Predicting project sales prices using machine learning techniques: a case study in a project consultancy

Marcelo Carneiro Gonçalves[1[0000-0002-4957-6057]], Alexandre dos Santos Pereira[1[0009-0006-3563-0493]], Thales de Freitas Ferraz[1[0009-0007-4946-3865]], Elpidio Oscar Benitez Nara[1[0000-0002-4947-953X]] Izamara Cristina Palheta Dias[1[0000-0001-5413-0423]]

[1] Pontifical Catholic University of Paraná, Curitiba PR 80215-901, Brazil
Carneiro.marcelo@pucpr.br, p.alexandre@pucpr.edu.br, thales.ferraz@pucpr.edu.br, elpidio.nara@pucpr.br, izamara.dias@pucpr.edu.br

Abstract. The article aims to apply a comparative analysis of machine learning techniques to predict project sales prices of a consulting company located in Curitiba/PR. The company has projects in the most diverse fields, such as in the strategic, productive, quality and innovation areas. Due to this diversity, company managers find it difficult to calculate the sale value of new projects, since they deal with different types of predictor variables such as: type of consultant, type of project and number of hours. In this sense, there is a need to use a method that predicts from a multivariate analysis and that results in sales values close to those expected by the company. To this end, a literature review was carried out on the research topics, namely: Production Planning and Control (PPC) and machine learning techniques; then, the company's current sales prospecting process was mapped; in addition, data were collected, analyzed and prepared, and then proceeded to the testing stage and selection of the best model; and finally, the improvement proposal was discussed with the organization. As a result, it was obtained that the application of the Gradient Boosting Machine (GBM) technique obtained the lowest error result among the Machine Learning techniques tested. The error was approximately 21%, which can be considered acceptable for the analyzed segment. Thus, this work met the expectations of stakeholders by presenting the possibility of pricing projects using computational algorithms for forecasting demand.

Keywords: Machine Learning, Gradient Boosting Machine, Computational Intelligence.

1 Introduction

In the global economic scenario, several companies in the service industry find it difficult to price their business. It is very difficult to determine how much a service "unit" is worth or its costs. The pricing strategy for services then is generally on based on the value observed by the customer, rather than cost, as employed in manufacturing companies. This value is determined by what the customer believes the service is worth, in addition to the prices charged by the competition [1].

The service sector represented 73.4% of GDP in Brazil in 2020, highlighting the importance of generating improvements in the ways in which services are priced and presented to the customer, given its enormous economic contribution to the country [2].

In this context, in addition to service quality to generate value from the customer's point of view, companies need to deal with demands internally, in the proper allocation and organization of their activities and employees. Production Planning and Control (PPC) has the mission of balancing the different interests of demand and supply of the organization. The PPC aligns the activities of the Commercial and Production areas, to do what is best for the company [3].

Forecasting, used in PPC actions, means using a methodological process to determine future data based on statistical, mathematical, or econometric models. This forecast allows the company to better prepare for future demands and to price it in a more assertive and agile way [4].

To make the prediction from these statistical models, machine learning is used. This technique is defined as a set of methods that identify patterns in the data, using them later to predict future data, with the objective of assisting decision making in cases of uncertainty [5].

Machine learning is being increasingly used in many areas and functions. In Brazil, this concept is still little known/used since many companies do not know about its benefits or do not yet have the necessary maturity to implement it. A data control and management system/method is necessary for its use, and these are often not present in the companies. In addition, there are problems when the opposite occurs, that is, the company generates a lot of data simultaneously, making it difficult to choose a model and limiting the hardware processing power of the computers available [6].

Machine learning can be used on small, medium and high complexity problems. In the case of pricing, it is noted that the biggest problem is in relation to the number of responsible predictors that directly influence the elaboration of the project price, which makes it necessary to use techniques of different natures of Machine Learning.

Company managers have some difficulty when calculating the sale value of new projects, given the different variables and possibilities that are present in the whole. In this sense, there is a need to use a method that foresees different scenarios and results in sales values within the expected range. It is expected that it will be possible to obtain project prices in different scenarios, based on predictor variables such as the type of consultant, type of project and number of hours.

Based on the scenario presented, this work seeks to answer the following guiding question: “From demand forecasting methods based on machine learning, is it possible to obtain results capable of assisting in decision making regarding the pricing of consulting projects?”

The present study aims to apply a comparative analysis of machine learning techniques to predict project sales prices, more specifically for a consulting company located in Curitiba. The company operates in several areas, such as in the strategic, productive, quality and innovation areas, having in its portfolio multinationals and medium and small companies throughout Brazil.

2 Methodological Procedures

To carry out the literature review, the databases used were Scopus and Science Direct to search for scientific papers on Production Planning and Control (PPC) and machine learning techniques. To meet the process mapping, the Bizagi Modeler tool was used, which allows the modeling of processes in BPMN (Business Process Model and Notation). For data analysis, python language codes were used, containing the prediction algorithms K-Nearest Neighbor, Random Forest, Support Vector Machines and Gradient Boosting Machine. These codes were pre-modeled, and then modified to meet the needs of the specific problem of this work. The choice of machine learning techniques was because they are the most classic in the literature. Therefore, a literature review was carried out over the last 5 years using the K-Nearest Neighbor, Random Forest, Support Vector Machines and Gradient Boosting Machine methods, in the Scopus database, as descriptors. The search resulted in 48 scientific articles. The most used methods were those presented in the research and as for the area of application, it can be noted that 48% of the sample represents cases of medicine, 8% biotechnology and 8% computer science. As an example, for the medicine sector, [7] were cited, as an example for the case of biotechnology, [8] were cited and for the case of computer science, [9] were cited. It was observed that no article dealt with the case of pricing in the area of project consulting, thus making the research innovative. The sample of 48 articles, considering the applied method, contribution and limitation, is found in Annex A of this research.

Regarding the selection of the best model, some metrics based on error, namely Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). These measures allow an evaluation of each of the models to indicate if they are being correctly applied and if they are generating satisfactory results, close to reality. These metrics were used because they are the most classic ones in the literature to deal with regression cases, according to a systematic review carried out in the sample of 48 articles. As an example of works that applied these metrics, we can mention the 3 articles highlighted in the previous paragraph. Finally, the results obtained were presented to the company's managers to verify whether the values presented were in accordance with those actually practiced for different types of projects and clients.

3 Theorical background

3.1 Production Planning and Control (PPC)

PPC aims to deal with internal demands, allocation, organization of its activities and employees, to meet the customer's wishes. Companies direct their activities in the direction in which they think their business will go. This course is usually based on forecasts, mainly on demand. In this way, the forecasts are used by the PPC at two different times: to plan the production system and the use of this system. In the first case, long-term forecasts are used to create a production strategy, defining which products and services will be offered to the market, among other items. For the use of the production system, detailed medium and short-term forecasts are used for master planning and production scheduling for the most adequate use of available resources [10][11].

3.2 Prediction Methods via Machine Learning

K-Nearest Neightbor (KNN)

The K-Nearest Neighbors (KNN) algorithm is a very simple and effective method for classifying values. It is used in cases where it is necessary to classify an observation z for a population X or Y, using training data. Z is assigned to a population X if at least 1/2 k of a population of k values in the training dataset close to z are from population X [12]. To calculate the distance between observations, the Euclidean method is more frequently used, limited to real values of vectors. Equation 1 below represents the calculation method:

$$d = \sqrt{\sum_{i=1}^{n} (x_n - y_n)^2} \quad (1)$$

In which:
d = distance between observations.
x_n = is the value of an observation "n" relative to a data population X.
y_n = is the value of an observation "n" relative to a data population Y.

Some recent articles in the literature can be highlighted using the KNN technique, for example: in the health care sector for the prevention of diabetes cases [13] and for the prediction of coronary artery diseases [14] and in Product Development Process [15].

Random Forest (RF)

The Random Forest method is a method belonging to the class of methods based on decision trees used in regression and classification, more specifically, applied in this work for machine learning. In most cases, to make a prediction of an observation, the mean or mode of a training observation in the region to which it belongs is used. Because the split rules used to segment the prediction space can be summarized in a tree, these approaches are known as decision tree methods [16].

The problem with the use of decision trees in general is the lack of precision in their results, due to the method learning irregular patterns, in which it adapts too much to the training sets, having low polarization, but high variances. Random Forest comes with the solution to calculate averages of several decision trees, trained in different parts of the same training set, seeking to reduce this variance. This practice increases the performance of the final model, with small losses in contextualization and a slight increase in bias [17]. Some recent articles in the literature can be highlighted using the Random Forest technique, for example: in the marketing sector to identify fake reviews [18] and in the Information Technology sector to detect abnormal traffic [19].

Support Vector Machine (SVM)

Support Vector Machines are a machine learning technique that aims to recognize patterns in data, used to solve classification and regression problems. SVMs are known to have a good generalization ability, in addition to being robust in the application of large databases. As limitations, SVMs have a high sensitivity in choosing parameter values and difficulty in interpreting the generated model [20].

The technique consists of initially separating two classes of data, from the infinite hyperplanes that can be chosen. The objective then is to find the plane that has the largest margin, that is, the maximum distance between the data of the two different classes. Some recent articles in the literature can be highlighted using the SVM technique, for example: in the Civil Construction sector to predict the compressive strength of concrete [21] and heath care to prevent cases of diabetes [13].

Gradient Boosting Machine (GBM)

Most methods that work with regression and machine learning data are based on the use of a single and robust model. The Gradient Boosting Machine is an "ensemble" method, that is, it uses a set of several prediction methods considered weaker, in order to produce better results. Common ensemble methods, such as Random Forest, use an average of the results of the contemplated models. GBM uses the boosting method, in which it is possible to add several new models to the ensemble sequentially. In this sense, at each iteration, a new weaker model is trained to "respect" the error that the set of ensemble models have learned so far, thus increasing the accuracy of the expected prediction [22].

Some recent articles in the literature can be highlighted using the GBM technique, for example: in the natural disaster sector for landslide disaster prevention and mitigation [23], in Civil Construction for mapping the use of landslides [23], in clay [24] and in the electricity sector to prevent energy consumption in residential buildings [25].

3.3 Decision Metric

Mean Absolute Error

Mean absolute error is a way of measuring the accuracy of a data prediction. The error is defined as the current or observed value minus the predicted value, and may have a positive or negative value, based on the average of the sums of the values [26]. The formula for this calculation can be seen in equation 2 below:

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|A_t - F_t| \tag{2}$$

In which:

$\mathbf{n}$ = number of iterations
$\mathbf{A_t}$ = Current value
$\mathbf{F_t}$ = Expected value

Mean Absolute Percentage Error

Mean absolute percentage error represents the average of the total percentage error in a forecast. The error is defined as the current or observed value minus the predicted value. In this way, it is possible to obtain an error value in percentage terms, facilitating the forecast analysis [26]. The formula 3 below corresponding to this concept is similar to the mean absolute error, with the inclusion of the percentage at the end:

$$\mathrm{MAPE} = \frac{1}{\mathrm{n}}\sum_{\mathrm{t}=1}^{\mathrm{n}}\left|\frac{\mathrm{A_t} - \mathrm{F_t}}{\mathrm{A_t}}\right| \times 100\% \tag{3}$$

Mean Squared Error

The mean squared error corresponds to the mean squared error of all errors, being used mainly to calculate the error in numerical prediction models [27]. Formula 3 is similar to the mean absolute percent error:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2 \tag{4}$$

Root Mean Squared Error

The root mean squared error is the root mean squared error of all errors and is mainly used to calculate the error in numerical prediction models. Formula 5 is similar to the mean squared error, represented in formula 4.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2} \tag{5}$$

4 Results

4.1 Process Mapping

The present work was developed based on the data and context of a business consulting company, which is specialized in the areas of strategic management, productivity, quality and innovation. For reasons of preserving the company name, in this paper the organization will be called XYZ company.

The XYZ company has been in the market for 15 years, having provided more than 350 consulting services and 1000 training sessions to more than 150 clients. XYZ's proposal is to have a genuine commitment to positive results, promoting customized and innovative solutions to meet the reality of its different customers. In addition, it uses a systemic view in its services based on methodologies, experience, and adaptability. It also promotes the transfer of know-how, avoiding the generation of dependence on the company and encouraging employees to take a leading role.

In this project, the focus was on the commercial area of the consulting firm, more specifically on prospecting, proposal generation and pricing. In the process mapping section below, you can see the flow of this activity.

The idea for the realization of this project came through Brainstorming, in which the difficulties of the XYZ company were thought, and then several suggestions of activities were generated to be developed in order to meet the identified needs. At the end of this process, the theme of the project was defined. The analyzed process was the prospection of the company XYZ. In the following figure, it is possible to see the mapping of this activity from a BPMN model built in Bizagi Modeler (see Fig. 1).

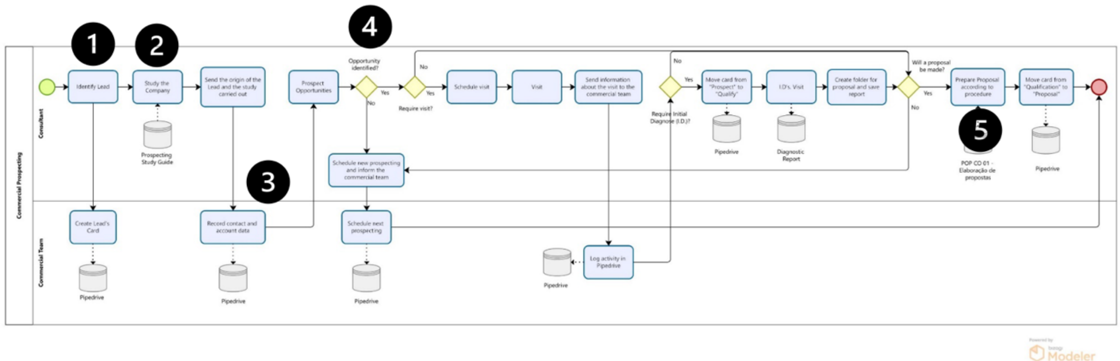

Fig. 1. Company process. Source: Authors, 2022.

In the figure 1, it is possible to observe the different activities that contemplate the prospecting process of the company XYZ. The process begins with the identification of leads (Detail 1 in Fig. 1), study of the company to which the project will be sold (Detail 2 in Fig. 1), and registration of all relevant data in the process (Detail 3 in Fig. 1). When an opportunity is identified (Detail 4 in Fig. 1), the need for a visit is verified or not, and then a proposal is prepared for the client (Detail 5 in Fig. 1), shaped according to the real demand of the client. It is worth mentioning that there are two elements involved in this process, the Commercial team and the Consultant responsible for preparing the proposal.

4.2 Data Collection

The data collected refer to the sales prices of different types of projects, from the year 2012 to 2022. These prices are conceived in the proposal preparation phase, as indicated in figure 1.

In order to preserve customer names, numbers from 1 to 531 were used for the companies contracting the service. For the preservation of project sales price data, a base X % multiplier was used for all values represented in the base.

The following table presents the descriptive statistics of the base:

Table 1. Base descriptive statistics. Source: Authors, 2022.

	Company	Consultant Type A	Consultant Type B	...	Consultant Type F	Value/hour	Class
Count	531	531	531	...	531	531	531
Mean	266	3.05	41.14	...	23.07	199.65	5.81
Standard Deviation	153.43	28.64	63.72	...	100.95	80.62	3.50
Min	1	0	0	...	0	64.84	1
25%	133.5	0	0	...	0	158.12	4
50%	266	0	12	...	0	188.12	10
75%	398.5	0	60	...	0	234.93	10
Max	531	450	402	...	904	1326	11

Note that for the response variable, which is about the price of the project, the total data value represents 531, with an average of R$ 199.65, and deviation of R$ 80.62, and a minimum and maximum value of R$ 64.84 and R$ 1.326,00. The evaluation of the predictive variables was summarized in the table; however, it is possible to perform the same analysis. The data are not normalized in the table to facilitate the understanding of the descriptive statistics of the sample.

Regarding the model variables, only the project price is the response variable, which is classified as continuous values.

As for the project's predictive variables, there were 3 variables. These variables were considered sufficient for the initial proposal of the project: (i) Consultant type: Represented by columns A to F, meaning the quantity of hours by type of consultant involved in the project. These categories were created for this work, in order to anonymize the actual categories of hours used by XYZ; (ii) Value/hour: Represents the amount charged for the sum of the internal hours plus the face-to-face hours of a project. Internal hours include material preparation activities, planning, internal meetings, among others. Face-to-face hours are the hours the client sees, that is, when the training, consultancy, lecture or training is actually applied to the client; (iii) Class: represents the type of project applied. The first letter of the acronym represents the type of service, including “Consulting”, “Training” and “Lecture”. The second letter represents the type of intelligence, that is, it can be “Quality”, “Productivity”, “Innovation” or “Strategy”. For example, the acronym “CS” stands for Strategy Consulting. These different types of project classes were numbered from 1 to 11 to enable the application in the machine learning model.

4.3 Comparative analysis of machine learning techniques

In this step, the forecast methods mentioned above were applied, namely KNN, RF, SVM and GBM. The value of 30% for the test base was used as a parameter for the models. This means that prediction methods use 70% of the data to train the model and 30% to test whether the model is generating solutions as expected. In the systematic review of this work, it is possible to observe the pattern of 70% of the sample dedicated to training and 30% dedicated to testing the model.

The parameters used in each model can be seen below (see Table 2). Table 3 is shown below, with the errors found by the different models.

Table 2. Model parameters. Source: Authors, 2022.

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"*

Table 3. Results of the errors in the prediction models. Source: Authors, 2022.

Metric	KNN	RF	SVM	GBM
MAE	45.98	50.93	50.38	**41.78**
MAPE	0.2425	0.2648	0.2543	**0.2188**
MSE	3813.41	5190.03	4792.38	**3449.49**
RMSE	61.75	72.04	69.23	**58.73**

For the prediction, machine learning techniques KNN, RF, SVM and GBM were used. According to table 01, each technique had as predictor variables (X) the following variables: (i) Consultant type, (ii) Value/hour and (iii) Class and as response variable (Y) the expected price of the project. Considering the total input data, 531, and applying the forecasting techniques, the technique that presented the lowest error, using the MAE, MAPE, MSE and RMSE metrics, was the GBM technique, according to table 03. A cross-validation method was applied to the model in order to avoid overfitting the results. Analyzing the performance of the models using the error metrics, it appears that the GBM method has the lowest error for all metrics. For example, considering MAPE, this method presented approximately 21.88% error when applied to the test set.

Choosing the best predicting technique

For the selection of the best predicting technique, the mean absolute percentage error (MAPE) was used as a reference. In this regard, the GBM model obtained the best result (21.88%), followed by the KNN (24.25%), SVM (25.43%) and RF (26.48%) models. Below is the comparative chart between the results generated by the GBM model and the 30% base test set (159 data out of 531 in total).

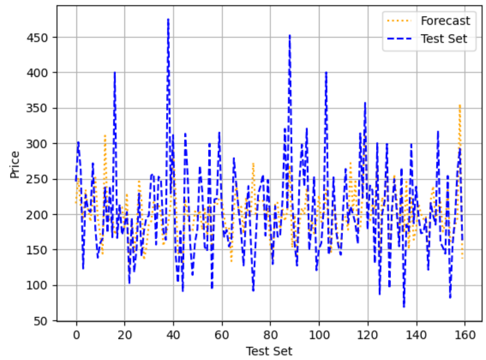

Fig. 2. Comparison between the model prediction and the test set. Source: Authors, 2022.

As shown, the GBM method obtained approximately 21.88% of error, considering the MAPE. The error graph can be seen below:

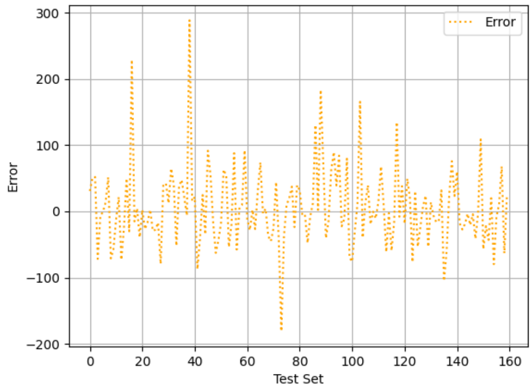

Fig. 3. Errors between the prediction model and the test set. Source: Authors, 2022.

It can be seen in the graph above that the vast majority of data remains with values between 100 and -100 of error, with few exceptions that are outside this set.

5 Final Considerations

As a contribution to the literature, this article presented an innovative dataset, which has not been used recently by other articles, and represents an opportunity to work in the area of price prediction in consulting services based on machine learning methods.

The article aimed to apply a comparative analysis of machine learning techniques to predict the sale prices of projects of a consulting company that operates in the strategic, productive, quality and innovation areas. To reach the objective of the work, the following steps were carried out: a literature review in search of articles and books on PPC topics and machine learning techniques to conceptualize the tools used; then the processes were mapped and data was collected and analyzed; in the sequence, tests and selection of the best model were carried out; finally, the results obtained were presented to the company's managers to verify whether the values presented were in accordance with those actually practiced for different types of projects and clients.

The application of the GBM technique obtained, as previously mentioned, the lowest error result among the Machine Learning techniques. This error, of 21.88%, can be considered acceptable for an initial forecast model in view of the amount of data used as input in the model. Therefore, this work meets the expectations of stakeholders by presenting the possibility of pricing projects using computational algorithms for forecasting demand. In this sense, it is possible to continue this project with the search and implementation of other forecasting techniques available in the literature, in addition to the experimentation and testing of different parameters in the techniques already used. The objective of this action is to increase reliability and decrease forecast error, in order to ensure that company employees use machine learning as a standard for project pricing.

As a limitation of the work, the model has not yet been considered adequate enough to be implemented as a standard for predicting the pricing of company XYZ. In this way, it is indicated, for future work, that the planned action plan be carried out, in order to increase the amount of data and consequently the reliability of the Machine Learning model.

Currently, the company already uses its own demand forecasting method that does not incorporate machine learning characteristics, and the forecasting process needs to go through different sectors of the company in order to obtain results close to reality for pricing. The time spent with this current process of the company was not measured in this work, however, we see that the machine learning method is an agile and intelligent alternative for the organization that has scarce resources and enough data to obtain forecasts with greater accuracy in the future.

References

1. Thomas, D.R.E. "Strategy is different in service business", Harvard Business Review, v. 56, p. 158-65 (1978).
2. Confederação Nacional do Comercio de Bens, serviços e turismo. CNC 75 anos. pp.107 (2020).
3. Buettgen, John Jackson. Administração da produção. UNIASSELVI, pp. 173-175 (2012).
4. Martins, Petrônio Garcia; Laugeni, Fernando P. Administração da Produção. 2nd edn, São Paulo (2005).
5. Murphy, Kevin. P. Machine learning: a probabilistic perspective. MIT press (2012).
6. Lee, In; Shin, Yong Jae. Machine learning for enterprises: Applications, algorithm selection, and challenges. Business Horizons, v. 63, p. 157-170 (2020).

7. Xu, X., Fairley, C.K., Chow, E.P.F., (...), 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. Scientific Reports, 12(1),8757 (2022)
8. 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. Life, 12(9),1443 (2022)
9. 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. Cluster Computing, In press (2022)
10. Tubino, Dalvio Ferrari. Planejamento e Controle da Produção: Teoria e Prática. São Paulo (2007).
11. Chinosi, Michele; Trombetta, Alberto. BPMN: An introduction to the standard. Computer Standards & Interfaces, v. 34, pp. 122-134 (2012).
12. Hall, Peter; Park, Byeong U; Samworth, Richard J. Choice of Neighbor Order in Nearest-Neighbor Classification. Annals of Statistics, p. 2135-2152 (2008).
13. Kaur, H; Kumari, V. Predictive modelling and analytics for diabetes using a machine learning approach. Applied Computing and Informatics, v. 18, p. 90-100 (2022).
14. Ayon, S.I.; Islam, M.M.; Hossain, M.R. Coronary artery heart disease prediction: A comparative study of computational intelligence techniques. IETE Journal of Research, v. 68, n. 4 (2022).
15. Qiu, et al. Sensor combination selection strategy for kayak cycle phase segmentation based on body sensor networks. IEEE Internet of Things Journal, v. 9, n. 6 (2022).
16. James, Gareth et al. An Introduction to Statistical Learning: with Applications in R. New York (2013).
17. Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2 ed. Springer (2016).
18. Alsubari, et al. Data analytics for the identification of fake reviews using supervised learning. Computers, Materials and Continua, v. 70, n. 2, (2022).
19. Salman, et al. A machine learning based framework for IoT device identification and abnormal traffic detection. Transactions on Emerging Telecommunications Technologies, v. 33, n. 3 (2022).
20. Lorena, Ana Carolina; Carvalho, André C.P.L.F de. Uma Introdução às Support Vector Machines. Revista de Informática Teórica e Aplicada, v. 14, p. 43-67 (2007).
21. Shariati, et al. A novel hybrid extreme learning machine–grey wolf optimizer (ELM-GWO) model to predict compressive strength of concrete with partial replacements for cement. Engineering with Computers, v. 38, n. 1, p. 757-779 (2022).
22. Natekin, Alexey. Knoll, Alois. Gradient Boosting Machines, A Tutorial. Frontiers in neurorobotics (2013).
23. Zhang, et al. Slope stability prediction using ensemble learning techniques: A case study in yunyang county, chongqing, china. Journal of Rock Mechanics and Geotechnical Engineering, v. 14, n. 4, p. 1089-1099 (2022).
24. Zhao, et al. Clay content mapping and uncertainty estimation using weighted model averaging. Catena, v. 209 (2022).
25. Olu-ajayi, et al. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. Journal of Building Engineering, v. 45 (2022).
26. Encyclopedia of Production and Manufacturing Management. Mean Absolute Percentage Error (MAPE). Boston (2000).

27. Christie, David; Neill, Simon P. Measuring and Observing the Ocean Renewable Energy Resource. In: Comprehensive Renewable Energy. 2 ed. Elsevier, pp. 149-175. cap. 8 (2022).