Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Comparative Analysis of Quantitative Predictive Models Utilizing Machine Learning for Business Decision Making

Pedro-Pablo Chambi-Condori[1], Telma-Raquel Saravia-Ticona [2], Miriam Chambi-Vásquez[3], Moisés-Ismael Escobedo-Dueñas[4]

[1]*Universidad Nacional Jorge Basadre Grohmann, Perú*
Orcid: https://orcid.org/0000-0002-8628-6825
Email: pchambic@unjbg.edu.pe
[2]*Universidad Nacional Jorge Basadre Grohmann, Perú*
Orcid: https://orcid.org/0000-0003-3890-6695
Email: tsaraviat@unjbg.edu.pe
[3]*Universidad Nacional Mayor de San Marcos*
Orcid: https://orcid.org/0000-0001-5501-6438
Email: miriam.chambi@unmsm.edu.pe
[4]*Universidad Nacional Jorge Basadre Grohmann*
Email: mescobedod@unjbg.edu.pe,
Orcid: https://orcid.org/000-0002-6799-2185

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

Multivariate analysis is a critical decision-making tool as it allows business managers to understand and predict complex behavior. This is an outline of how various quantitative approaches under the Machine Learning paradigm contribute towards optimizing business decisions. The research sought to evaluate, with measurement of accuracy, the results of the predictive model developed by eight quantitative techniques applied with Machine Learning algorithms executed in Python with the aim of finding the best model to adequately estimate the output variable: price. In data processing, they were loaded, cleaned and then divided into two categories: 70% for training and 30% for testing. The models used in the comparative study were: multiple regression, Ridge regression, Lasso regression, Decision Tree, Gradient Boosting, Random Forest, Support Vector Regression and Neural Networks. Benchmark performance, according to measures of accuracy such as mean square error, mean absolute error, and R-squared, showed that the best fits were for the model developed using Gradient Boosting with an R^2 of 0.8175, followed by the Random Forest model with an R^2 of 0.7889. The two models mentioned above were the most stunning, as they provided the most ideal key indicators for business decision making.

Keywords: Quantitative models, machine learning, decision making, companies.

INTRODUCTION

With a more complex and competitive business world, well-informed, data-driven decisions are key to the success and economic and financial sustainability of businesses. Managerial decision-making backed by credible analysis of information, pattern recognition, and the identification of meaningful trends contributes to the quality of such decisions. In this, Machine Learning has emerged as a valuable Artificial Intelligence (AI) tool for the development of predictive models that can support business managers in making better and well-informed decisions.

1.1. Quantitative Techniques and Machine Learning Literature Review

The study of Kaggwa et al. (2023) recognizes that AI is not only a technology, but a strategic capital significantly re-fashioning decision-making in business. The use of AI in business strategy plays a crucial role in improving business performance and promotes sustainable practices. The study of Ali et al. (2024) shows that AI has become a bulwark of business transformation with unparalleled promise for innovation and efficiency, especially by means of dynamic analysis applications.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

In the meantime, Fatnassi et al. (2025) posit that the application of AI techniques in prediction enhances decision-making and hold the opinion that there is a necessity to strike a balance in its application, highlighting companies' necessity to synchronize AI with their strategic objectives and core values. As AI continues to develop, its impact on business decision-making is anticipated to profoundly shape the corporate environment.

Apart from that, Ghorban et al. (2024) state that, in order to manage their customers efficiently, companies must thoroughly scrutinize the cost and benefit of various spends and other investments and realize the best way of distributing resources to marketing and sales activities in the long run. Choice support models based on AI will benefit decision-makers through predictions of customer portfolio value and linking spending to consumer purchasing behavior.

Mano et al. (2023) claim that machine learning models, given historic data and plenty of input features, can predict the impact of new systems on metrics such as revenue growth, customer behavior, and inventory handling. Firms might, with these models, assess new investment in new equipment and systems with greater precision. Al-Anqoudi et al. (2021) on the other hand address process discovery, process behavior prediction, process improvement, and process optimization. Similarly, Emadi et al. (2024) also propose a simple marketing optimization method on the basis of support vector regression (SVR) in a least squares setting and conducted some numerical experiments to validate the suggested model and procedure.

According to Bagnato (2022), Machine Learning (ML) or "aprendizaje automático" in Spanish refers to one of the Artificial Intelligence areas studying large data processing using statistical models to enable computers to identify patterns from enormous databases and make forecasts, and thereby, it represents an effective means to support business decision-making. The relationship between supervised Machine Learning and Artificial Intelligence is schematically shown in Figure 1.

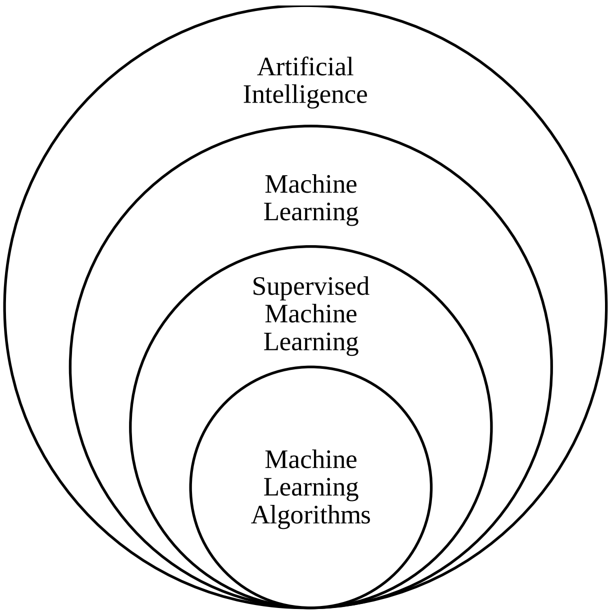

Figure 1. Belonging Relationship between Artificial Intelligence and Machine Learning

Note. Belonging relationship between Artificial Intelligence, Machine Learning, and Machine Learning algorithms.

Yet, Artificial Intelligence (AI), as described by Luger (2009), is the computer science field that aims at automating intelligent behavior. As such, one of the best-known subdomains in AI is Machine Learning (ML), a set of methods and algorithms allowing machines to learn from experience and enhance their performance autonomously. Among the many Machine Learning techniques that exist are multiple regression models, Ridge regression, Lasso regression, decision trees, random forests, support vector machines (SVM), and neural networks, each of which has proven to be very popular in predictive applications. Each of the models possesses strengths and weaknesses that must be carefully weighed based on the nature of the data and the desired objectives. Selecting an optimal model can not only improve

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

predictive effectiveness but can also contribute to the maximization of revenue, cost minimization, and, ultimately, greater general decision effectiveness.

In addition, as stated by Russell (2018) and Huyen (2023), Machine Learning is a broad and ever-evolving discipline. It is divided into four classes: supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. Machine Learning also has a wide range and variety of applications, such as sales and asset price prediction in equity markets, risk and opportunity assessment in investments, voice recognition, online fraud detection, autonomous cars, etc.

Sarkar et al. (2018) argue that the primary challenge that firms have faced in recent decades has been finding ways of utilizing tools to make sense of all of the information that they possess at hand and make practical conclusions that will enable more successful decision-making. From this perspective, impressive development in computer systems and low-cost, gigantic storages are inducing a context for domains such as artificial intelligence, machine learning, and deep learning. In such a situation, data scientists and engineers are using smart systems and models that can predict events and carry out complex analyses.

Supervised regression models are machine learning models employed to forecast continuous values from input data. Supervised regression models are trained using a labeled dataset, where patterns between independent variables and the dependent variable are learned to make accurate predictions on new data. Application of these models in management problems can transform the manner in which companies address challenges such as sales estimation, demand forecasting, risk identification, and resource optimization.

1.2. Quantitative Approaches in Prediction

1.2.1. Linear Regression

Quantitative regression models are analytical tools used for studying and explaining the manner in which one or many explanatory variables have an effect on the dependent variable. Quantitative regression models provide firms with a strong tool for forecasting, analyzing, and measuring important indicators influencing their decision-making process, which helps to make more strategic and informed decisions to ensure maximum efficiency, profitability, and growth. Linear regression is not so complex and can easily be utilized. The model is based on a simple linear equation, and as such, it is easy to understand. Linear regression is computationally inexpensive, and thus best for scenarios where there is vast data and more advanced algorithms may be too costly to train. Linear regression allows quantifying and determining the relationship between independent variables and the dependent variable. Linear regression will be unsuitable if the relationship isn't linear or in case there are complex interactions of variables, where more sophisticated models may be required.

The model specification according to Gujarati and Porter (2010) is as follows:

$Y = f(X_1, X_2, X_3, \dots, X_n)$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \varepsilon$$

Where:

Y: is the dependent variable, impacted by factors that influence its behavior, in this case, the price of a house.
β: are the coefficients that denote the degree of impact of each independent variable on Y, explaining how each factor affects the value of the dependent variable.
i:1, …n: represents the number of factors involved in the model.
ε: is the error or white noise (Gujarati and Porter, p. 188).

1.2.2. Ridge Regression

According to Wu (2020), Ridge regression is a form of regularization, meaning it adds a penalty to the model to avoid overfitting and improve the model's generalization to unseen data. In traditional linear regression, the model is fit to minimize the mean squared error (MSE) between the model's predictions and the actual values. Ridge regression adds a penalty term to the cost function, limiting the size of the coefficients, which helps reduce overfitting.

Minimize $\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2$, where

$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$: is the mean squared error (MSE) between the model's predictions and actual values.
$\sum_{j=1}^{p} \beta_j^2$: is the penalty (also called the regularization term), which is the sum of the squares of the predictor variable coefficients.

λ is the regularization parameter (also known as the Ridge coefficient). This value controls the amount of penalty applied to the coefficients.

1.2.3. Lasso Regression

As per Ramstam and Cook (2018), Lasso Regression (Least Absolute Shrinkage and Selection Operator) is a regularization method that, similar to Ridge regression, is applied to avoid overfitting in linear regression models, particularly when there are numerous predictor variables or multicollinearity. Lasso applies absolute penalization. This modification has significant implications for the way the model treats variables and its capacity to conduct variable selection.

$$RSS_{Lasso} = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 + \lambda \sum_{j=1}^{p}\left|\beta_j\right|$$

$\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$ is the mean squared error (MSE) between the model's predictions and actual values.

$\sum_{j=1}^{p}\left|\beta_j\right|$ is the L1 penalty, which is the sum of the absolute values of the coefficients of the variables.

λ is the regularization parameter, which controls the strength of the penalty. When λ is higher, the penalty is stronger, and the model tends to reduce the coefficients toward zero.

1.2.4. Support Vector Regression

As per Zhang and O'Donnell (2020), Support Vector Regression (SVR) is a statistical approach to examining the interaction between one or more predictor variables and a dependent variable with a real (continuous) value. SVR is a machine learning approach in which a model learns the importance of a variable to explain the interaction between input and output. This approach has the ability to handle nonlinear relationships and provides high accuracy in complex regression problems.

$Min\left(\frac{1}{2}\right)\|w\|^2 + C\sum_{i=1}^{N}(\xi + \xi^*)$, donde

$Min\left(\frac{1}{2}\right)\|w\|^2$ maximizes the margin, and $C\sum_{i=1}^{N}(\xi + \xi^*)$ minimizes the training error.

w is the magnitude of the vector, C is a constant greater than zero and ξ and ξ^* are the variables that control the regression error by the approximation to the bands.

1.2.5. Decision Trees

According to Kotsiantis (2013), the construction of a decision tree starts with all the data at the root node. Then, the algorithm selects the feature that best splits the data into two or more subsets at each step. This is recursively done for each new node created until a stopping criterion is met. Decision trees are used in classification and regression problems.

1.2.6. Random Forest

According to Parmar et al. (2019), the Random Forest algorithm uses a lot of randomly selected trees and averages their prediction. Its high performance and accuracy have been greatly praised. It is an ensemble model, in reference to the combination of decision trees, to perform classification and regression. Gradient Boosting

In accordance with Campillo et al. (2018), the Gradient Boosting is a supervised learning algorithm that chooses successive models to fit the errors of the previous model. It is mainly used for classification and regression and is

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

renowned for providing extremely accurate models. Its basis is on gradient descent, and through this, it can optimize a loss function iteratively by a chain of weak models.

1.2.7. Neural Networks

According to Mishra and Srivastava (2014), neural networks are machine learning algorithms based on the human brain, consisting of artificial neurons stacked on top of one another, which attempt to learn complex patterns in data to perform tasks such as classification, regression, and so much more. They are able and very adaptive, but they require vast amounts of data and computational power. Since they can learn complex representations, they have played a key role in significant advances in areas such as computer vision, natural language processing, and recommendation systems.

1.3. Artificial Intelligence and Machine Learning in Decision-Making

Machine Learning-based research on credit risk supports robust risk analysis to inform prudent credit decisions, note Chang et al. (2024). Predictive accuracy has immensely been enhanced through the use of machine learning and deep learning techniques in this area. In turn, Gadekar and Bhagyashree (2023) note that the integration of machine learning techniques into business processes has been of interest due to its potential to enhance functioning and yield better outcomes. In their experiments with actual company data across various industries, they evaluated the performance of the proposed approach. The results proved the influential role of decision-making through Support Vector Machines (SVM) on organizational performance. Through the application of patterns and insights using SVM, companies can make wise decisions that translate into cost reductions, increased efficiency, and enhanced customer satisfaction.

Muhammad et al. (2025), in their work on customer buying behavior using linear regression machine learning algorithms, utilized the machine learning approach's error rates to assess accuracy, which comprised Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-squared, whose value was 0.3994. On the other hand, Bawa et al. (2025) posit that marketers are utilizing Artificial Intelligence (AI) to track and foretell future consumer purchasing behavior of their target market, and learn more about their experience. Marketers are able to forecast consumer trends and track, evaluate, and analyze consumer spending and behavior. Also, they emphasize that automated systems can recognize concepts and trends in various datasets instantly, comprehend human emotions and behavior, and provide suitable customer support. These systems can predict customer decisions and actions instantly and utilize that data to solve future problems.

Wen et al. (2024) in the sales forecasting, known as sales, revenue forecasting or revenue forecasting described that this role means accurate and timely estimations of future manufacturers', distributors', and retailers' future revenues, and providing it to them in valuable insight. Sales forecasting plays a basic role in various businesses, such as retail businesses, automobile leasing businesses, sale or purchase transactions in real estates, and other conventional businesses. To determine the accuracy of their projections, they got an R^2 value of 0.6181 in forecasting Big Mart sales using the XGBoost regression algorithm. By contrast, Prasad et al. (2024) had provided a comparison of the projection of the U.S. real estate market by regression and ensemble models, investigating the model most appropriate. The aim of this paper was to have accurate forecasts in the real estate industry based on multivariate regression, random forest regressors, decision trees, XGBoosting, and CatBoost. The current research brings useful information when making smart and wise choices within the real estate industry. Decision tree algorithms have also been applied by Mohd. Ariffin et al. (2024) to identify the best model to study topics of internet security.

Rajagopal et al. (2022) in their paper established the difference in policy-making based on human beings and artificial intelligence in view of five considerable contextual factors namely, accuracy within the domain of alternatives search, addition to process innovation and policy output formulation process, number of alternative sets, pace of policymaking, and generalization. They showed that in a world where AI systems exist, customer expectations, industry norms, and collaborative management all complement each other to lead to better strategic business decisions. They argue that technology tools allow entrepreneurs to make better decisions and point out the limitless potential of AI systems. They identify AI's revolutionary role in shaping the future of business culture.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Organizations are decision machines, and organizational results are largely a function of the decisions they take, according to Tabesh (2022). There exists a point of decision before anything in an organization happens. Whether they invest in a new line of products, hire a new worker, enter a new market, respond to a competitor's action, or revive a business model, managers must decide.

Figure 2 illustrates the logical decision-making process and problem-solving in companies under the influence of Artificial Intelligence (AI) and Machine Learning (ML). The issues to be studied and analyzed with the aid of the ML approach are initially determined. Data to be used for the study are subsequently collected from various sources. The modeling process is then carried out using various ML algorithms. The result of the predictions obtained through these techniques is analyzed by human experts, who evaluate the sensibility of the predictions. The predictions are utilized as useful inputs by decision-makers, who use them to find the best solutions based on specified decision criteria.

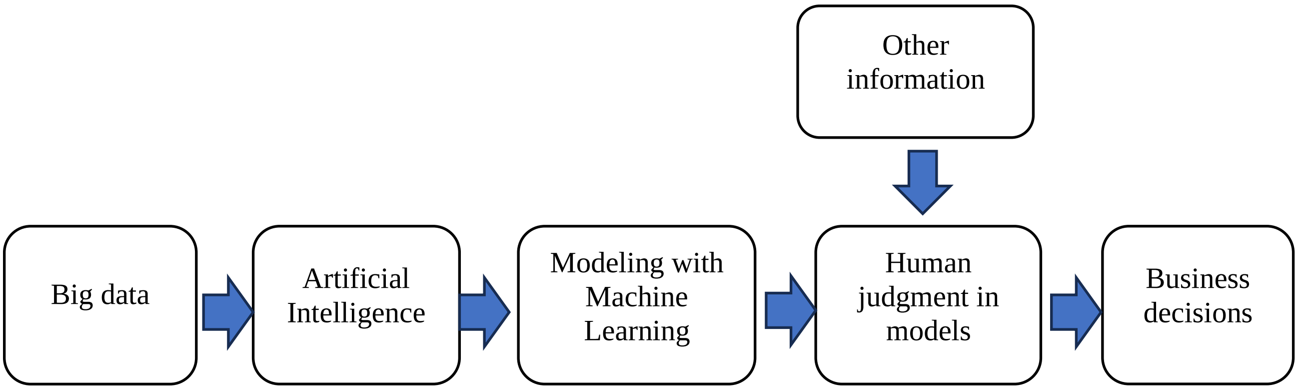

Figure 2. The Confluence Between Data, Machine Learning, and Decision-Making

Note: Self-developed diagram representing the machine learning and decision-making process.

Sarkar et al. (2018) say that the biggest challenge businesses have been grappling with over the last decade has been employing methods to make sense of all the data they hold and derive meaningful information from it to aid in improved decision-making. In such a context, major innovations in computer technology and affordable, immense storage have created an ecosystem for fields such as artificial intelligence, machine learning, and deep learning. Here, data scientists and engineers are using smart models and systems that have the capability to forecast events and perform advanced analysis.

Apart from the context provided in the previous paragraphs, Table 1 gives an overview of the state of the art of multivariate modeling and Machine Learning.

Table 1. *State of the Art in Multivariate Analysis and Machine Learning*

Document Title	Authors	Year	Publication Medium
Econometrics	Gujarati, D. & Porter, C.	2010	Mc Graw Hill
Introduction to Machine Learning with Python	Muller, A. & Guido, S.	2016	O'reilly Inc.
University Dropout: Evaluation of Different Machine Learning Algorithms for Its Prediction	Valero et al.	2021	Revista de Ciencias Sociales (Ve)

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Comparative Assessment of Regression Models Based On Model Evaluation Metrics	Tatachar, A.	2021	International Research Journal of Engineering and Technology
Comparative Analysis of Machine Learning Techniques for Predicting University Dropout Cases	Aco et al.	2023	Revista Ibérica de Sistemas e Tecnologias de Informação
Introduction to the Use of Linear and Nonlinear Regression Analysis in Quantitative Biological Assays	Jarantow et al.	2023	Current Protocols
An Introduction to Statistical Learning wirh Python	James et al.	2023	Springer
A Comparative Analysis of Machine Learning Algorithms for Predictive Analytics in Healthcare	Pal, S.	2024	Heritage Research Journal
Machine learning in business process management: A systematic literature review	Weinzier et al.	2024	Expert Systems With Applications
Comparison of Regression Analysis with Machine Learning Supervised Predictive Model Techniques	Sihombing et al.	2024	Jurnal Ekonomi Dan Statistik Indonesia
Analyzing the Impact of Machine Learning Algorithms on Risk Management and Fraud Detection in Financial Institution	Kumar, D. & Singh, S.	2024	International Journal of Research Publication and Reviews

Note: Prepared by the author based on publication data.

OBJECTIVES

The objective of this article was to perform a comparative evaluation of various Machine Learning models and their performance for outcome prediction, as well as their applicability for business decision-making. By this analysis, it is expected to provide a better understanding of the advantages and disadvantages of every model and make recommendations for its implementation in specific business scenarios. This approach offers a clear description of the context, the appropriateness of data-driven decision-making, and the purpose of the article: comparing and evaluating Machine Learning models in a business environment.

There was a comparative evaluation of various predictive Machine Learning models so as to evaluate their statistical accuracy measures, key to business decision-making. The purpose is to establish the strengths and weaknesses of the models using statistical measures of statistical accuracy with the objective of evaluating the efficiency of various supervised Machine Learning models, which have business decision contexts applications. By a detailed comparison of techniques such as linear regression, Ridge regression, Lasso regression, support vector regression, decision trees, Gradient Boosting, Random Forest, and neural networks, among others, the purpose is to identify which of these models has the best performance in terms of accuracy, stability, and usability in real-world decision-making issues. This comparative analysis attempts to determine the accuracy of each model and how accurate they can be in a bid

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

to optimize business processes, like what influences major decision variables like prices of goods and services offered by companies. It also attempts to provide recommendations on which model is best suited depending on the special character of the data as well as the strategic objectives of the companies.

METHODS

The dataset for comparative analysis was downloaded from scikit-learn.org as 'housing_data_for_regression.csv'. The dataset contains 500 observations and 10 independent variables that are input variables in the models and affect the dependent variable, the output variable: price.

The economic incidents that are modeled using the multiple regression model in Machine Learning may also be handled using other approaches, such as Ridge regression, Lasso regression, Gradient Boosting, decision trees, support vector machines, and neural networks, with the aim of establishing the best-fit model for the event's behavior, as is the case with this study.

The Lasso method, which is defined by Tibshirani (1996), assumes that the coefficients of the linear model are sparse, i.e., predominantly equal to zero. Lasso is a method of minimizing the sum of the squared residuals subject to the constraint that the sum of the coefficients' absolute values is less than some constant.

Conversely, the selected accuracy measures are the Mean Squared Error (MSE), which in Machine Learning calculates the accuracy of predictions that the model produces. A low MSE indicates that the model predicts well because the difference between predictions and actual values is low, whereas a high MSE suggests enormous errors and worse predictions. Mean Absolute Error (MAE) is a statistical measure that calculates the average magnitude of the errors produced by a model. It is the average of the absolute differences between the actual and fitted values. The coefficient of determination (R^2) is another statistical measure that assesses the goodness of fit of the data to the fitted regression line. In multiple regression, it is also called the multiple determination coefficient. This value is between 0 and 1, and the more it approaches 1, the better the model represents the variable. If it is close to 0, then the fit of the model is weaker and, therefore, less reliable. Therefore, if the value of R^2 is high, then the fit is correct and the equation produced does reflect the quantitative relationship between the variables accurately, hence the model can be used to estimate the value of one variable using the others (Almazan et al., 2016, pp.49).

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y}_i)^2}$$

n: total number of data points

MSE : mean squared error

MAE: mean absolute error

Y: actual value

$\hat{Y}$: predicted value

$\bar{Y}$: mean of the actual values

R^2: coefficient of determination

In accordance with instructions provided by Muller and Guido (2016) and James et al. (2023), once in the Python editor, necessary libraries were loaded, and the dataset was imported into the Jupyter Notebook or Google Colab environment. Data preprocessing was then carried out, where the data were cleaned such that all values were numeric and free of blank spaces. The data were normalized afterward. The dataset was then split into two groups: 70% for training and 30% for testing. Python code was executed to generate the evaluation metrics of different models and

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

identify the best-performing indicators. The models utilized were: multiple linear regression, Ridge regression, Lasso regression, decision trees, Gradient Boosting, support vector machines, and neural networks. MSE (mean squared error), MAE (difference between the predicted and actual value at every point), and R^2 (coefficient of determination) were calculated in order to confirm the correctness of the best model for the case under consideration.

RESULTS

The results obtained in the study first present the descriptive features of each of the variables employed in the multivariate model. For this purpose, the statistical distribution of the explanatory variables for housing prices was observed. These variables presented a distribution pattern similar to that of a discrete uniform distribution, between a minimum and a maximum. For example, "housing area" ranges from 50 to 300 square meters and the floors from 1 to 3 with a greater number of houses of two stories. The distribution of the variable price, in its turn, seems normal, that is why it is possible to find no sharp jumps of housing prices.

The mean values of all the variables that were used in the explanatory model of housing prices were also investigated. It was observed that the bedroom number has a mean of 2 to 4. Garage area has a mean of 20 to 40 square meters, and the mean city center distance is 6 to 15 kilometers. Finally, the average value of housing prices ranges from 500 to 650 thousand U.S. dollars, due to the impact of each explanatory variable included in the housing pricing model.

The correlation matrix among the variables of the model was also taken into account. To begin with, a correlation was found between the built-up area and housing price, r = 0.56, and between the number of bedrooms and housing price, r = 0.56. The age of construction was found to be significantly correlated with housing price (r = 0.42), whereas the number of floors was found to have a poor correlation (r = 0.18). The size of the correlation between explanatory variables and the price variable is significant to the analyst as it will indicate which variables are more relevant in explaining housing prices. Secondly, correlation analysis was also conducted to anticipate multicollinearity or autocorrelation issues that may be present in the regression model. No high correlation values were observed among these variables in this case.

In machine learning model construction, the performance of a model needs to be evaluated in order to gauge its effectiveness and ability to generalize. For this, different algorithms were employed under the machine learning approach, and the results are indicated in Table 2 and Figures 3 and 4. From the comparative analysis of three measures — MSE, MAE, and R^2 — it was observed that, in this specific case, the Gradient Boosting model exhibited the highest accuracy values in all three measures with an R^2 of 0.8174. This indicates that 81.74% of the housing price change is explained by the variables employed in modeling, and 18.26% by other variables beyond the model. Another machine learning model that was also very good was Random Forest with an R^2 of 0.7858.

MSE computes the mean of squared differences between predicted and actual values, penalizing large errors more than small ones. This makes it particularly suitable when minimizing large errors is most critical. MAE, however, computes the mean of absolute differences between actual and predicted values. MAE does not penalize large errors as much as MSE, making it more robust when there are outliers in the data.

Whereas, R^2 measures the proportion of the variance of the dependent variable accounted for by the model. Therefore, by juxtaposing MSE, MAE, and R^2, one obtains a complete and more realistic picture of a machine learning model's performance in order to make more informed decisions regarding the quality and utility of the model. Therefore, all three of them have been used in this study.

Table 2. *Accuracy Metrics of the Applied Models*

MODELO	MSE	MAE	R^2
Linear Regression	16.4954	3.0550	0.7433
Ridge Regression	15.8970	3.0537	0.7531
Lasso Regression	23.1552	3.7510	0.6405
Decisión Tree	17.1882	2.9920	0.7331
Random Forest	13.7979	2.4350	0.7858

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Gradient Boosting	11.7578	2.2979	0.8174
Support Vector Regression	48.1511	4.7270	0.2524
Neural Network	31.3921	3.8606	0.5126

Note. Accuracy metrics of machine learning algorithms based on sample data

MSE : Mean Squared Error

MAE: Mean Absolute Error

R^2: Determination Coefficient, or the model's measure of accuracy.

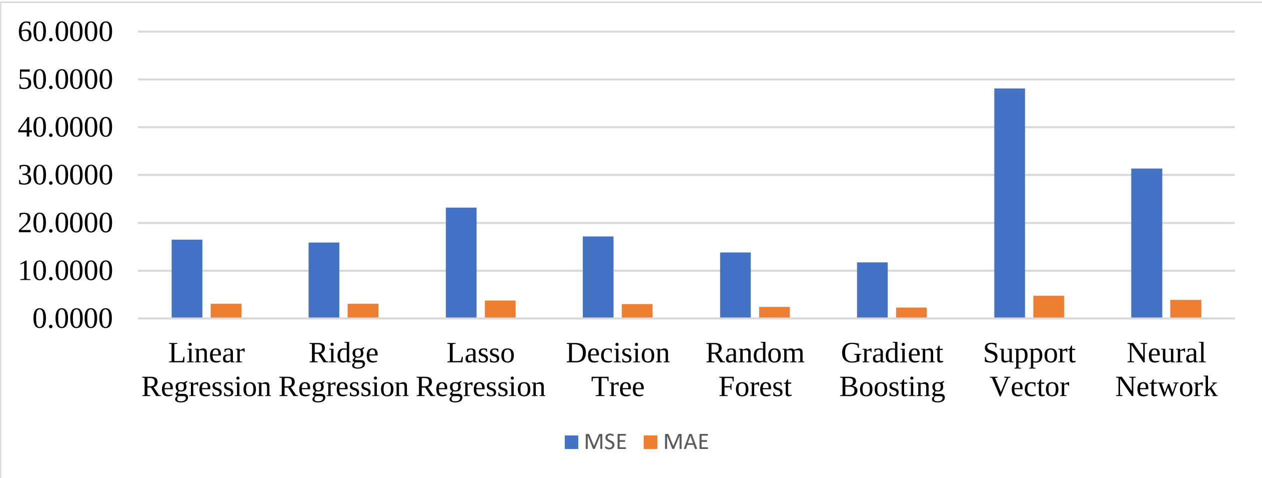

Figure 3. Comparative Evaluation of the Accuracy Metrics MSE and MAE Obtained from Modeling Using the Studied Machine Learning Algorithms

Note. Prepared by the author based on the results presented in Table 2.

The most important figure to utilize while modeling using machine learning algorithms and comparing their performances is R^2, as derived through forecasting based on multiple algorithms, with 70% of historical data being available for training and 30% for testing. The measure of accuracy was derived using the machine learning algorithms researched within this research. R^2 quantifies the predictive power of a machine learning algorithm and provides feedback on the level of fit between predicted values and real values.

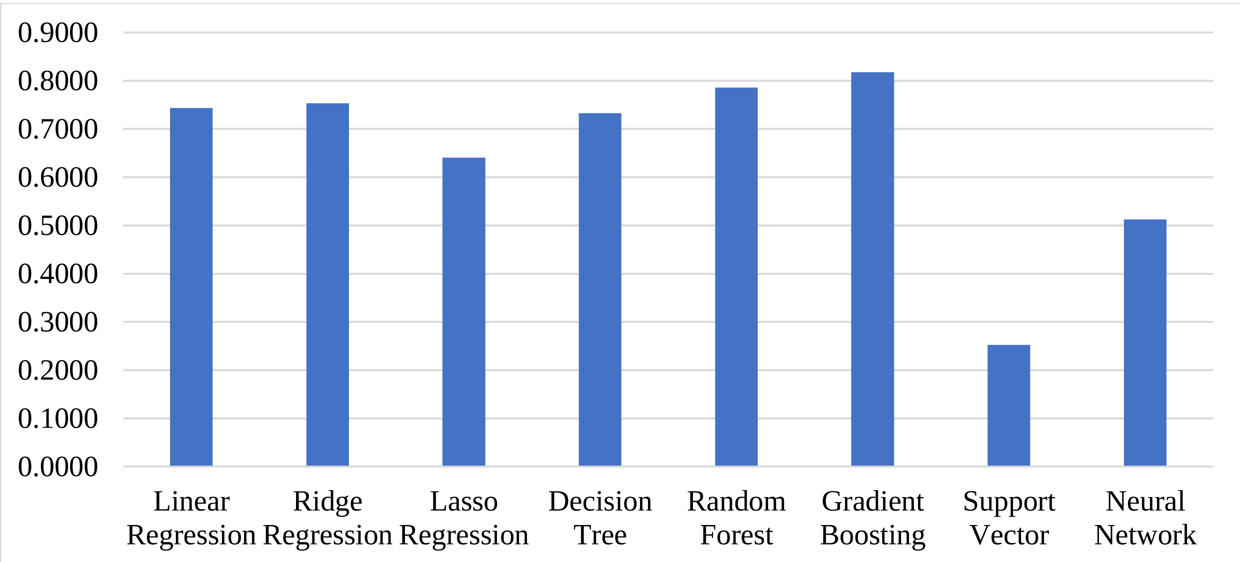

Figure 4. Comparative Evaluation of the Statistical Accuracy Metric R^2

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Note. Comparative evaluation of the R^2 accuracy metric obtained from modeling using the studied machine learning algorithms.

DISCUSSION

In the current study, eight machine learning models were used to provide the accuracy metrics MSE, MAE, and R^2 for predictions in a multiple regression problem. A comparative study based on these three metrics was then conducted. In this current case, the best-performing model was Gradient Boosting with an R^2 of 0.8174, followed by the Random Forest model with an R^2 of 0.7858. These results provide a solid ground for selecting a proper model for the process of real estate pricing. On the statistical measures part, Tatachar (2021) highlighted that regression models are effective machine learning techniques for fitting and predicting relationships among variables on the basis of performance measures MSE, MAE, and R^2.

Similarly, Kumar and Singh (2024) and Sihombing et al. (2024) also evaluate favorably the use of linear regression models, random forests, and support vector regression (SVR) in machine learning modeling. Aco et al. (2023) used MSE, MAE, and R^2 to compare regression models, neural networks, decision trees, and SVR in predicting university dropouts.

Chico et al. (2021) describe that regression analysis is a part of supervised machine learning because it involves the forecasting of a continuous dependent variable from more than one predictor variable. The authors applied three measures of forecast evaluation — MSE, MAE, and R^2 — in their research and obtained an R^2 of 0.756 while modeling using the Random Forest algorithm. Furthermore, Wen et al. (2024) achieved an R^2 value of 0.6181 when they used the XGBoost regression model to fit Big Mart sales. With this experience, they suggest the use of R^2 as a default metric for measuring regression model predictions in any scientific field.

With regard to the values that were achieved in this study, Campillo et al. (2018) opined that Gradient Boosting method, due to the characteristic of correcting prediction mistakes, obtains very accurate outcomes for regression tasks. In the same vein, Parmer et al. (2019) commented that the Random Forest method also has the capability of offering accurate results and is widely used in cases of regression modeling.

The differences between the results of this study and those conducted by Chicco et al. (2021) and Wen et al. (2024) for the same measures and tasks indicate that performances of metrics depend on several factors: the quantity and quality of the data, the data cleaning and normalization process, the selection of an algorithm able to handle nonlinearities, and the selection of accuracy measures used, which are conditioned by the type of regression task being undertaken. Other relevant factors are the available computational resources and the generalizability of the model. All these must be considered when selecting the most appropriate model—one that fits the data well but is also realistic and useful for the specific problem at hand. Briefly, this is why the predictive process is run through different algorithms, as in this specific case, to see which one provides the best precision fit.

CONCLUSIONS

The main objective of this research was to identify the best-performing algorithm for outcome prediction in a regression modeling context and its application in business decision-making. In this context, given the size and quality of the data—and the data cleaning and normalization process—the predictions were done using eight quantitative models based on the machine learning paradigm, for a housing price prediction task within a multiple regression model. It was found that, owing to various factors which influence the quality of prediction, the best precise metrics were gained with the application of the Gradient Boosting algorithm because it allows for corrections as well as creation of highly accurate models, then the Random Forest algorithm.

AI decision-making in a company is priceless, particularly if data processing is entirely conducted by an AI platform. Automatically, without any direct intervention from people, such activity guarantees the data is counted correctly, which makes predictions and decision making easy. AI is capable of performing anomaly detection, data processing, complex analysis, decision making with best results, and trend identification. The last decision can either be automated or done manually.

Journal of Information Systems Engineering and Management
2025, 10(45s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Overall, application of supervised machine learning models to managerial decision-making can confer unparalleled benefits if the right models are selected and deployed. The comparative analysis establishes grounds for the identification of each model's strengths and weaknesses and for making better-informed decisions in a rapidly changing business environment. Besides, based on the data collected in this study and other referenced experiences such as those of Wen et al. (2024), it is clear the prediction processes with the application of various algorithms are necessary because, in this specific case, the best algorithm of maximum accuracy fit is chosen, adding effectively to business decision-making

REFRENCES

[1] Aco, A., Hancco, B. & Pérez, Y. (2023). Análisis comparativo de Técnicas de Machine Learning para la predicción de casos de deserción universitaria. *Risti-Revista Ibérica de Sistemas e Tecnologias de Informacao*. 50(9). https://doi.org/10.17013/risti.51.84-98

[2] Ali, R., Xu, J., Baig, M.H., Rehman, H.S.U., Waqas Aslam, M., Qasim, K.U. (2024). From data to decisions: enhancing financial forecasts with LSTM for AI token prices. *Journal of Economic Studies,* https://doi.org/10.1108/JES-01-2024-0022

[3] Al-Anqoudi, Y.; Al-Hamdani, A.; Al-Badawi, M.& Hedjam, R. (2021). Using Machine Learning in Business Process Re-Engineering. *Big Data and Cognitive Computing*. 5(4),61. https://doi.org/10.3390/bdcc5040061

[4] Almazan, A., Arribas, J., Camarero, L., Mañas, B. & Vallejos, A. (2016). Análisis Estadístico para la Investigación Social. Editorial Garceta – Madrid, España.

[5] Bagnato, J. (2022). Aprende Machine Learning. *Ediciones Leanpub*. http://leanpub.com/aprendem

[6] Bawa, S., Kunal, K., Kaur, K., Sharma, J., Srivastava, V. & Tikku, P. (2025). An Analysis of Artificial Intelligence Implications and its Impact on Marketing. A Systematic Review. *Communications on Applied Nonlinear Analysis*. *32(1s)*. e-ISSN: 1074-133X. https://doi.org/10.52783/cana.v32.2115

[7] Campillo, J. P., Vargas, J. M., & Ibáñez, P. C. (2018). Análisis de la utilidad del algoritmo Gradient Boosting Machine (GBM) en la predicción del fracaso empresarial. *Spanish Journal of Finance and Accounting / Revista Española de Financiación y Contabilidad*, *47*(4), 507–532. https://doi.org/10.1080/02102412.2018.1442039

[8] Chang, V., Sivakulasingam S., Wang, H., Wong, S., Ganatra, M. & Luo, J. (2024). Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers. *Risks*. 12(11). https://doi.org/10.3390/risks12110174

[9] Chicco D, Warrens MJ, Jurman G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Comput Sci*. 2021 Jul 5;7:e623. doi: https://doi.org/10.7717/peerj-cs.623

[10] Emadi, S. H., Sadeghian, A., Rabbani, M., & Dehghan Dehnavi, H. (2024). A machine learning approach to the optimal control of the customer dynamics. Journal of applied research on industrial engineering, 11(2), 212-228. https://doi.org/10.22105/jarie.2024.426657.1578

[11] Fatnassi, K., Zahaf, S. & Gargouri, F. (2025). Artificial intelligence integration for extension of big data for decision-making. *Future Generation Computer Systems*. *166(1)*. https://doi.org/10.1016/j.future.2024.107635

[12] Gadekar , B. ., & Hiwarkar, T. . (2023). A Proposed Business Improvement Model Utilizing Machine Learning: Enhancing Decision-Making and Performance. International Journal of Intelligent Systems and Applications in Engineering, 12(1s),557–568. https://ijisae.org/index.php/IJISAE/article/view/3491

[13] Ghorban, H., Boozary, P., Sheykhan, S., Rabiese, M. & Rahmani, F. (2024). Predictive analytics in customer behavior: Anticipating trends and preferences. *Results in Control and Optimization*. *17(1)*. https://doi.org/10.1016/j.rico.2024.100462

[14] Gujarati, D. & Porter, C. (2010). *Econometría*. Quinta edición. Mc Graw Hill, México DF. https://fvela.wordpress.com/wp-content/uploads/2012/10/econometria-damodar-n-gujarati-5ta-ed.pdf

[15] Huyen, C. (2023). *Diseño de sistemas de Machine Learning: Un proceso iterativo para aplicaciones listas para funcionar*. Editorial Marcombo. Madrid España.

[16] James, G., Witten, D., Hastie, T. & Tibshirani, R. (2023). *An Introduction to Statistical Learning with Python*. Springer. https://hastie.su.domains/ISLP/ISLP_website.pdf.download.html

[17] Jarantow, S. W., Pisors, E. D., & Chiu, M. L. (2023). Introduction to the use of linear and nonlinear regression analysis in quantitative biological assays. *Current Protocols*, 3, e801. https://doi.org/10.1002/cpz1.801
[18] Kaggwa, , E. , Okonkwo, F. , Farayola, O. , Uwaoma, P. & Akinoso, A. (2023). AI in Decision Making: Transforming Business Strategies. *International Journal of Research and Scientific Innovation (IJRSI)*. DOI: https://doi.org/10.51244/IJRSI.2023.1012032
[19] Kotsiantis, S.B. (2013). Decision trees: a recent overview. *Artif Intell Rev* 39, 261–283. https://doi.org/10.1007/s10462-011-9272-4
[20] Kumar, D. & Singh, S. (2024). Analyzing the Impact of Machine Learning Algorithms on Risk Management and Fraud Detection in Financial Institution. *International Journal of Research Publication and Reviews*. http://dx.doi.org/10.55248/gengpi.5.0524.1135
[21] Luger, G. & Stubblefield, W. (2009*). Artificial Intelligence*. Pearson Education Inc. Boston, USA.
[22] Mano A., Kilaru, M., Prasad, K., Ninay, K., Banitamani, M. & Kaylan, Ch. (2023). Machine learning models for evaluating the benefits of business intelligence systems. The Journal of High Technology Management Research. 34(2). https://doi.org/10.1016/j.hitech.2023.100470
[23] Mishra, M. & Srivastava, M. (2014). A view of Artificial Neural Network, *International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)*, Unnao, India, 2014, pp. 1-3, doi: https://doi.org/10.1109/ICAETR.2014.7012785
[24] Mohd.Ariffin, N. Mohamed.Iqbal, M, Yusoff, M. & Mohd, N. (2024). A Study on the Best Classification Method for an Intelligent Phishing Website Detection System. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. 48(2). https://doi.org/10.37934/araset.48.2.197210
[25] Muhammad, N., Amin, M., Adnan, R. & Nunkaw, O, (2025). Predictive Modelling on Competitor Analysis Performance by using Generalised Linear Models and Machine Learning Approach. *Journal of Advanced Research in Applied Sciences and Engineering Technology*. *45(1)*. https://doi.org/10.37934/araset.45.1.5159
[26] Muller, A. & Guido, S. (2016). *Introduction to Machine Learning with Python*. O´Reilly Books. https://www.nrigroupindia.com/e-book/Introduction%20to%20Machine%20Learning%20with%20Python%20(%20PDFDrive.com%20)-min.pdf
[27] Pal, S. (2024). A Comparative Analysis of Machine Learning Algorithms for Predictive Analytics in Healthcare. *Heritage Research Journal* ISSN No: 0474-9030. https://heritageresearchjournal.com/
[28] Parmar, A., Katariya, R., Patel, V. (2019). A Review on Random Forest: An Ensemble Classifier. In: Hemanth, J., Fernando, X., Lafata, P., Baig, Z. (eds) International Conference on Intelligent Data Communication Technologies and Internet of Things (ICICI) 2018. ICICI 2018. Lecture Notes on Data Engineering and Communications Technologies, vol 26. Springer, Cham. https://doi.org/10.1007/978-3-030-03146-6_86
[29] Prasad, R.D., Varma, V.J., Jyothi, U.P., Shankar, S.S., Deenakonda, M., Narasimharao, K. (2024). Optimizing Real Estate Prediction - A Comparative Analysis of Ensemble and Regression Models. In: Pareek, P., Gupta, N., Reis, M.J.C.S. (eds) Cognitive Computing and Cyber Physical Systems. IC4S 2023. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol 536. Springer, Cham. https://doi.org/10.1007/978-3-031-48888-7_12
[30] Rajagopal, N., Qureshi, N., Durga,S., Ramírez, E., Huerta, R. Gupta, S. & Deepak, S. (2022). *Future of Business Cultures: An Artificial Intelligence-Driven Digital framework for Organization Decision-Making Process*. Wiley. USA. https://doi.org/10.1155/2022/7796507
[31] Ranstam, J. & Cook, J. (2018). LASSO regression, *British Journal of Surgery*, Volume 105, Issue 10, September 2018, Page 1348, https://doi.org/10.1002/bjs.10895
[32] Russell, R. (2018*). Machine Learning: Guía detallada para implementar algoritmos de Machine Learning con Python*. CreateSpace Independent Publishing Platform, USA.
[33] Sarkar, D., Bali, R. & Sharma, T. (2018). *Practical Machine Learning with Python*. Springer. New York. https://doi.org/10.1007/978-1-4842-3207-1
[34] Sihombing, P. R., Budiantono, S., Arsani, A. M., Aritonang, T. M., & Kurniawan, M. A. (2023). Comparison of Regression Analysis with Machine Learning Supervised Predictive Model Techniques. *Jurnal Ekonomi Dan Statistik Indonesia*, *3*(2), 113-118. https://doi.org/10.1594/jesi.03.02.03

[35] Tabesh, P. (2022). Who's making the decisions? How managers can harness artificial intelligence and remain in charge. *Journal of Business Strategy*. Vol. 43 NO. 6 2022, pp. 373-380. DOI: https://doi.org/10.1108/JBS-05-2021-0090
[36] Tatachar, A. (2021). Comparative Assessment of Regression Models Based On Model Evaluation Metrics. *International Research Journal of Engineering and Technology (IRJET).* https://www.irjet.net/archives/V8/i9/IRJET-V8I9127.pdf
[37] Tibshirani, R. (1996). Regresión, contracción y selección mediante el lazo, *Journal of the Royal Statistical Society: Serie B (Metodológica)*, Volumen 58, Número 1, enero de 1996, páginas 267-288, https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
[38] Valero, J., Navarro, A., Larios, A. & Julca, J. (2021). Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción. *Revista de Ciencias Sociales (Ve).* https://www.redalyc.org/articulo.oa?id=28071865024
[39] Weinzierl, S., Zilker, S., Dunzer, S. & Matzner, M, (2024). Machine learning in business process management: A systematic literature review. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2024.124181
[40] Wen, K. Y., Joseph, M. H., & Sivakumar, V. (2024). Big Mart Sales Prediction using Machine Learning. *EAI Endorsed Transactions on Internet of Things*, *10*. https://doi.org/10.4108/eetiot.6453
[41] Wu, Y. (2020). Can't Ridge Regression Perform Variable Selection? *Technometrics*, *63*(2), 263–271. https://doi.org/10.1080/00401706.2020.1791254
[42] Zhang, F. & O'Donnell, L. (2020). Suport Vector Regression. Chapter 7. Editor(s): Andrea Mechelli, Sandra Vieira, Machine Learning, Academic Press, Pages 123-140, ISBN: 9780128157398, https://doi.org/10.1016/B978-0-12-815739-8.00007-9