PRODUTO & PRODUÇÃO, vol. 24, n.2, p. 1-12. 2023*.

Application of machine learning methods in forecasting sales prices in a project consultancy

Alexandre dos Santos Pereira
Pontifícia Universidade Católica do Paraná (PUCPR)
e-mail: ale952@gmail.com

Marcelo Carneiro Gonçalves
Pontifícia Universidade Católica do Paraná (PUCPR)
e-mail: carneiro.marcelo@pucpr.br

Elpidio Oscar Benitez Nara
Pontifícia Universidade Católica do Paraná (PUCPR)
e-mail: elpidio.nara@pucpr.edu.br

Thales de Freitas Ferraz
Pontifícia Universidade Católica do Paraná (PUCPR)
e-mail: thalesdefreitasferraz@gmail.com

* RECEBIDO em 26/09/2022. ACEITO em 09/06/2023.

Abstract

The objective of this article is to apply a comparative analysis of machine learning techniques to predict project sales prices for a consulting company in the South of Brazil. The company is involved in various fields such as strategy, production, quality, and innovation. Due to this diverse range of projects, the company managers face challenges in accurately determining the sales value of new projects, as they deal with different types of predictor variables such as consultant type, project type, and number of hours. Hence, there is a need to utilize a method that can predict sales values through multivariate analysis and yield results close to the company's expectations. To achieve this goal, the article conducted a literature review on two research topics: Production Planning and Control (PPC) and machine learning techniques. Subsequently, the current sales prospecting process of the company was mapped out. Data were collected, analyzed, and prepared, followed by testing and selection of the best model. Finally, the proposed improvement was discussed with the organization. The results revealed that the application of Gradient Boosting Machine (GBM) technique achieved the lowest error rate among the tested machine learning techniques. The error rate was approximately 22%, which is deemed acceptable within the analyzed segment. Consequently, this study successfully met stakeholders' expectations by demonstrating the potential of utilizing computational algorithms for demand forecasting and project pricing.

Keyword: Machine Learning; Gradient Boosting Machine; Computational Intelligence.

1. Introduction.

In the current global economic landscape, numerous companies in the service industry face challenges in determining the appropriate pricing for their offerings. Pricing services is complex as it involves assessing the value perceived by customers rather than focusing solely on costs, as seen in manufacturing companies. The perceived value is influenced by customer beliefs about the worth of the service and the pricing strategies employed by competitors (THOMAS, 1978).

In Brazil, the service sector played a significant role, contributing 73.4% to the country's GDP in 2020. This highlights the importance of enhancing the pricing and presentation of services to customers, considering their substantial economic impact (CNC, 2020).

In addition to delivering high-quality services to generate customer value, companies must also address internal demands by efficiently allocating and organizing activities and employees. Production Planning and Control (PPC) plays a crucial role in balancing the interests of demand and supply within an organization. PPC aligns the activities of commercial and production departments to optimize company performance (BUETTGEN, 2012).

Forecasting, an essential aspect of PPC, involves utilizing statistical, mathematical, or econometric models to project future data. Accurate forecasting enables companies to better prepare for future demand and facilitate more precise and agile pricing strategies (MARTINS *et al*., 2005).

Machine learning techniques are employed to make predictions based on these statistical models. Machine learning refers to a collection of methods that identify patterns in data and use them to predict future outcomes, assisting decision-making in uncertain scenarios (MURPHY, 2012).

Although machine learning is increasingly utilized across various domains and functions, its adoption in Brazil is still limited due to a lack of awareness among companies about its benefits or a lack of maturity in implementing it. Effective utilization of machine learning requires a robust data control and management system, which is often absent in many companies. Conversely, when companies generate a significant amount of data simultaneously, challenges arise in selecting appropriate models and the computational capacity of available hardware (LEE; SHIN, 2020).

Machine learning can be applied to address problems of varying complexity, including pricing challenges. In the case of pricing, a major hurdle lies in the multitude of predictors that directly impact project pricing, necessitating the use of diverse machine learning techniques (XU, *et al*., 2022).

Managers often encounter difficulties when calculating the sales value of new projects due to the numerous variables and possibilities involved. Consequently, there is a need for a method that can forecast different scenarios and yield sales values within the expected range. By considering predictor variables such as consultant type, project type, and hours required, it becomes possible to determine project prices across various scenarios (CHOI, et al., 2022).

This study aims to address the following guiding question: "Can machine learning-based demand forecasting methods provide valuable insights for decision-making in pricing consulting projects?" It seeks to apply a comparative analysis of machine learning techniques to predict project sales prices, specifically within a consulting company located in Curitiba. The company operates across multiple areas, including strategy, production, quality, and innovation, and serves a portfolio of both multinational corporations and medium and small-sized companies throughout Brazil.

2. Methodology.

For the literature review, the research utilized Scopus and Science Direct databases to search for scientific papers on Production Planning and Control (PPC) and machine learning techniques. Python language codes were employed for data analysis, incorporating prediction algorithms such as K-Nearest Neighbor, Random Forest, Support Vector Machines, Gradient Boosting Machine, and Linear Regression. These codes were initially pre-modeled and subsequently modified to address the specific problem of this study. The selection of machine learning techniques was based on their prominence in the literature, and a literature review was conducted over the past five years, utilizing the descriptors K-Nearest Neighbor, Random Forest, Support Vector Machines, Gradient Boosting Machine, and Linear Regression. This search yielded 48 scientific articles, with the most frequently employed methods being those presented in the research. The application areas varied, with medicine accounting for 49% of the sample, biotechnology for 7%, computer science for 8%, and the remaining 36% representing different fields. Specific articles were referenced as examples for each sector. Notably, no articles were found that addressed pricing in the domain of project consulting, making this research innovative. Annex A of this study contains the sample of 48 articles, considering the applied methods, contributions, and limitations.

To select the best model, various error-based metrics were employed, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics enable the evaluation of each model's performance and the determination of whether they are being appropriately applied, generating satisfactory results that align with reality. These metrics were selected based on their prevalence in the literature for regression cases, as indicated by a systematic review of the 48-article sample. Noteworthy studies applying these metrics were referenced, as mentioned in the preceding paragraph. Finally, the obtained results were presented to the company's managers to assess their alignment with actual pricing practices for different project types and clients.

3. Theoretical foundation.

The theoretical foundation of this scientific paper encompasses three main themes: Production Planning and Control, Prediction Methods via Machine Learning, and Decision Metrics. Firstly, Production Planning and Control will be explored, focusing on its role in balancing the demands and supply within an organization, and its significance in optimizing activities and resource allocation. Secondly, the paper will delve into various prediction methods facilitated by Machine Learning, including Gradient Boosting Machine, Support Vector Machine, Random Forest, K-Nearest Neighbor, and linear regression. These methods are widely used for forecasting purposes and will be discussed in terms of their applicability and effectiveness in predicting project sales prices. Lastly, the paper will present and analyze different decision metrics such as Mean Absolute Error, Mean Absolute Percentage Error, Mean Squared Error, and Root Mean Squared Error. These metrics serve as evaluation criteria for assessing the accuracy and performance of the prediction models. By examining these themes within the theoretical framework, this paper aims to provide a comprehensive understanding of the subject matter and contribute to the advancement of decision-making processes in the pricing of consulting projects.

3.1. Sales and Operations Planning (S&OP) and Production Planning and Control (PPC).

Sales and Operations Planning (S&OP) is a crucial component of the Production Planning and Control (PPC) framework, enabling organizations to align their sales and production activities for optimal results. S&OP integrates sales planning with production planning, bringing together cross-functional teams to collaboratively develop a unified plan that balances demand and supply.

In the S&OP process, organizations review sales forecasts, customer demand, and production capabilities to create a comprehensive plan that meets customer requirements while effectively utilizing available resources. By involving key stakeholders from sales, operations, finance, and other relevant departments, S&OP facilitates communication, collaboration, and decision-making across the organization.

Through S&OP, organizations can effectively manage demand fluctuations, identify potential capacity constraints, and mitigate risks associated with supply chain disruptions. The process involves regular meetings and discussions to assess market trends, customer feedback, and operational constraints, allowing organizations to make informed decisions regarding production volumes, inventory levels, and resource allocation (TUBINO, 2007).

The ultimate goal of S&OP is to achieve a well-balanced plan that optimizes customer service, minimizes costs, and maximizes profitability. By aligning sales and production plans, organizations can improve forecast accuracy, reduce lead times, and enhance customer satisfaction. S&OP serves as a dynamic framework that enables organizations to proactively respond to changes in market demand and business conditions, ensuring long-term success and competitiveness (CHINOSI; TROMBETTA, 2012).

3.2. Machine Learning Techniques.

In this section, we will delve into five prominent machine learning techniques that are widely used in various domains, including sales planning and forecasting. These techniques are Gradient Boosting Machine, Support Vector Machine, Random Forest, K-Nearest Neighbor, and linear regression.

Gradient Boosting Machine (GBM) is an ensemble learning method that combines multiple weak prediction models, typically decision trees, to create a strong predictive model. GBM iteratively builds models by focusing on the errors made by the previous models, effectively improving prediction accuracy (NATEKIN; KNOLL, 2013). 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, in Civil Construction for mapping the use of landslides (Zhang *et al*., 2022), in clay (Zhao *et al*., 2022) and in the electricity sector to prevent energy consumption in residential buildings (Olu-ajayi et al., 2022).

Support Vector Machine (SVM) is a powerful algorithm used for both classification and regression tasks. It works by finding the optimal hyperplane that separates data points into different classes while maximizing the margin between them. SVM is particularly effective in handling high-dimensional data and can also handle non-linear relationships through the use of kernel functions (LORENA et al., 2007). 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 (Shariati, *et al*., 2022) and heath care to prevent cases of diabetes (Kaur; Kumari, 2022).

Random Forest is another ensemble learning technique that constructs a multitude of decision trees and outputs predictions based on the majority vote or average prediction of the individual trees. By combining multiple trees, Random Forest improves prediction accuracy and reduces the risk of overfitting (James *et al*., 2013; Hastie *et al*., 2016). Some recent articles in the literature can be highlighted using the Random Forest technique, for example: in the marketing sector to identify fake reviews (Alsubari et al., 2022) and in the Information Technology sector to detect abnormal traffic (Salman *et al*., 2022).

K-Nearest Neighbor (KNN) is a simple yet effective algorithm for both classification and regression. It works by identifying the K nearest neighbors to a given data point and making predictions based on their majority class or average value. KNN is flexible, as it can handle numerical and categorical data, and is particularly useful when there are clear patterns or clusters in the data (HALL *et al*., 2008). 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 (Kaur; Kumari, 2022) and for the prediction of coronary artery diseases (Ayon *et al*., 2022) and in Product Development Process (Qiu *et al*., 2022).

Linear regression is a widely used statistical technique that models the relationship between a dependent variable and one or more independent variables. It assumes a linear relationship between the variables and estimates the coefficients that best fit the data. Linear regression is interpretable and provides insights into the direction and magnitude of the relationships between variables (NATEKIN; KNOLL, 2013).

By exploring these machine learning techniques in the context of sales planning and forecasting, we aim to understand their strengths, limitations, and suitability for different scenarios. Each technique has its own unique characteristics and choosing the most appropriate one depends on the specific requirements of the sales forecasting problem at hand.

3.3. Evaluation criteria for assessing the accuracy.

The evaluation of predictive models is crucial in sales planning and forecasting to ensure accurate and reliable predictions. To assess the performance and effectiveness of these models, various decision metrics are commonly employed. In this section, we will delve into the discussion of key decision metrics, namely Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). These metrics provide quantitative measures to evaluate the accuracy and precision of predictive models and play a pivotal role in guiding decision-making processes in sales planning and forecasting.

In the realm of sales planning and forecasting, it is essential to evaluate the performance and accuracy of predictive models. To assess the quality of these models, several decision metrics are commonly used, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Mean Absolute Error (MAE) is a straightforward metric that calculates the average absolute difference between predicted and actual values. It provides a measure of the average magnitude of errors, regardless of their direction. MAE is useful for understanding the overall error magnitude but does not consider the relative impact of errors (SWAMIDASS, 2000).

Mean Absolute Percentage Error (MAPE) is a percentage-based metric that calculates the average absolute percentage difference between predicted and actual values. It offers insights into the relative error magnitude and is particularly valuable when dealing with datasets with varying scales. However, MAPE can be sensitive to extreme values and can result in undefined or infinite values when actual values are zero (SWAMIDASS, 2000).

Mean Squared Error (MSE) is a widely used metric that measures the average squared difference between predicted and actual values. It penalizes larger errors more than MAE, making it more sensitive to outliers. MSE is suitable for models that prioritize accuracy and want to account for the variability of erros (CHRISTIE; NEILL, 2022).

Root Mean Squared Error (RMSE) is the square root of MSE, which provides an interpretable measure of the average magnitude of errors. By taking the square root, RMSE brings the metric back to the original unit of measurement, making it easier to understand and compare across different models. RMSE is widely used in sales planning and forecasting, as it combines the benefits of MSE while maintaining interpretability (CHRISTIE; NEILL, 2022).

These decision metrics serve as valuable tools to evaluate and compare the performance of predictive models. The choice of metric depends on the specific objectives and requirements of the sales forecasting task. Whether the focus is on the magnitude of errors, relative error percentages, or a balance between accuracy and interpretability, these metrics provide quantitative measures to assess the effectiveness of the models and guide decision-making processes in sales planning and forecasting.

4. Results.

In this section, we will delve into the crucial aspects of data collection, data analysis, and the application of machine learning techniques. Collecting and analyzing data is fundamental in understanding patterns, trends, and relationships within a given domain. By harnessing the power of machine learning, we can extract meaningful insights and make accurate predictions, thereby enhancing decision-making processes.

4.1. Collecting and analyzing data.

The collected dataset comprises sales prices of various project types spanning from 2012 to 2022. To protect the confidentiality of customer information, a numerical coding system ranging from 1 to 531 was assigned to the contracting companies. Additionally, to ensure the privacy of project sales price data, all values were multiplied by a factor of X% across the dataset.

The descriptive statistics of the dataset are presented in the table below:

Table 1 – Base descriptive statistics.

	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 the response variable, which pertains to the project price, consists of a total of 531 data points. The average price is R$ 199.65, with a standard deviation of R$ 80.62. The minimum and maximum prices are R$ 64.84 and R$ 1,326.00, respectively. The analysis of the predictive variables is summarized in the table, although it can also be further examined. The data in the table are not normalized to facilitate comprehension of the descriptive statistics of the sample.

Regarding the model variables, the response variable is the project price, classified as continuous values. The predictive variables for the project include three main factors. Firstly, the "Consultant type" variable, represented by columns A to F, signifies the number of hours contributed by each consultant type involved in the project. These categories were specifically created for this study to anonymize the actual hour categories utilized by XYZ. Secondly, the "Value/hour" variable represents the charge amount for the combined internal hours and face-to-face hours of a project. The internal hours encompass activities such as material preparation, planning, and internal meetings, while face-to-face hours refer to the client-visible hours when training, consulting, lectures, or other services are directly provided to the client. Lastly, the "Class" variable indicates the applied project type. The acronym's first letter denotes the service type, including "Consulting," "Training," and "Lecture." The second letter denotes the type of intelligence, such as "Quality," "Productivity," "Innovation," or "Strategy." For instance, the acronym "CS" represents Strategy Consulting. To facilitate the application in the machine learning model, these different project classes were assigned numerical values ranging from 1 to 11.

Subsequently, a one-variable analysis was conducted utilizing box plots to identify outliers and gain insights into the distribution of the data. Box plots provide a visual representation of the data's central tendency, dispersion, and skewness. They consist of a box that represents the interquartile range (IQR), with a line indicating the median. The whiskers extend to the minimum and maximum values within a certain range. Outliers, if present, are depicted as individual points outside the whiskers.

This analysis allowed us to assess the presence of any extreme values or data points that deviate significantly from the majority. By examining the box plots, we could identify potential outliers that might have a significant impact on the results and warrant further investigation. Additionally, the box plots provided valuable insights into the overall distribution of the data, such as skewness, symmetry, and the presence of any data clusters or gaps.

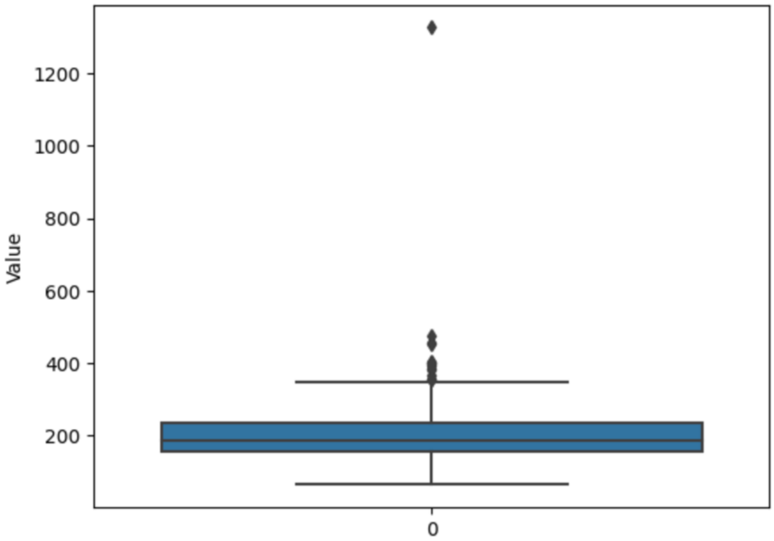

Figure2 – Box-plot analysis.

During the data analysis, we identified an outlier in the dataset, with a project value exceeding R$ 1200. This value stood out significantly when compared to the average project value of R$ 199.65, as shown in Table 1. Given the extreme nature of this outlier, we made the decision to replace it with the average value.

By replacing the outlier with the average, we aimed to mitigate the potential impact of this extreme value on the overall analysis. This approach helps to maintain a more representative and balanced dataset, reducing the potential distortion caused by outliers. This adjustment allows for a more accurate understanding of the central tendency and overall patterns within the data.

After that, we conducted a histogram analysis, which is a valuable tool in data analysis.

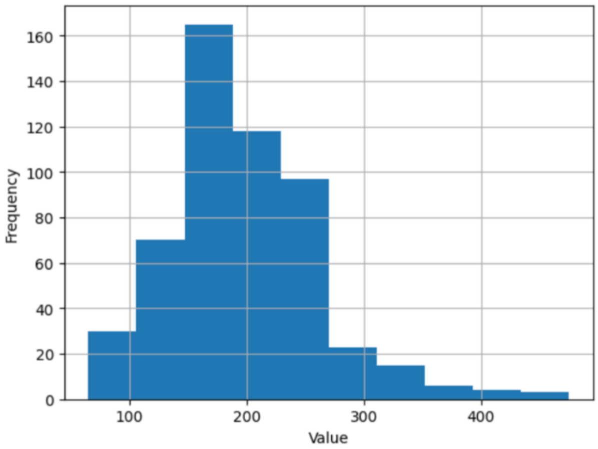

Figure 2 – Histogram analysis.

As we observe, there exists a positive asymmetry, which signifies that the distribution or pattern under examination demonstrates an unequal extent or intensity in its positive aspects compared to its negative aspects. This asymmetry highlights a tendency towards greater positive outcomes or attributes, reflecting a skewed or imbalanced distribution in favor of favorable outcomes.

4.2. Application of Machine Learning techniques.

In this step, the mentioned forecast methods, namely KNN (K-Nearest Neighbors), RF (Random Forest), SVM (Support Vector Machine), GBM (Gradient Boosting Machine) and Linear Regression (LR) were employed. To evaluate the performance of these models, a test base consisting of 30% of the data was utilized as a parameter. This implies that 70% of the data was used for training the models, while the remaining 30% was reserved for testing the models' ability to generate expected solutions. Through a systematic review of this study, the training-testing split of 70% and 30% respectively can be observed as the prevailing pattern.

The specific parameters employed in each model were as follows: For KNN, we used "n_neighbors=20, weights='uniform', algorithm='auto', leaf_size=30, p=2, metric='minkowski', metric_params=None, n_jobs=None". For RF, the parameters included "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 utilized "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", and GBR had "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". Lastly, for LR (Linear Regression), the parameters used were "it_intercept=bool, copy_X=bool, n_jobs=int, bool positivo=none". These parameter configurations were applied to their respective models during the analysis.

For the applied techniques, hyperparameter tuning was employed to select the optimal parameters. Through various methodologies such as grid search, random search, or Bayesian optimization, the parameters of each model were systematically explored to identify the configurations that maximize the model's performance. This iterative process allowed us to fine-tune the hyperparameters and find the most suitable settings for KNN, RF, SVM, GBM, and LR. By leveraging hyperparameter tuning techniques, we aimed to enhance the predictive capabilities and overall effectiveness of these models in generating accurate and reliable results.

Each machine learning technique was applied to the dataset, and the analysis of errors was conducted to evaluate their performance (table 2). The metrics analysis allowed us to assess the accuracy and reliability of the predictions generated by KNN, RF, SVM, GBM, and LR. By quantifying the errors using appropriate evaluation measures such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). We were able to gauge the effectiveness of each technique in capturing and minimizing prediction discrepancies. This approach provided valuable insights into the strengths and weaknesses of the different models, enabling us to make informed decisions and select the most suitable machine learning technique for the given task.

Table 2 – Analysis of errors.

Metric	KNN	RF	SVM	GBM	LR
MAE	44.28	57.02	50.11	**43.11**	48.15
MAPE	0.23	0.29	0.25	**0.22**	0.25
MSE	3737.26	6517.48	4732.19	**3387.86**	4250.52
RMSE	61.13	80.20	68.79	**58.21**	65.20

In order to make predictions, the machine learning techniques employed in this study included KNN, RF, SVM, GBM and LR. Table 01 reveals that each technique utilized certain predictor variables (X), namely Consultant type, Value/hour, and Class, while the response variable (Y) represented the expected price of the project. With a total of 531 input data points, the forecasting techniques were applied, and the resulting errors were evaluated using metrics such as MAE, MAPE, MSE, and RMSE, as shown in table 1. Among the techniques, GBM demonstrated the lowest error across all metrics, indicating its superior performance. To mitigate the risk of overfitting and ensure the generalizability of the results, a cross-validation method was applied to the GBM model. Notably, when assessing the error using MAPE, the GBM technique exhibited an approximate error of 22% when applied to the test set. This demonstrates the significance of cross-validation in assessing and comparing the performance of different models, highlighting GBM as the most accurate and reliable technique in this particular scenario.

The comparative chart provided below depicts the results generated by the GBM model in comparison to the 30% base test set, consisting of 159 data points out of the total 531. It showcases the goodness of fit of the GBM model by visually illustrating the alignment between the predicted values and the actual values observed in the test set.

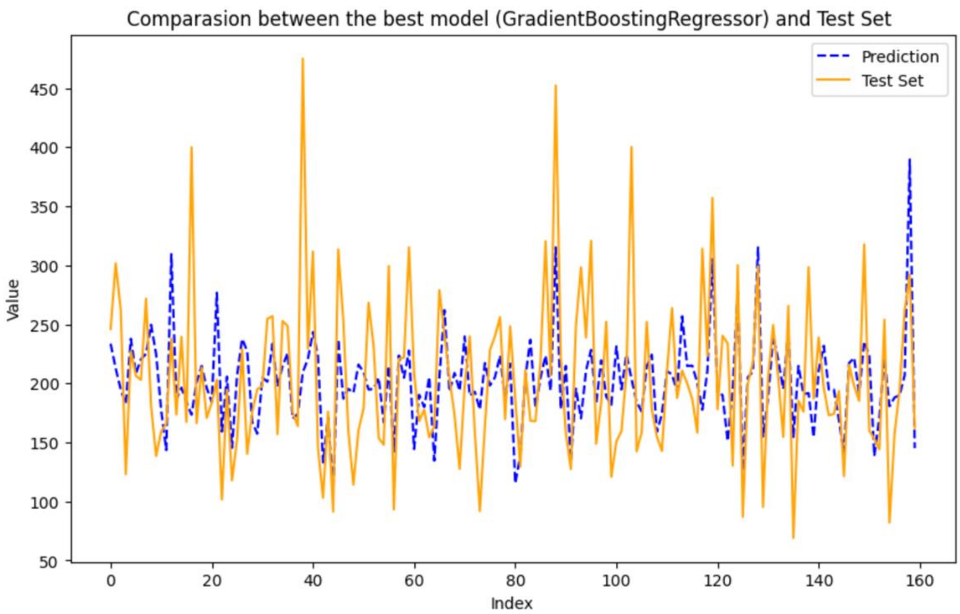

Figure 3 – Results generated by the GBM model.

This chart serves as a valuable tool to evaluate the accuracy and reliability of the GBM model, offering insights into its ability to effectively capture the underlying patterns and trends in the data.

5. Concluding remarks.

This article contributes to the existing literature by presenting an innovative dataset that has not been previously utilized by recent studies. It represents an opportunity to explore price prediction in consulting services using machine learning methods. The objective of this work was to conduct a comparative analysis of machine learning techniques for predicting the sale prices of projects undertaken by a consulting company operating in strategic, productive, quality, and innovation domains.

To achieve this objective, the study followed several steps. A comprehensive literature review was conducted to gather insights from existing articles and books on price prediction and machine learning techniques, establishing the conceptual framework for the employed tools. Subsequently, data collection and analysis were carried out, followed by tests and the selection of the best-performing model. Finally, the obtained results were presented to the company's managers to verify their alignment with the actual prices practiced for different project types and clients.

Among the applied machine learning techniques, the GBM method demonstrated the lowest error, as mentioned earlier. With an error rate of 22%, this outcome can be considered acceptable for an initial forecasting model, given the substantial amount of input data used. Therefore, this work meets the expectations of stakeholders by presenting the potential for pricing projects using computational algorithms for demand forecasting. This opens up possibilities for further exploration of alternative forecasting techniques from the literature, as well as experimentation with different parameters in the already utilized techniques. These efforts aim to increase the reliability and decrease the forecast error, ultimately establishing machine learning as a standard approach for project pricing within the company.

At present, the company relies on its own demand forecasting method, which does not incorporate machine learning capabilities. The current process requires involvement from different sectors within the company to obtain pricing results close to reality. Although the time spent on this existing process was not measured in this study, it is evident that the adoption of machine learning methods offers an agile and intelligent alternative for the organization. By leveraging scarce resources and leveraging available data, the company can achieve more accurate forecasts in a more efficient manner, thus improving pricing outcomes in the future.

References

ALSUBARI, et al. Data analytics for the identification of fake reviews using supervised learning. Computers, **Materials and Continua**, v. 70, n.2, 2022.

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.
Buettgen, John Jackson. Administração da produção. **UNIASSELVI**, p. 173-175 (2012).

CHINOSI, MICHELE; TROMBETTA, ALBERTO. BPMN: An introduction to the standard. **Computer Standards & Interfaces**, v. 34, p. 122-134, 2012.

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**, v. 12, n. 9, p. 1443, 2022.

CHRISTIE, DAVID; NEILL, SIMON P. Measuring and Observing the Ocean Renewable Energy Resource. In: **Comprehensive Renewable Energy. 2 ed. Elsevier**, p. 149-175. cap. 8, 2022.

CONFEDERAÇÃO NACIONAL DO COMERCIO DE BENS, SERVIÇOS E TURISMO. **CNC 75 anos**. p.107, 2020.

HALL, PETER; PARK, BYEONG U; SAMWORTH, RICHARD J. Choice of Neighbor Order in Nearest-Neighbor Classification. **Annals of Statistics**, p. 2135-2152, 2008.

HASTIE, TREVOR; TIBSHIRANI, ROBERT; FRIEDMAN, JEROME. **The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2 ed.** Springer, 2016.

JAMES, GARETH. An Introduction to Statistical Learning: with Applications in R. New York (2013).

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.

LEE, IN; SHIN, YONG JAE. Machine learning for enterprises: Applications, algorithm selection, and challenges. **Business Horizons**, v. 63, p. 157-170, 2020.

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.

MARTINS, PETRÔNIO GARCIA; LAUGENI, FERNANDO P. **Administração da Produção. 2nd** edn, São Paulo, 2005.

MURPHY, KEVIN. P. Machine learning: a probabilistic perspective. **MIT press**, 2012.

NATEKIN, ALEXEY. KNOLL, ALOIS. **Gradient Boosting Machines, A Tutorial. Frontiers in neurorobotics**, 2013.

OLU-AJAYI. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. **Journal of Building Engineering**, v. 45, 2022.

QIU. Sensor combination selection strategy for kayak cycle phase segmentation based on body sensor networks. **IEEE Internet of Things Journal**, v. 9, n. 6, 2022.

SALMAN. A machine learning based framework for IoT device identification and abnormal traffic detection. **Transactions on Emerging Telecommunications Technologies**, v. 33, n. 3, 2022.

SHARIATI. 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.

SWAMIDASS, P. Mean Absolute Percentage Error (MAPE). Encyclopedia of Production and Manufacturing Management. Springer: Boston (2000).

THOMAS, D.R.E. “Strategy is different in service business”, **Harvard Business Review**, v. 56, p. 158-65, 1978.

TUBINO, DALVIO FERRARI. **Planejamento e Controle da Produção: Teoria e Prática**. São Paulo, 2007.

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**, v. 12, n. 1, p. 8757, 2022.

ZHANG. 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.

ZHAO. Clay content mapping and uncertainty estimation using weighted model averaging. **Catena**, v. 209, 2022.