine (GBM), Linear Regression (LR), and Linear Programming were applied. The choice of the best model was based on continuous error metrics.

4. Mathematical Model and Optimization: Mathematical modelling through Linear Programming was chosen due to the linearity of constraints and the objective function. Parameters were collected, and the machine learning demand results were inputs. The optimization aimed to minimize total costs in different production scenarios.

5. Validation with Stakeholders: Results were validated with stakeholders by comparing them with the actual company operation, ensuring practical applicability.

IV. RESULTS

IV.1 COLLECTING THE DATA

A company named "M," founded in 2012, operates primarily in the electronics manufacturing sector, specializing in the assembly of electronic boards. The assembly of these boards involves soldering components onto the board to create a functional electronic board. These components can be divided into SMD (Surface Mount Device) and PTH (Plated Through-Hole). SMD components are smaller and are typically assembled by a machine called an "insertion machine." In contrast, PTH components, which are larger, can also be machine-assembled but are often done manually.

The company is headquartered in Curitiba, located in Paraná, Brazil, and currently employs 90 people. The company's business model focuses on enabling the production of electronic projects for its clients, either through an industrialization model or a turnkey sales model, as referred to internally. According to information provided by the company's employees, the industrialization model accounts for 95% of the monthly demand, while the remaining 5% comprises the sales model.

The company has manufactured over 3,000 different models for various clients, encompassing a wide range of projects, including diverse product segments and varying levels of complexity.

Initially, there were seven different products. However, due to the high variability of their client's projects, the company sought to identify two products in high-demand representation in their portfolio collaboration with the Production Planning Control (PCP) department. As identified employees, chosen products, board 371000490 and board 37100493, have a monthly output of two batches of 1200 pieces each. These two products are called Product 1 and Product 2 throughout the work.

Since the database initially included seven different products, reducing the focus on two products and simplifying presentation results is necessary. The collected product characteristics included order number, quantity ordered, unit price, quantity sold, month sale, year sale, and product. Therefore, considering subsequent application machine learning methods, features representing predictor variables were order number, quantity ordered, unit price, month, year, and product; feature representing response variable quantity sold. This differentiation between variables is necessary to determine variable-dependent others. In this case, sales depend on predictor variables; for example, sales quantity depends on unit price—higher prices tend to result in lower sales, and vice versa.

Considering the response variable "sales," a total of 2823 data points, an average of 3553.9 units, and a standard deviation of 1841.85 units. Maximum sales quantity observed 14,082 units, minimum 482.13 units.

IV.2 COMPARISON BETWEEN MACHINE LEARNING TECHNIQUES

In this stage of the process, five machine learning techniques were applied, among them linear regression (LR), gradient boost machine (GBM), random forest (RF), SVM and KNN. The choice was based on the study by [4], which supported the literature review. The demand forecasting model uses 70% of the data to train the model and 30% to test whether the model generates assertive solutions compared to what is already available. The parameters of each technique were as follows:

Model	Parameters
KNN	*“n_neighbors=20, *, weights='uniform', algorithm='auto', leaf_size=30, p=2, metric='minkowski', metric_params=None, n_jobs=None”*
RF	*“criterion='squared_error', splitter='best', max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=1, max_leaf_nodes=None, min_impurity_decrease=0.0, ccp_alpha=0.0”*
SVR	*“kernel='rbf', degree=3, gamma='scale', coef0=0.0, tol=0.001, C=1.0, epsilon=0.1, shrinking=True, cache_size=200, verbose=False, max_iter=-1”*
GBR	*“loss='squared_error', learning_rate=0.1, n_estimators=100, subsample=1.0, criterion='friedman_mse', min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_depth=3, min_impurity_decrease=0.0, init=None, random_state=None, max_features=None, alpha=0.9, verbose=0, max_leaf_nodes=None, warm_start=False, validation_fraction=0.1, n_iter_no_change=None, tol=0.0001, ccp_alpha=0.0”*

Figure 1: Parameters of Machine Learning models.
Source: Authors, (2023).

Using the Google Colab software with Python language, the results of the performance of the errors for each technique are in the following table:

Table 1: Result of prediction errors by model.

Metrics	KNN	GBM	RF	SVM	LR
MAE	559,94	321,23	409,81	1142,75	660,94
MAPE	0,19	**0,09**	0,11	0,52	0,26
MSE	640197,92	341596,68	842517,78	2108672,17	648472,63
RMSE	800,12	584,46	917,89	1452,13	805,28

Source: Authors, (2023).

Applying the forecasting models to 2823 data points and performing filtering, it was possible to extract, according to Table 1, that the best forecasting technique is GBM (Gradient Boosting Machine). Considering the MAPE (Mean Absolute Percentage Error), which is a model to calculate the error in percentage, it showed an approximate 9% error when applied in these settings.

The following figure presents the performance of demand forecasting using the GBM technique on the test set, representing 30% of the total dataset (846).

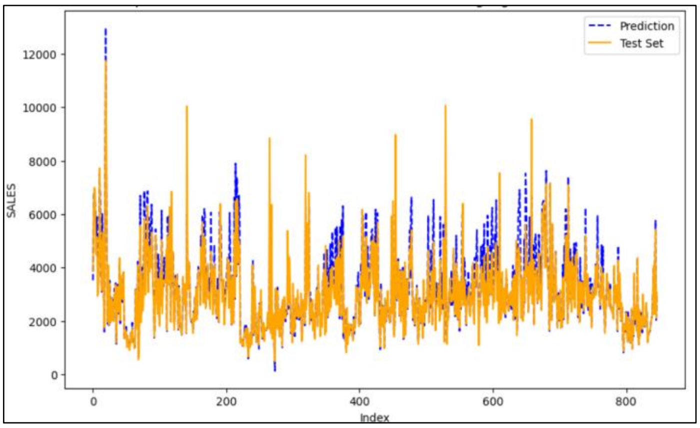

Figure 2: Comparison between the predictions with the best model of the GBM.
Source: Authors, (2023).

In dashed lines, one has the values predicted by the GBM technique and the actual values (only of the test set) in continuous lines. It can be observed that the prediction is close to the actual results of the test set, and as presented, only a 9% error was obtained.

IV.3 MATHEMATICAL MODEL

The following are the parameters of the mathematical model of linear programming. The technique was chosen due to all constraints and objective functions being of the linear type. The parameters of the model are The Costs, divided into cost with normal hours (chn_t), cost with overtime (che_t) and cost with subcontracted hours (chs_t); The Demand, where dt represents the predicted demand using the GBM technique, and Capacity, where C_t represents the productive capacity in normal hours of the period. The decision variables of the mathematical model of linear programming are: A xhn_t, represents the quantity produced of the product using normal hours in the period t; The xhe_t, represents the quantity produced of the product using overtime in the period t; The xhs_t, represents the quantity produced of the product using subcontracted hours in the period t. The following is the mathematical model of minimization of total costs for the planning of the production of the factory:

$$\min CT = \sum_{t=1}^{6} \{chn_t \times xhn_t + che_t \times xhe_t + chs_t \times xhs_t\}$$

s.t:

Inventory balance: $$E_t = E_{t-1} + (xhn_t + xhe_t + xhs_t) - d_t$$

Productive capacity: $$xhn_t \leq C_t$$

Demand: $$xhn_t + xhe_t + xhs_t = d_t$$

$$xhn_t, xhe_t, xhs_t, E_t \geq 0$$

Figure 3: Mathematical model.
Source: Authors, (2023).

The objective function represents the choice of the quantity to be produced using regular hours, overtime hours, and subcontracted hours to minimize the total cost. The first constraint represents the inventory balance, where the current period's inventory (et) is equal to the previous period's inventory (et-1) plus the quantity produced (whether by regular hours, overtime hours, or subcontracted hours) minus the forecasted demand for the period using the GBM technique. The second constraint refers to the production capacity for the period, meaning that production cannot exceed the quantity possible within regular hours. The third constraint relates to the period's demand (dt). This period's demand, estimated by the GBM, must be met with production using regular, overtime, and subcontracted hours. The last constraint is non-negativity, meaning all variables belong to the set of positive real numbers (the first quadrant).

As for the cost parameters, for producing Product 1, the regular hour cost is R$ 27.00, the overtime hour cost is R$ 99.00, and the subcontracting cost is R$ 207.00. For Product 2, the costs are regular hour cost of R$ 19.60, overtime hour cost of R$ 91.60, and subcontracting cost of R$ 199.60.

Regarding the demand parameters, with the choice of the forecasting model, future forecasts can be calculated for six months for Product 1 and Product 2. The calculated demands are as follows: for the first three months of Product 1, 5244 pieces, and the next three months, 5235 pieces; for Product 2, the first three months have a demand of 2036 pieces, and for the next three months, 2027 pieces.

As for the capacity parameters, Product 1 can produce 2217 pieces per month, and Product 2 has a capacity of 5580 pieces per month.

Based on the presented mathematical model and considering the collected input parameters, Microsoft Excel with the Solver add-in was used to obtain the optimal solution for production levels using regular hours, overtime hours, and subcontracted hours. Below is an Excel figure displaying the input parameters and the result of the decision variables for each product.

Parameters	Planning horizon 1	2	3	4	5	6
Production costs (normal hour)	R$ 27,00	R$ 27,00	R$ 27,00	R$ 27,00	R$ 27,00	R$ 27,00
Production costs (overtime)	R$ 99,00	R$ 99,00	R$ 99,00	R$ 99,00	R$ 99,00	R$ 99,00
Production costs (subcontracting)	R$ 207,00	R$ 207,00	R$ 207,00	R$ 207,00	R$ 207,00	R$ 207,00

Period		Planning horizon 1	2	3	4	5	6	Total
Demand		5243,00	5243,00	5243,00	5235,00	5235,00	5235,00	31434,00
Production capacity (TD/TC)		2217,00	2217,00	2217,00	2217,00	2217,00	2217,00	13302,00
Total production		5243,00	5243,00	5243,00	5235,00	5235,00	5235,00	31434,00
Production: regular hours		2217,00	2217,00	2217,00	2217,00	2217,00	2217,00	13302,00
Production: overtime hours		3026,00	3026,00	3026,00	3018,00	3018,00	3018,00	18132,00
Production: subcontrated hours		0,00	0,00	0,00	0,00	0,00	0,00	0,00
Stock level	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
Total cost								2.154.222,00

Figure 4: Model result for product 1.
Source: Authors, (2023).

Parameters	Planning horizon 1	2	3	4	5	6
Production costs (normal hour)	R$ 19,60	R$ 19,60	R$ 19,60	R$ 19,60	R$ 19,60	R$ 19,60
Production costs (overtime)	R$ 91,60	R$ 91,60	R$ 91,60	R$ 91,60	R$ 91,60	R$ 91,60
Production costs (subcontracting)	R$ 199,60	R$ 199,60	R$ 199,60	R$ 199,60	R$ 199,60	R$ 199,60

Period		Planning horizon 1	2	3	4	5	6	Total
Demand		2036,00	2036,00	2036,00	2036,00	2036,00	2036,00	12216,00
Production capacity (TD/TC)		5580,00	5580,00	5580,00	5580,00	5580,00	5580,00	33480,00
Total production		2036,00	2036,00	2036,00	2036,00	2036,00	2036,00	12216,00
Production: regular hours		2036,00	2036,00	2036,00	2036,00	2036,00	2036,00	12216,00
Production: overtime hours		0,00	0,00	0,00	0,00	0,00	0,00	0,00
Production: subcontrated hours		0,00	0,00	0,00	0,00	0,00	0,00	0,00
Stock level	0,00	0,00	0,00	0,00	0,00	0,00	0,00	
Total cost								R$ 239.433,60

Figure 5: Model result for product 2.
Source: Authors, (2023).

To exemplify the analysis of the results, for product 1 with the aid of the solver tool, the optimal solution was to manufacture 2217 pieces using normal hours and 3026 pieces using overtime to meet the demand of 5243 pieces, considering only the first period. At the end of the six periods, the total produced 13302 pieces using normal hours and 18132 using overtime. This decision had a total cost of R$ 2,154,222.00. The same idea is replicated in the analysis of product 2.

The next step was to send the results obtained from the model for analysis by the company. After a certain period, the feedback was positive with some considerations, such as incorporating more products and production lines in the mathematical model to represent reality better.

V. DISCUSSIONS

The results obtained by this research reveal a significant impact, providing practical support to production managers in optimizing factory organization. The focus was on improving the monthly production of two key products, with particular attention to cost minimization. The research successfully predicted sales quantities, enabling an accurate view of future demands based on Machine Learning techniques.

An important observation made by the company's stakeholders is the need to strengthen the model further. This implies expanding the mathematical model and the machine learning forecasts. This quest for robustness suggests including a wider range of parameters and variables, allowing the model to capture the inherent complexities of the production process more comprehensively.

Furthermore, this research adds valuable contributions to the literature. The use of an innovative dataset, framed within the literature review presented by [4], stands out as a distinctive element of this study. Combining effective Operations Research and Machine Learning approaches is a significant step. This fusion of strong disciplines offers remarkable potential for addressing the intricate challenges of planning and optimization.

The problem question of the article, which considered the link between operations research and Machine Learning for production optimization, was validated through the positive results of this study. This proves that the joint application of these approaches can generate highly effective results and be oriented towards improving operational efficiency.

In conclusion, the results of this research not only provide practical insights to the company but also enrich the literature by exploring innovative interdisciplinary approaches to optimize production planning [15][16][17]. The continued pursuit of greater robustness and the continued application of these approaches can give the organization a lasting competitive advantage, enhancing its operational effectiveness and ability to make informed decisions more efficiently [18].

VI. CONCLUSION

The results obtained through machine learning and the mathematical linear programming models have fully achieved the objectives outlined in this research. These results have provided crucial insights to guide the company's decision-making process based on carefully considering the selected parameters and variables. The application of the machine learning model was executed using the Python programming language in Google Colab, while optimization through the linear programming model was performed using the Excel tool. It is worth noting that both selected tools are widely accessible, eliminating the need to acquire new software and aligning with the organization's daily practices.

The results generated by the mathematical model offered an accurate perspective on the required quantities of regular hours, overtime hours, and subcontracted hours to produce each product to meet the forecasted demands. Product 1 demonstrated the need for both hours due to its high demand and limited production capacity. In contrast, Product 2 relied solely on regular hours, reflecting its lower demand and higher production capacity.

Although the study presents inherent limitations, such as building models based on specific parameters, the complexity of the real operational scenario, and the quality of available data, it provides a solid foundation for future improvements. Incorporating stakeholder feedback emerges as a promising path for adding additional variables and parameters, aiming to represent the organization's dynamics accurately. As we look ahead, exploring various combinations of machine learning techniques proves to be a promising direction to further enrich the operational and strategic effectiveness of the company in the context of production planning [19][20][21]. In summary, this study has fully achieved its specific objectives, highlighting the fundamental potential of the developed models to enhance decision-making practices and organizational efficiency continuously.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Lucas Vianna Vaz.
Methodology: Lucas Vianna Vaz.
Investigation: Lucas Vianna Vaz.
Discussion of results: Marcelo Carneiro Gonçalves.
Writing – Original Draft: Izamara Cristina Palheta Dias.
Writing – Review and Editing: Izamara Cristina Palheta Dias.
Resources: Elpídio Oscar Benitez Nara.
Supervision: Elpídio Oscar Benitez Nara.
Approval of the final text: Lucas Vianna Vaz, Marcelo Carneiro Gonçalves, Izamara Cristina Palheta Dias and Elpídio Oscar Benitez Nara.

VIII. ACKNOWLEDGMENTS

The development of this research received funds from the Coordination Foundation for the Improvement of Higher Education Personnel (CAPES) and the Pontifical Catholic University of Paraná (PUCPR).

IX. REFERENCES

[1] I.C. Baierle, J.L. Schaefer, M.A. Sellitto, J.C. Furtado, and E.O.B. Nara, "Moona software for survey classification and evaluation of criteria to support decision-making for properties portfolio", International Journal of Strategic Property Management, vol.24, no.2, pp.226–236, 2020. DOI: https://doi.org/10.3846/ijspm.2020.12338

[2] L.D.D. Nora, J.C.M. Siluk, A.L.N. Júnior, E.O.B. Nara, and J.C. Furtado, "The performance measurement of innovation and competitiveness in the telecommunications services sector". International Journal of Business Excellence, 9(2), pp.210–224, 2016. DOI: https://doi.org/10.1504/IJBEX.2016.074844.

[3] J.L. Schaefer, I.C. Baierle, M.A. Sellitto, J.C. Furtado, and E.O.B. Nara, "Competitiveness Scale as a Basis for Brazilian Small and Medium-Sized Enterprises". EMJ - Engineering Management Journal, vol.33. no. 4, pp.255–271 2021. DOI: https://doi.org/10.1080/10429247.2020.1800385.

[4] A. S. Pereira, M. C. Goncalves, E. O. B. Nara, andT. F. Ferraz, "Application of machine learning methods in forecasting sales prices in a project consultancy". Produto e Produção. vol. 24, no. 2, pp. 1-12, 2023. DOI: https://doi.org/10.22456/1983-8026.127422.

[5] M. C. Gonçalves, E. O. B. Nara, I. M. Dos Santos, I. B. Mateus, and L. M. B. do Amaral, "Comparative analysis of machine learning techniques via data mining in a railroad company In Deschamps, F.; Pinheiro de Lima, E.; Gouvêa da Costa, S.E.; G. Trentin, M. (Org.)". Proceedings of the 11th International Conference on Production Research Americas. ICPR 2022. 1ed.Berlin: Springer Nature, vol. 1, pp. 655-664, 2022. Doi: 10.1007/978-3-031-36121-0_83.

[6] M. Goncalves, A. Pereira, T. Ferraz, E. Nara, and I. Dias, "Predicting project sales prices using machine learning techniques: a case study in a project consultancy". International Joint Conference on Industrial Engineering and Operations Management, IJCIEOM 2022. Springer Proceedings in Mathematics and Statistics 2023. Doi: http://dx.doi.org/10.14488.

[7] B. Topu, and N. Cakici Alp, "Machine learning in architecture", vol. 154, 2023.

[8] S. Pandey, N. J. Basisth, T. Sachan, N. Kumari and P. Pakray, "Quantum machine learning for natural language processing application", Physic A: Statistical Mechanics and its applications, vol. 627, 2023. doi: https://doi.org/10.1016/j.physa.2023.129123

[9] H. Frederick and J. Lieberman. "Introduction to operations research". Available at: My Library, (9th edition). Group A, 2013.

[10] K. Hamasaki, M.C. Gonçalves, O. C. Junior, E. O. B. Nara, and R.R.G Wollmann, "Robust linear programming application for the production planning problem" In Deschamps, F.; Pinheiro de Lima, E.; Gouvêa da Costa, S.E.; G. Trentin, M. (Org.). Deschamps, F., Pinheiro de Lima, E., Gouvêa da Costa, S.E., G. Trentin, M. (eds) Proceedings of the 11th International Conference on Production Research Americas. ICPR 2022.. 1ed.Berlin: Springer Nature, 2022, vol. 1, pp. 647-654, 2022 Doi: 10.1007/978-3-031-36121-0_82.

[11] I. Dias, R. Sampaio, R. Wollmann, M. Goncalves, and E. Nara, "A Decomposition Scheme in Production Planning Based on Linear Programming that Incorporates the Concept of a Dynamic Planning Environment". International Joint Conference on Industrial Engineering and Operations Management, IJCIEOM 2022. Springer Proceedings in Mathematics and Statistics, 2023. http://dx.doi.org/10.14488/IJCIEOM2023_FULL_0031_37690.

[12] M. Goncalves, R. Sampaio, R. Wollmann, E. Nara, and I. Dias, "Using robust approach concept to solve the production planning problem in manufacturing systems". International Joint Conference on Industrial Engineering and Operations Management, IJCIEOM 2022. Springer Proceedings in Mathematics and Statistics 2023. http://dx.doi.org/10.14488/IJCIEOM2023_FULL_0031_37686.

[13] M.C. Gonçalves, R. R. G. Wollmann, and R. J. B. Sampaio, "Proposal of a Numerical Approximation Theory to Solve the Robust Convex Problem of Production Planning", Internacional Journal of Operational Research, 2023. Doi: 10.1504/IJOR.2022.10049618.

[14] M. Gonçalves, A. Canciglieri, K. Strobel, M. Antunes, and R. Zanellato, "Application of operational research in process optimization in the cement industry", JETIA, vol. 6, no. 24, pp. 36-40, Aug. 2020. Doi: https://doi.org/10.5935/jetia.v6i24.677.

[15] G. de Faria, J. Tulik, and M. Gonçalves, "Proposition of A Lean Flow of Processes Based on The Concept of Process Mapping For A Bubalinocultura Based Dairy", JETIA, vol. 5, no. 18, pp. 23-28, Jun. 2019. Doi 10.5935/2447-0228.20190022 (2019).

[16] O. J. Junior and M. Gonçalves, "Application of Quality and Productivity Improvement Tools in A Potato Chips Production Line", JETIA, vol. 5, no. 18, pp. 65-72, Jun. 2019.

[17] F. Serra, E. Nara, M. Gonçalves, S. Costa, and S. Bortoluzzi, "Preliminary construct for decision making in organizations: a systemic approach". In: Deschamps, F.; Pinheiro de Lima, E.; Gouvêa da Costa, S.E.; G. Trentin, M. (Org.). Proceedings of the 11th International Conference on Production Research Americas, 1ed.Berlin: Springer Nature, vol. 1, pp. 639-646, 2022. Doi: 10.1007/978-3-031-36121-0_81.

[18] F. Lourenço, E. Nara, M. Goncalves, O. Canciglieri, "Preliminary Construct of Sustainable Product Development with a Focus on the Brazilian Reality: A Review and Bibliometric Analysis". In: Leal Filho, W.; Frankenberger, F.;Tortato, U.. (Org.). Sustainability in Practice. 1ed.Suiça: Springer Nature, vol. 1, pp. 197-220, 2023. Doi: 10.1007/978-3-031-34436-7_12.

[19] P.R.Tardio, J.L. Schaefer, E.O.B. Nara, M.C. Gonçalves, I.C.P. Dias, G.B. Benitez, and A. Castro e Silva, "The link between lean manufacturing and Industry 4.0 for product development process: a systemic approach", Journal of Manufacturing Technology Management, Vol. ahead-of-print No. ahead-of-print. 2023. https://doi.org/10.1108/JMTM-03-2023-0118.

[20] G. Cardoso and I. C. Dias, "Mapping Process Improvement and Sequencing Analysis for Productive Definitions:" JETIA, vol. 6, no. 21, pp. 66-71, Feb. 2020.

[21] F. Stankevecz and I. C. Dias, "System Integrated Management for Stock Management in a Beverage Distributor: A Proposal Based on A Case Study", JETIA, vol. 5, no. 18, pp. 58-64, Jun. 2019.

X. ANNEX

Appendix A – List of literature review articles.

Index	Authors	Title	Machine Learning Models	Contributions	Area
1	Xu, X., Fairley, C.K., Chow, E.P.F., Lee, David., Aung, Ei T., Zhang, L., Ong, J.J.	Using machine learning approaches to predict timely clinic attendance and the uptake of HIV/STI testing post-clinic reminder messages	logistic regression, lasso regression, ridge regression, elastic net regression, support vector machine, k-nearest neighbour, naïve Bayes, random forest, Gradient boosting machine, XGBoost, and multilayer perceptron.	The machine learning approach helps predict timely clinic attendance and HIV/STI re-testing. The predictive models could be incorporated into clinic websites to inform sexual health care or follow-up services.	Healthcare
2	Jiang, K., Liang, Y., Zhao, O.	Machine-learning-based design of high-strength steel bolted connections	Decision Tree, Random Forest, Support Vector Machine, K-Nearest Neighbour, Adaptive Boosting, Light Gradient Boosting Machine, Extreme Gradient Boosting and Cat Boosting.	Using a machine learning framework to predict the total failure modes. The machine-learning-based approach can accurately predict 97.2%, while the standard framework can only accurately predict 67.9%-85.3%.	Civil Engineering
3	Choi, T.-J., An, H.-E., Kim, C.-B.	Machine Learning Models for Identification and Prediction of Toxic Organic Compounds Using Daphnia magna Transcriptomic Profiles	Learning Vector Quantization, Random Forest, Support Vector Machines with a Linear kernel, Linear Discriminant Analysis, Classification And Regression Trees, K-nearest neighbours, Boosted C5.0, Gradient Boosting Machine, eXtreme Gradient Boosting with tree, and eXtreme Gradient Boosting with DART booster	Study to establish a model for monitoring aquatic toxic substances by machine learning. The best ensemble model had an accuracy of 95.7%. This model could be an effective tool to manage contaminants and toxic organic compounds in aquatic systems.	Biotechnology
4	Kim, Y.J.	Machine Learning Model Based on Radiomic Features for Differentiation between COVID-19 and Pneumonia on Chest X-ray	logistic regression, naive Bayes, support vector machine, k-nearest neighbour, bagging, random forest, extreme gradient boosting, and light Gradient boosting machine.	The study confirmed that the radiomic features in chest X-rays can be used as indicators to differentiate between COVID-19 and pneumonia using machine learning.	Biomedical
5	Song, J., Huang, F., Chen, L., Feng, KaiYan., Jian, Fangfang., Huang, T., Cai, Y.-D.	Identification of methylation signatures associated with CAR T cell in B-cell acute lymphoblastic Leukemia and non-Hodgkin's lymphoma	Monte Carlo feature selection, light gradient boosting machine and least absolute shrinkage and selection operator	Using advanced machine learning approaches to the high-throughput data, investigating the mechanism of CAR T cells to establish the theoretical foundation for modifying CAR T cells.	Biotechnology
6	Ljubobratović, D., Vuković, M., Brkić Bakarić, M., Jemrić, T., Matetić, M.	Assessment of Various Machin e Learning Models for Peach Maturity Prediction Using Non-Destructive Sensor Data	least absolute shrinkage and selection operator, artificial neural network, linear discriminant analysis, logistic regression, Gradient boosting machine, random forest, support vector machines, a classification and regression trees model, and k-nearest neighbours.	The Artificial Neural Network model proved the most accurate for preaching maturity prediction on the given dataset (AUC of 0.766).	Biotechnology
7	Tsutsui, K., Matsumoto, K., Maeda, M., Takatsu, Terusato.,Moriguchi, Koji.,Hayashi, Kohtaro., Morito, S., Terasaki, H.	Mixing effects of SEM imaging conditions on convolutional neural network-based low-carbon steel classification	ResNet50, support vector machine, k-nearest neighbour, random forest, Gradient boosting machine, and multilayer perceptron.	Using machine learning models to identify the SEM sources for the images of the microstructures. Accuracies are in the range between 0.91 and 0.96.	Mechanical Engineering
8	Shah, A.A., Devana, S.K., Lee, C., Bugarin, Amador.,; Lord, Elizabeth.,	Machine learning-driven identification of novel patient	decision tree, Gradient boosting machine, k nearest neighbour, logistic regression, random forest, and support vector machine.	An ensemble Machine learning model for predicting major complications and readmission after posterior cervical fusion	Health care

Index	Authors	Title	Machine Learning Models	Contributions	Area
	Shamie, Arya., Park, Don., van der Schaar, M., SooHoo, N.F.	factors for prediction of major complications after posterior cervical spinal fusion		with a modest risk prediction advantage compared to logistic regression and benchmark Machine learning models.	
9	Tiwari, A., Chugh, A., Sharma, A.	Ensemble framework for cardiovascular disease prediction	ExtraTrees Classifier, Random Forest, k nearest neighbour, support vector machine and XGBoost.	A model that can predict cardiovascular disease before it becomes a critical situation. The highest accuracy was 92.34%.	Health care
10	Mao, Y., Huang, Y., Xu, L., Liang, J., Lin, W., Huang, H., Li, L., Wen, J., Chen, G.	Surgical Methods and Social Factors Are Associated With Long-Term Survival in Follicular Thyroid Carcinoma: Construction and Validation of a Prognostic Model Based on Machine Lea rning Algorithms	eXtreme Gradient Boosting, Light Gradient Boosting Machine, Random Forests, Logistic Regression, Adaptive Boosting, Gaussian Naive Bayes, K-Nearest Neighbor, Support Vector Machine and Multilayer Perceptron.	Build a machine learning model to predict the prognosis of follicular thyroid cancer (FTC). The XGBoost model has relatively better prediction accuracy and clinical usage	Health care
11	Mao, N., Shi, Y., Lian, C., Wang, Z., Zhang, K., Xie, H., Zhang, H., Chen, Q, Cheng, G., Xu, C., Dai, Y.	Intratumoral and peritumoral radiomics for preoperative prediction of neoadjuvant chemotherapy effect in breast cancer based on contrast-enhanced spectral mammography	gradient boosting machine, k nearest neighbour, least absolute shrinkage and selection operator, random forest, and support vector machine.	Using machine learning to investigate the performance of intratumoral and peritumoral radiomics based on contrast-enhanced spectral mammography (CESM) to preoperatively predict the effect of the neoadjuvant chemotherapy (NAC) of breast cancers	Medicine
12	Wie, J.H., Lee, S.J., Choi, S.K., Jo, Y.S., Hwang, H.S, Park, M.H., Kim, Y.H., Shin, Ko, HS, Na, S.	Prediction of Emergency Cesarean Section Using Machine Learning Methods: Development and External Validation of a Nationwide Multicenter Dataset in the Republic of Korea	logistic regression, random forest, Support Vector Machine, Gradient boosting, extreme gradient boosting, light Gradient boosting machine, k-nearest neighbour.	Machine learning algorithms with clinical and sonographic parameters in the near term could be useful tools to predict individual risks of emergent cesarean section during active labour in nulliparous women.	Medicine
13	Alfi, I.A., Rahman, M.M., Shorfuzzaman , M., Nazir, A.	A Non-Invasive Interpretable Diagnosis of Melanoma Skin Cancer Using Deep Learning and Ensemble Stacking of Machine Lea rning Models	logistic regression, support vector machine, random forest, k- k-nearest neighbor, Gradient boosting machine, MobileNet, Xception, ResNet50, ResNet50V2, and DenseNet121.	Deep learning models and machine learning models play an essential role in the detection of skin lesions. This study introduces an interpretable method for the non-invasive diagnosis of melanoma skin cancer using deep learning and ensemble stacking of machine learning models.	Health care
14	Patel, D., Hall, G.L., Broadhurst, D., Smith, A., Schultz, A., Foong, RE.	Does machine learning have a role in the prediction of asthma in children?	Gradient boosting machine, k- k-nearest neighbor, random forest, and support vector machine.	Asthma prediction traditional tools use conventional statistical models with modest accuracy, sensitivity, and positive predictive value. Few studies have utilized machine learning as an approach. This study is a review of these studies.	Medicine
15	Cai, Y., Xu, D., Shi, H.	Rapid identification of	k-nearest neighbor, support vector machine, random forest,	This paper demonstrates an extremely fast and accurate	Spectroscopy

Index	Authors	Title	Machine Learning Models	Contributions	Area
		ore minerals using multi-scale dilated convolutional attention network associated with portable Raman spectroscopy	cosine similarity, extreme Gradient boosting machine, Alexnet and ResNet 18	method (using machine learning) for identifying unknown ore mineral samples by portable Raman spectroscopy.	
16	Mishra, A.K., Paliwal, S.	Mitigating cyber threats through integration of feature selection and stacking ensemble learning: the LGBM and random forest intrusion detection perspective	light Gradient boosting machine, random forest, stochastic gradient descent, Gaussian Naive Bayes, support vector machine, bagging + reduced error pruning, K nearest neighbour and AdaBoost.	Machine learning-based approaches successfully quell modern-day attacks by analyzing the patterns in the encrypted network traffic—a stacking of Light gradient boosting machine and random forest given the highest predictions.	Computer Science
17	Andrian, B., Simanungkalit, T., Budi, I., Wicaksono, A.F.	Sentiment Analysis on Customer Satisfaction of Digital Banking in Indonesia	naïve Bayes, Logistic Regression, k-nearest neighbours, support vector machines, Random Forest, Decision Tree, Adaptive Boosting, eXtreme Gradient Boosting and Light Gradient Boosting Machine	This research aims to obtain customer satisfaction with digital banking in Indonesia based on sentiment analysis from Twitter. The results of this study show that SVM, among other stand-alone classifiers, performs best when used to predict sentiments, with a value score of 73.34%.	Computer Science
18	Lee, Y.-H., Tsai, T.-H., Chen, J.-H., Huang, C.-J, Chen, C.-H., Cheng, H.-M.	Machine learning of treadmill exercise test to improve selection for testing for coronary artery disease	Support vector machine, logistic regression, random forest, k-nearest neighbor and extreme Gradient boosting machine	Using the information obtained from conventional treadmill exercise tests, a more accurate diagnosis can be made by incorporating an artificial intelligence-based model.	Health Care
19	Hasan, M.S., Kordijazi, A., Rohatgi, P.K., Nosonovsky, M.	Triboinformatics approach for friction and wear prediction of Al-graphite composites using machine learning methods	artificial neural network (ANN), K nearest neighbor (KNN), support vector machine (SVM), Gradient boosting machine (GBM), and random forest (RF).	Machine learning models can predict tribological behaviour from material variables and test conditions. The analysis identified graphite content and hardness as the most significant variables in predicting the COF, while graphite content and sliding speed were the most dominant variables for wear rates.	Mechanical Engineering
20	Olier, I., Orhobor, O.I., Dash, T., David, A.M., Soldatova, L., Vanschoren, J., King, R.D.	Transformational machine learning: Learning how to learn from many related scientific problems	Random forests, Gradient boosting machines, support vector machines, k-nearest neighbors, and neural networks	Using Transformational Machine Learning (TML) to understand and improve the performance of machine learning models. The study found that TML significantly improved the predictive performance of all the ML methods in all the domains (4 to 50% average improvements) and that TML features generally outperformed intrinsic features.	Computer Science
21	Kingsmore, K.M., Puglisi, C.E., Grammer, A.C., Lipsky, P.E.	An introduction to machine learning and analysis of its use in rheumatic diseases	gradient boosting machine, random forest, support vector machine, k nearest neighbor	The study does a review that introduces the basic principles of ML and discusses its current strengths and weaknesses in the classification of patients with rheumatic autoimmune inflammatory diseases (RAIDs)	Biomedicine
22	Jen, K.-Y., Albahra, S., Yen, F., Sageshima, J., Chen, L-.X., Tran, N., Rashidi, H.H.	Automated en masse machine learning model generation shows comparable	gradient boosting machine, k-nearest neighbor, logistic regression, neural network, naive Bayes, random forest, support vector machine	The automated en masse machine learning modelling approach rapidly generated machine learning models for DGF prediction. The performance of the machine	Medicine

Index	Authors	Title	Machine Learning Models	Contributions	Area
		performance as classic regression models for predicting delayed graft function in renal allografts		learning models was comparable with classic logistic regression models.	
23	Shim, J.-G., Ryu, K.-H., Cho, E.-A., Kim, H.K., Lee, Y.-J., Lee, S.H.	Machine learning approaches to predict chronic lower back pain in people aged over 50 years	logistic regression (LR), k-nearest neighbors (KNN), naïve Bayes (NB), decision tree (DT), random forest (RF), Gradient boosting machine (GBM), support vector machine (SVM), and artificial neural network (ANN).	The ANN model was identified as the best machine learning classification model for predicting the occurrence of Chronic Lower Back Pain (LBP). Therefore, machine learning could be effectively applied in identifying populations at high risk of chronic LBP.	Health Care
24	Agrawal, S., Sisodia, D.S., Nagwani, N.K.	Augmented sequence features and subcellular localization for functional characterization of unknown protein sequences	Decision tree (C 4.5), k-nearest neighbor (k-NN), multilayer perceptron (MLP), Naïve Bayes (NB), support vector machine (SVM), AdaBoost, Gradient boosting machine (GBM), and random forest (RF)	This paper describes two feature augmentations through sequence-induced, physicochemical, and evolutionary information on the amino acid residues.	Medicine / Protein sequeces
25	Islam, M.A., Subramanyam Rallabandi, V.P., Mohammed, S., Srinivasa, S., Natarajan, S., Dudekula, D.B., Park, J.	Screening of β1- and β2-adrenergic receptor modulators through advanced pharmacy informatics and machine learning approaches	gradient boosting machine, k nearest neighbor, support vector machine, decision tree, random forest	In the present study, structure-based virtual screening, machine learning, and a ligand-based similarity search were conducted for the PubChem database against both β1-and β2-AR(β-Adrenergic receptors)	Biotechnology
26	Achu, A.L., Thomas, J., Aju, C.D., Gopinath, G., Kumar, S., Reghunath, R.	Machine-learning modelling of fire susceptibility in a forest-agriculture mosaic landscape of southern India	artificial neural network (ANN), generalized linear model (GLM), multivariate adaptive regression splines (MARS), Naïve Bayesian classifier (NBC), K-nearest neighbour (KNN), support vector machine (SVM), random forest (RF), gradient boosting machine (GBM), adaptive boosting (AdaBoost) and maximum entropy (MaxEnt)	The study proposes a weighted approach to characterize the forest fire susceptibility of the region using the outputs of the different Machine Learning Techniques. Besides that, this study suggests that roughly one-third of the study area is highly susceptible to the occurrence of forest fires, implying the severity of the disturbance regime.	Agriculture / geospatial data
27	Hasan, M.S., Kordijazi, A., Rohatgi, P.K., Nosonovsky, M.	Triboinformatic modelling of dry friction and wear of aluminium base alloys using machine learning algorithms	K Nearest Neighbor (KNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), and Gradient Boosting Machine (GBM)	The study demonstrated that the machine learning models could satisfactorily predict friction and wear of aluminium (Al) alloys from material and tribological test variables data.	Mechanical Engineering
28	Krishnan, P.T., Joseph Raj, A.N., Rajangam, V.	Emotion classification from speech signal based on empirical mode decomposition and non-linear features: Speech emotion recognition	Linear Discriminant Analysis (LDA), Naïve Bayes, K-Nearest Neighbor, Support Vector Machine, Random Forest, and Gradient Boosting Machine	Created a machine learning model to recognize the emotional states from speech signals. The model presents a peak balanced accuracy of 93.3%, an F1 score of 87.9%, and an area under the curve value of 0.995 in recognising emotions from speech signals of native English speakers.	Computer Science
29	Xu, Z., Kurek, A., Cannon, S.B., Beavis, W.D.	Predictions from algorithmic modelling result in better decisions than	Random Forest, Gradient Boosting Machine, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes, and Artificial Neural Network. We found that a Support Vector Machine model.	Developed a model to compare and select continuous plant traits.	Agriculture

Index	Authors	Title	Machine Learning Models	Contributions	Area
		from data modelling for soybean iron deficiency chlorosis			
30	Wu, Z., Zhu, M., Kang, Y., Leung, E. L-H., Lei, T., Shen, C., Jiang, D., Wang, Z., Cao, D., Hou, T.	Do we need different machine learning algorithms for QSAR modelling? A comprehensive assessment of 16 machine learning algorithms on 14 QSAR data sets	[linear function Gaussian process regression (linear-GPR), linear function support vector machine (linear-SVM), partial least squares regression (PLSR), multiple linear regression (MLR) and principal component regression (PCR)], [radial basis function support vector machine (RBF-SVM), K-nearest neighbor (KNN) and radial basis function Gaussian process regression (rbf-GPR)], [extreme gradient boosting (XGBoost), Cubist, random forest (RF), multiple adaptive regression splines (MARS), Gradient boosting machine (GBM), and classification and regression tree (CART)], [principal component analysis artificial neural network (pca-ANN) and deep neural network (DNN)]	Some machine learning models were employed to learn the regression-based quantitative structure-activity relationships (QSAR) models for 14 public data sets comprising nine physicochemical properties and five toxicity endpoints.	Medicine
31	Hussain, A., Choi, H.-E., Kim, H.-J., Aich, S., Saqlain, M., Kim, H.-C.	Forecast the exacerbation in patients of chronic obstructive pulmonary disease with clinical indicators using machine learning techniques	random forests (RF), support vector machine (SVM), Gradient boosting machine (GBM), XGboost (XGB), and K-nearest neighbor (KNN).	This paper proposes a voting ensemble classifier with 24 features to identify the severity of chronic obstructive pulmonary disease patients.	Health Care
32	Dos Santos Santana, I.V., Da Silveira, A.C.M., Sobrinho, A., Silva, L.C.E.,Da Silva, L.D., Gurjão, E.C., Perkusich, A.	Classification models for COVID-19 test prioritization in Brazil: Machine learning approach	Supervised learning; and the algorithms multilayer perceptron (MLP), Gradient boosting machine (GBM), decision tree (DT), random forest (RF), extreme gradient boosting (XGBoost), k-nearest neighbors (KNN), support vector machine (SVM), and logistic regression (LR)	The Decision Tree classification model can effectively (with a mean accuracy of 89.12%) assist COVID-19 test prioritization in Brazil. The model can be applied to recommend prioritizing a patient who is symptomatic for COVID-19 testing.	Health Care
33	Aktar, S., Ahamad, M.M., Rashed-Al-Mahfuz, M., Azad, A.K.M., Uddin, S., Kamal, A.H.M., Alyami, S. A., Lin, P-I., Islam, S.M.S, Quinn, J.M.W., Eapen, V., Moni, M.A.	Machine learning approach to predicting COVID-19 disease severity based on clinical blood test data: Statistical analysis and model development	decision tree, random forest, variants of Gradient boosting machine, support vector machine, k-nearest neighbor, and deep learning methods.	This paper revealed that several measurable clinical parameters in blood samples are factors that can discriminate between healthy people and COVID-19-positive patients. Besides that, this paper showed the value of these parameters in predicting the later severity of COVID-19 symptoms.	Health Care
34	Liu, X., Tian, Z., Chen, C.	Total Organic Carbon Content Prediction in Lacustrine Shale Using Extreme Gradient Boosting Machine Learning Based on Bayesian Optimization	random forest, support vector machine, K-nearest neighbors, and multiple linear regression.	This study proposed an approach that was applied to predict the total organic carbon (TOC) curves of 20 exploration wells in the Damintun Sag. Besides that, it obtained quantitative contour maps of the TOC content of this block for the first time. The results of this study facilitate the rapid detection of the sweet spots of the lacustrine shale oil.	Geology
35	Deif, M.A., Hammam, R.E., Solyman, A.A.A.	Gradient Boosting Machine Based on PSO for Prediction of Leukemia after a Breast Cancer Diagnosis	Gradient Boosting Machine (GBM), KNN (k-Nearest Neighbor), SVM (Support Vector Machine), and RF (Random Forest).	The results proved the implemented Classifier's abClassifierlassify breast cancer disease and predict patients with Leukemia developed after having breast cancer. These results are promising as they show the integral role of the GBM classifier in classifying and predicting the tumour with high	Health Care

Index	Authors	Title	Machine Learning Models	Contributions	Area
				accuracy and efficiency, which will further help in better cancer diagnosis and treatment of the disease.	
36	Verma, R., Maheshwari, S., Shukla, A.	Feature engineering combined with a 1-D convolutional neural network for improved mortality prediction	XGBoost classifier, Light GradientClassifierMachine (LGBM) classifier, Support Vector Machine (SVM), Decision Tree (DT), K-Neighbours Classifier (K-NN), Random Forest Classifier (RF) and Long Short-Term Memory (LSTM).	The objective of the research is to utilize the relations among the clinical variables and construct new variables with a Dimensional Convolutional Neural Network (1-D CNN).	Health Care
37	Shim, J.-G., Kim, D.W., Ryu, K.-H., Cho, E-A., Ahn, J-H., Kim, J.-I., Lee, S.H.	Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women	k-nearest neighbors (KNN), decision tree (DT), random forest (RF), Gradient boosting machine (GBM), support vector machine (SVM), artificial neural networks (ANN), and logistic regression (LR)	This study developed and compared seven machine learning models to accurately predict osteoporosis risk. The ANN model performed best compared to the other models, having the highest AUROC value. Applying the ANN model in the clinical environment could help primary care providers stratify osteoporosis patients and improve osteoporosis prevention, detection, and early treatment.	Health Care
38	Nusinovici, S., Tham, Y.C., Chak Yan, M.Y., Wei Ting, D.S., Li, J., Sabanayagam, C., Wong, T.Y., Cheng, C.-Y.	Logistic regression was as good as machine learning for predicting major chronic diseases	ingle-hidden-layer neural network, support vector machine, random forest, Gradient boosting machine, and k-nearest neighbor	Logistic regression performs as well as machine learning models to predict the risk of major chronic diseases with low incidence and simple clinical predictors.	Health Care
39	Hou, P., Jolliet, O., Zhu, J., Xu, M.	Using machine learning models, estimate ecotoxicity characterization factors for chemicals in life cycle assessment.	Decision trees, Random forests, adaptive boosting, Gradient boosting machine, k nearest neighbor, support vector machine.	This study develops machine learning models to estimate ecotoxicity hazardous concentrations 50% (HC50) in USEtox to calculate chemical characterisation factors based on their physical-chemical properties in EPA's CompTox Chemical Dashboard and their mode of action classification.	Sustainability
40	Ribeiro, M.H.D.M., dos Santos Coelho, L.	Ensemble approach based on bagging, boosting and stacking for short-term prediction in agribusiness time series	Random forests, Gradient boosting machine, extreme Gradient boosting machine, support vector machine, k-nearest neighbor, multilayer perceptron neural network.	The ensemble approach presents statistically significant gains, reducing prediction errors for the price series studied. Ensembles are recommended to forecast agricultural commodities prices one month ahead since a more assertive performance is observed, which increases the accuracy of the constructed model and reduces decision-making risk.	Agriculture
41	Zhang, Y., Xie, R., Wang, J., Chou, K.-C., Song, J.	Computational analysis and prediction of lysine malonylation sites by exploiting informative features in an integrative machine-learning framework	random forest, support vector machines, K-nearest neighbor, logistic regression and Light Gradient Boosting Machine (LightGBM).	This study reviews, analyzes and compares 11 different feature encoding methods to extract key patterns and characteristics from residue sequences of Kmal sites.	Medicine
42	Cho, G., Yim, J., Choi, Y., Ko, J., Lee, S.-H.	Review of machine learning algorithms for diagnosing mental illness	Support Vector Machines (SVM), Gradient Boosting Machine (GBM), Random Forest, Naïve Bayes, and K-Nearest Neighborhood (KNN).	This paper provides useful information on the properties and limitations of each ML algorithm in mental health practice.	Health Care

Index	Authors	Title	Machine Learning Models	Contributions	Area
43	Lu, Y., Yan, H., Zhang, L., Liu, J.	A Comparative Study on the Prediction of Occupational Diseases in China with Hybrid Algorithm Combing Models	KNN, SVM, RF, GBM, and ANN.	A machine learning model can be used to precisely predict occupational diseases in China, which may provide valuable information for the future prevention and control of occupational diseases.	Sustainability
44	Ashraf, I., Hur, S., Park, Y.	MagIO: Magnetic field strength-based indoor-outdoor detection with a commercial smartphone	Naive Bayes (NB), Support Vector Machine (SVM), Random Induction (RI), Gradient Boosting Machine (GBM), Random Forests (RF), K-Nearest Neighbor (kNN) and Decision Trees(DT).	This approach can achieve an accuracy of 85.30% using the magnetic data of the smartphone magnetic sensor. Moreover, with increased training data, the accuracy of the stacking scheme can be elevated by 0.83%. The performance of the proposed approach is compared with GPS-, Wi-Fi- and light sensor-based IO detection.	Electronic Engineering
45	Miettinen, O.	Protostellar classification using supervised machine learning algorithms	decision tree, random forest, Gradient boosting machine (GBM), logistic regression, naïve Bayes classifier, k-nearest neighbour classifier, support vector machine, and neural network	The application of machine learning is expected to be very useful in the era of big astronomical data, for example, to assemble interesting target source samples for follow-up studies quickly.	Astronomy
46	Kastrin, A., Ferk, P., Leskošek, B.	Predicting potential drug-drug interactions on topological and semantic similarity features using statistical learning	Classification tree, k-nearest neighbors, support vector machine, random forest, and Gradient boosting machine.	The applied methodology can be used to help researchers identify potential drug-drug interactions (DDIs). The supervised link prediction approach proved promising for potential DDI prediction and may facilitate the identification of potential DDIs in clinical research.	Biostatistics
47	Kendale, S., Kulkarni, P., Rosenberg, A.D., Wang, J.	Supervised Machine-learning Predictive Analytics for Prediction of Postinduction Hypotension	Support Vector Machines, Naive Bayes, K-nearest Neighbor, Linear Discriminant Analysis, Random Forest and Gradient Boosting Machine.	This technique in predicting postinduction hypotension demonstrates the feasibility of machine-learning models for predictive analytics in anesthesiology, with performance dependent on model selection and appropriate tuning.	Medicine