Associate Professor Adrian Nicolae COZGAREA, PhD
E-mail: adrian.cozgarea@cig.ase.ro
Department of Management Information Systems
Bucharest University of Economic Studies, Romania

Associate Professor Gabriel COZGAREA, PhD
E-mail: gabriel.cozgarea@cig.ase.ro
Department of Management Information Systems
Bucharest University of Economic Studies, Romania

Associate Professor Dana Maria BOLDEANU, PhD (corresponding author)
E-mail: boldeanud@cig.ase.ro
Department of Management Information Systems
Bucharest University of Economic Studies, Romania

Professor Irina PUGNA, PhD
E-mail: irina.pugna@cig.ase.ro
Department of Management Information Systems
Bucharest University of Economic Studies, Romania

Professor Mirela GHEORGHE, PhD
E-mail: mirela.gheorghe@cig.ase.ro
Department of Management Information Systems
Bucharest University of Economic Studies, Romania

PREDICTING ECONOMIC AND FINANCIAL PERFORMANCE THROUGH MACHINE LEARNING

Abstract. *The aim of this paper is to demonstrate the usefulness of supervised machine learning algorithms in predicting the profitability of Romanian companies applying International Financial Reporting Standards (IFRS), both by regression and classification methods. The algorithms used in this research are linear regression (LinR), logistic regression (LogR), decision tree (DT), random forest (RF), K-nearest neighbor (KNN), and multi-layer perceptron (MLP). The results showed that both methods can produce models with high accuracy in profitability prediction. Thus, for regression, the best estimates were generated by the MLP model, and for classification, by the RF model. These results can be used to obtain sustainable models for predicting economic and financial performance, with a major impact on the management decisions of companies.*

Keywords*: Profitability prediction, Machine learning, Regression, Classification*

JEL Classification: C53, C45, C38, M15

DOI: 10.24818/18423264/57.2.23.06

1. Introduction

Predicting financial performance is a particularly important issue for any company. Accurately evaluating and predicting financial performance, using various financial indicators, could help both analysts and business owners, as well as managers, in taking appropriate measures for profit (Ecer, 2013a).

Over the years, several studies have been carried out related to the prediction of economic performance of organisations using different machine learning (ML) algorithms (Budak & Sarvari, 2021; Ecer, 2013a; Özlem & Tan, 2022). Following our analysis, we observed that many researchers have used in their studies either classification methods (Ecer, 2013a; Gregova et al., 2020) or regression methods (Budak & Sarvari, 2021; Özlem & Tan, 2022), and less frequently both methods.

The aim of this study is to demonstrate the usefulness of both regression and classification methods as supervised ML methods in predicting the profitability of companies. The ML algorithms used in our study are linear regression (LinR), logistic regression (LogR), decision tree (DT), random forest (RF), K-nearest neighbor (KNN) and multi-layer perceptron (MLP). More specifically, we will try to prove the viability of these algorithms in predicting a company's financial performance by comparing continuous values obtained by regressions and discontinuous values obtained by classification methods.

This study is based on data provided by companies in Romania that apply the International Financial Reporting Standards (IFRS). The paper is organised as follows: Section 2 introduces an extensive literature review on the use of ML algorithms to predict the economic and financial performance of organisations. Section 3 presents the proposed methodology with theoretical descriptions of the ML tools and algorithms used in the study. Section 4 provides the empirical results obtained by applying ML regression and classification methods, and Section 5 contains the authors' discussions and conclusions on the study, as well as comments on the limits of our research and some future research directions.

2. Literature review

In recent years, considering the constant growth of ML tools to 'unlock new value or boost efficiency' (Brown, 2021) in different industries, such as business and finance (Bussmann et al., 2021; Ecer, 2013a), health care (Young & Steele, 2022), transportation and utilities (Dia et al., 2022) and the fact that according to a Deloitte survey 67% of companies are using ML, and 97% are using or planning to use it in the next year (Brown, 2021), several articles have been conducted in the literature to identify solutions to predict the firm's performance, as a topic of great interest for decision makers (Delen et al., 2013). According to Husmann et al., the application of ML in the finance field enables researchers and practitioners to gain new insights into financial data that is useful for optimal decision making (Husmann et al., 2022). There are different studies performed to accurately predict financial performance, mentioning the usage of both traditional statistical method LinR (Qi & Deng, 2019), and ML methods, among which we underline LogR (Gregova et al., 2020), Artificial

Neural Network (ANN) (Lam, 2004), DT (Hoang & Wiegratz, 2022), RF (Gregova et al., 2020) and KNN (Li & Sun, 2009).

There are mixed research results in predicting the firms' financial health, on the one hand, approaching financial performance (Delen et al., 2013; Lam, 2004), on the other hand, financial distress or bankruptcy (Clement, 2020; Krusinskas et al., 2022). Based on Manogna and Mishra's work, the variables used in the prediction models are financial ratios considered traditional tools to have a better understanding of the financial health of a company, rather than the absolute values which are identified in the financial statements (Manogna & Mishra, 2021). (Delen et al., 2013; Manogna & Mishra, 2021) showed that the identification of the best combination of financial measures or ratios that can accurately predict the firms' performance between industries, within the groups, and across the departments in the company itself or to considering the size of the firm, is of great interest to any decision maker.

Several authors (Alaka et al., 2018; Clement, 2020) have analysed in the last years the prediction of the business performance models based on parametric (LogR) and non-parametric (ANN, SVM, DT and KNN) algorithms applied on different industries, origin country, analysed sources of data, sample size and timelines to emphasise, according to Clement, the 'comparative performance of the techniques by presenting the accuracy' (Clement, 2020). Most studies focus on their proposed models on inputs or predictors defined based on financial data extracted from the financial statements of the analysed companies in the form of financial ratios (Ecer, 2013b; Gregova et al., 2020; Lam, 2004). There were identified in the models used different financial ratios based on their frequency of occurrence the models having various number and type of variables or predictors, i.e., profitability ratios (ROA – return on assets, ROE – return on equity, Gross/Net profit margin), liquidity ratios to show debt paying abilities (Current ratio, Quick ratio, Solvency ratio) etc. (Gregova et al., 2020).

However, in the literature review, there is a lack of relevant analysis to detect financial performance assessment based on simple variables retrieved from either the balance sheet or the income statement to be included as inputs in the proposed models or applied in different available algorithms (Ecer, 2013a; Wei et al., 2021). Ecer (Ecer, 2013a) used DT and MLP as intelligent techniques to predict the financial performance of 500 Turkish companies having in its model predictors such as sales, equity, assets, export, and number of employees and profit before tax as output and showed that MLP model outperformed DT in classifying the companies in terms of good or poor performance 'with an accuracy of more than 86%'(Ecer, 2013a). The prediction of business performance can also be seen in grouping the companies into two main financial categories: healthy or unhealthy considering, at the same time, only the financial ratios' implications (Yeh et al., 2010).

Based on Athey and Imbens research, DT and, their extension, RF 'have become very popular and effective methods for flexibly estimating regression functions in settings where out-of-sample predictive power is important' (Athey & Imbens, 2019). Delen et al. analysed the dimensions and impact of financial ratios, mainly earnings before tax to equity, on the firm performance using four DT algorithms (CHAID, C5.0, QUEST, and C&RT) on all the Turkish listed companies and proved as a final result that only CHAID – Chi-squared automatic interaction detector and C5.0 algorithms produced the best prediction accuracy (Delen et al., 2013).

Gregova et al. used ML LogR, RF, and ANN algorithms to identify the model with the highest predictive accuracy of financial difficulties for Slovak Enterprises (Gregova et al., 2020). Özlem and Tan have used Multiple linear regression (MLR), KNN, support vector regression (SVR), DT, extreme gradient boosting algorithm (XGBoost) and multi-layer neural networks (MLNN) algorithms for the prediction of cash holdings for 211 companies listed on Borsa Istanbul, analysed made between 2006-2019 (Özlem & Tan, 2022).

Having a superlative learning ability, ANNs are used in financial matters, including the prediction of the stock market, bankruptcy prediction, and corporate bond rating, thus becoming 'a popular tool for financial decision-making' (Lam, 2004). Given the power of ANN, extensive research and empirical studies are carried out using regression and classification methods with applicability in the economic-financial field (Ecer, 2013a; Gregova et al., 2020; Lam, 2004).

3. Methodology

In this research, we aim to demonstrate the extent to which ML algorithms can predict the economic and financial performance of companies, using both regression and classification methods. Profitability of a company can be expressed as an absolute value (profit/loss) or as a relative value, through rates of return (Return on Assets-ROA, Return on Equity-ROE, Return on Capital Employed-ROCE etc.).

In our study, we focused on predicting enterprise performance using two supervised machine learning methods: regression, as a method of predicting continuous values, and classification, as a method used in predictions of discrete values. Regression was used to forecast the financial result in absolute terms and the classification was applied to determine the category into which a company can be classified according to its return on assets (ROA).

3.1. Dataset

This study is based on public data provided by the Romanian Ministry of Finance *(https://data.gov.ro).* The dataset includes information taken from the IFRS financial statements of Romanian companies in the period 2013-2021 (about 870 records). Data preparation covered:

- Eliminate companies with zero values in all columns (two records).
- Replacing missing values (blanks). After a rigorous analysis, each missing value in the dataset was replaced by zero, considering that the data were taken from financial statements (e.g., balance sheet, profit and loss account) certified by state institutions, where any missing value can be assimilated to zero.

- Conversion of all financial data into euro to ensure a uniform valuation.
- Data shuffling, the operation being required by the initial retrieval of data grouped over the 9 years of reference.
- Partitioning the dataset into two subsets necessary for training and validating the supervised regression and classification algorithms: 80% (for training) and 20% (for testing).

3.2 Variables

The data collection published by the Ministry of Finance contains a number of indicators taken largely from the financial statements (balance sheet, profit and loss account). In order to use only data with a high degree of independence, among the financial indicators we have chosen those that are not directly reflected in the result of a company and omitted/ignored indicators that directly influence this result (e.g., revenue, costs). In addition, since by its nature the balance sheet contains patrimonial elements represented in two ways (assets and sources of financing of assets), we avoided using indicators from both categories, thus trying to remove redundant information that could have distorted the results of the study. Therefore, we considered as independent (input) variables the following indicators: liabilities, provisions, total capital, and average number of employees. Since the study attempts to demonstrate the viability/utility of ML in predicting the economic/financial results of companies both through regression and classification algorithms, it was necessary to establish two dependent (output) variables:

- Gross profit was set as the dependent variable for the regression algorithms. Gross profit was taken directly from the financial statements.
- Profitability class, determined on the basis of return on assets (ROA), as a dependent variable for the classification algorithms. ROA was calculated based on existing data in the financial statements according to the formula:

$$ROA = \frac{Net\ income}{Total\ assets} * 100 \quad (1)$$

Based on the ROA value, companies were grouped into three profitability classes which will require a multinomial classification machine learning method (Table 1).

Table 1. Profitability classes used in this study

Class	ROA
0 (*poor*)	under 5%
1 (*good*)	between 5% and 20%
2 (*excellent*)	over 20%

In conclusion, the ML models in our study use the following data (Table 2):

Table 2. Input and output variables related to *ML* models

Independent variables	Dependent variable	*ML* method
Liabilities (L) Provisions (P) Total capital (TC) Average number of employees (ANE)	Gross profit (GP)	Regression
	Class of profitability, determined based on *ROA* (CP)	Classification

In Table 3, a statistical description of the variables used in the research is presented. A large dispersion of values is evident, which, in our opinion, will seriously affect some specific metrics of ML algorithms (e.g., mean absolute error, root mean squared error).

Table 3. Statistical characteristics of variables

Variable	Minimum	Maximum	Average	Standard deviation
L	-6,719.84	14,998,888,223.27	131,625,962.27	850,826,658.62
P	0.00	2,522,209,692.24	32,793,019.32	208,534,208.96
TC	-1,238,500,917.70	6,817,615,827.15	183,225,638.04	729,408,899.19
ANE	0.00	23,404.00	1,170.35	2,917.68
GP	-384,866,476.0179	1,282,436,765.11	20,522,025.45	91,653,440.72
CP	0	2	-	-

Before applying the ML algorithms, we carried out an analysis of the correlations existing in the dataset, to identify possible relationships that would support the results generated by the ML models. For this purpose, we calculated *Pearson* correlation coefficients between the input and output variables, coefficients presented in Table 4 (where, maximum values are underlined). For the target variable used in the classification (profitability class, determined according to ROA), the correlations are established in relation to the values used in its determination (net income – NI, total assets – TA). It can be said that, in principle, the input data are relatively correlated with the output data, which is an additional reason for their use in ML algorithms.

Table 4. Correlation coefficients between variables

Inputs / **Outputs**	GP	NI	TA
L	0.0843	0.0737	<u>0.7975</u>
P	0.4670	0.4417	0.6804
TC	<u>0.7807</u>	<u>0.7588</u>	0.7163
ANE	0.3060	0.2821	0.5353

3.3. Machine learning algorithms

In our study, we turned to some of the most popular supervised ML algorithms. Supervised learning methods are regression and classification. Regression is known as a method of quantitative prediction of continuous values (e.g., numerical values), and classification is the method by which discrete values (classes or categories) are predicted. The ML algorithms used in this study are those specified in Section 1 (LinR, LogR, KNN, DT, RF, and MLP), and their application to the regression and/or classification method is described in Table 5.

Table 5. Algorithms used according to ML method

***ML* algorithm**	***ML* method**
LinR	Regression
LogR	Classification
KNN	Regression + Classification
DT	Regression + Classification
RF	Regression + Classification
MLP	Regression + Classification

Linear regression (LinR) predicts the target value (Y) by calling a linear function consisting of the independent variables (X), the corresponding weights (w) adjusted in the model training step and the model error term (ε):

$$Y = w_0 + w_1 X_1 + \cdots + w_n X_n + \varepsilon \quad (2)$$

Logistic regression (LogR) is an algorithm that, despite its name, can be used exclusively in classifications. LogR determines the probability with which a value can be placed in a class (category) by means of a classification function: *sigmoid* (for binary classifications) or *softmax* (for multinomial classifications).

K-nearest neighbor (KNN) is a supervised, nonparametric algorithm that can be used both in regressions and classifications. For regression, the algorithm returns the mean value of the K nearest elements, and in the case of classification, the output value is set as the most frequent value found among the K neighboring elements.

Decision Tree (DT) is an induction-based algorithm, similar to several alternative structures (IF-Then-Else), which can be used for regression or classification. As its name implies, a DT is a graph with a tree structure, where nodes represent conditions imposed on the data, branches indicate concrete values, and end nodes (also called leaves) are outcomes.

Random forest (RF) represents what is also called *ensemble learning*, i.e. a set of decision trees that provides higher accuracy in predictions. RF is a flexible and robust machine learning algorithm, which often generates excellent results even without additional adjustments.

Multi-layer Perceptron (MLP) is a type of artificial neural network (ANN) in which the artificial neurons are arranged in several layers: an input layer, one or more hidden layers, and an output layer. Neurons in one layer communicate with all neurons in the next layer via weighted connections (*w*). The input data is taken by the neurons of the first layer, then processed, and the results are transmitted to the next layer, and so on until reaching the output layer that generates the final results. The outputs of one layer are inputs to the next layer, and communication between layers is allowed in one direction (*feedforward network*).

All the algorithms described above were used in our study via the *scikit-learn* package, implemented in Python. *Scikit-learn* is a very popular library among researchers and offers a wide range of algorithms intended for solving supervised and unsupervised machine learning problems (Pedregosa et al., 2011). For most of the algorithms used, a single hyperparameter with a different value than the default one has been set: *random_state=0*, in order to obtain the same results every time. In the case of MLP, it was necessary to set some essential hyperparameters (number of layers and related neurons), for which purpose we used an iterative *Grid Search* process to obtain the best values, displayed in Table 6.

Table 6. MLP hyperparameters

ML method	MLP layers	Neurons/layer
Regression	2	60 | 29
Classification	2	82 | 11

3.4. Models evaluation

In general, the performance of a machine learning model is determined by comparing the predictions with the actual (known) results in the test set. In this study, the quality of the obtained models was measured by different metrics, depending on the ML method used (regression or classification).

For the regression models, we used the following metrics well known to statisticians: *coefficient of determination* (R^2), *mean absolute error* (MAE), and *root mean squared error* (RMSE). These metrics are commonly used in regression-based models (Chicco et al., 2021; Gregova et al., 2020; Özlem & Tan, 2022).

In measuring the performance of a classification model, *True Positive*, *True Negative*, *False Positive* and *False Negative* predictions must be taken into account. The results of a classification can be visualised and analysed by the confusion matrix, which, in the case of a binary classification model, can be represented as in Table 7:

Table 7. Confusion matrix

Predicted class	**Actual class**	
	Positive	**Negative**
Positive	*True positive (TP)*	*False positive (FP)*
Negative	*False negative (FN)*	*True negative (TN)*

Models for classifications were evaluated by the *accuracy*, *precision*, *recall*, and *F-measure* metrics, considered to be the most popular and relevant (Cano-Ortiz et al., 2022; Hossin & Sulaiman, 2015).

4. Empirical results

Since our study was based on two ML methods (regression and classification), the results will be presented in relation to these two techniques.
In the regression case, we followed the prediction of gross profit (GP) as a function of the values of the independent variables (L, P, TC, ANE). Figure 1 contains graphical representations of the values predicted by each regression model (LinR, KNN, DT, RF, MLP) compared to the actual values of the firms in the test set.

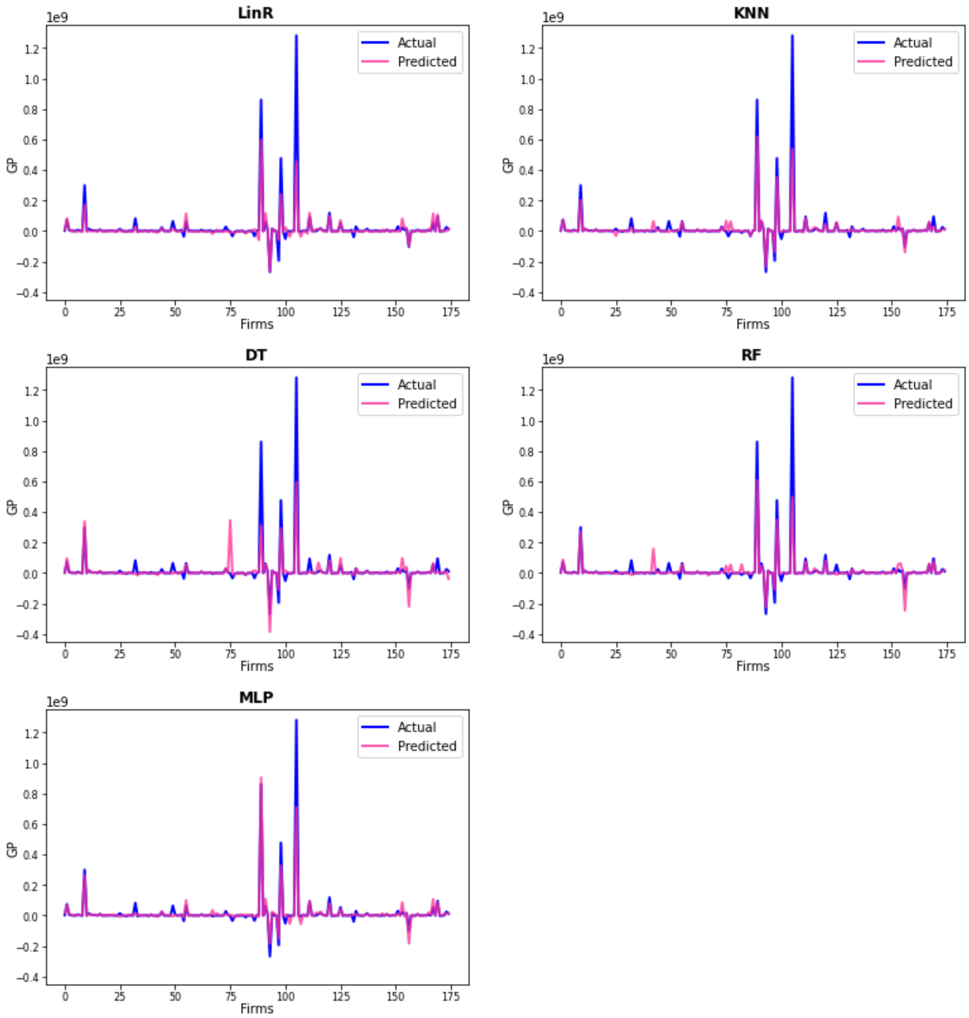

Figure 1. Actual versus predicted gross profit per company

As it can be seen in Figure 1, the MLP model generally gives better predictions both for the usual GP values and for the extreme values. The same conclusion can be drawn from Table 8, which shows the R^2, MAE, and RMSE metrics calculated for the five regression models.

Table 8. Metrics for regression

Rank	Model	R^2	MAE	RMSE
1	MLP	0.856034	1.271078e+07	4.832281e+07
2	KNN	0.758316	1.378153e+07	6.261039e+07
3	RF	0.719451	1.550318e+07	6.745689e+07
4	LinR	0.688324	1.699976e+07	7.110066e+07
5	DT	0.639918	1.910776e+07	7.642277e+07

The coefficient of determination (R^2) is much higher for MLP (about 0.85) compared to the other algorithms, which indicates a very good prediction rate of the model. More precisely, this means that, based on the five predictors, at least 85% of the gross profit dynamics can be estimated by the MLP model. The values for the MAE and RMSE metrics support the same case: the MLP has significantly lower (and better) values than the other models.

The next ranked models are KNN and RF which offer good accuracy in GP prediction (R^2 =0.75, respectively R^2 =0.71), which is also evidenced by the MAE and RMSE values, which indicate positions just below those of MLP.

DT and LinR generate the worst accuracies, which is to be expected, given the above: LinR generally provides a poorer accuracy than other algorithms, and DT is outperformed in predictions by RF. The MAE and RMSE metrics also hold the highest values, justifying once again the lower position occupied by the two models in this ranking. However, judged individually, the R^2 coefficient values (0.63 and 0.68) are not so low as to suggest that, in this case, the DT and LinR algorithms are totally inefficient and unusable.

LogR, KNN, DT, RF, and MLP classification algorithms were used to predict the class (0=*poor*, 1=*good*, 2=*excellent*) in which a company falls according to the value of the ROA indicator.

In Figure 2 are shown the confusion matrices for the five ML models. Each element m_{ij}, represents the number of companies (in the test set) that belong to class i but are predicted in class j. Thus, for class 0 (*poor profitability*), the best predictions are generated by RF and MLP (114 firms), followed by KNN with 112 firms, the last ranked being DT and LogR, which recognise 98 and 85 firms in this category, respectively. For class 1 (*good profitability*), good predictions are generated by RF and DT (30 and 29 firms, respectively), followed by MLP (21 firms), KNN and LogR (with 20 firms each). Regarding class 2 (*excellent profitability*) it can be said that it is the least targeted by the models analysed, most of them including only one company (out of six) in this category, MLP not being able to recognise any.

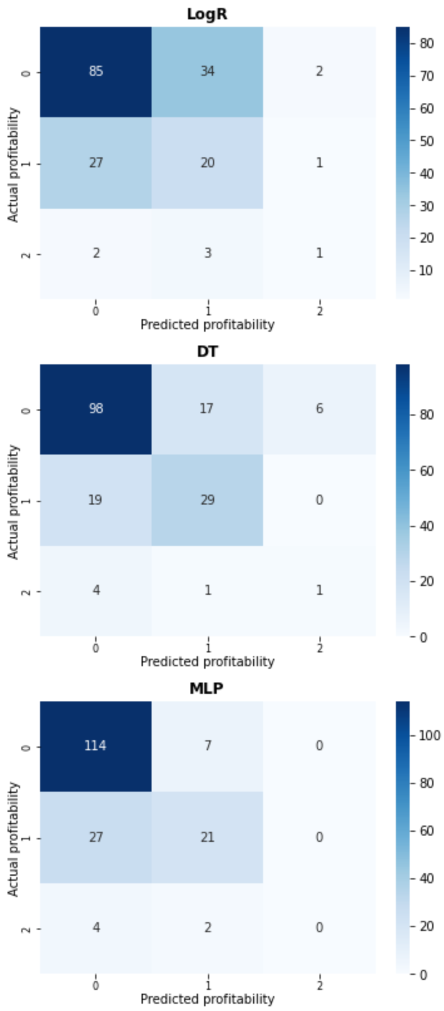

Figure 2. Confusion matrix

The metrics calculated for the five classification models are shown in Table *9*.

Table 9. Metrics of classification

Rank	Model	Accuracy	FM	Precison	Recall
1	RF	0.828571	0.625106	0.880976	0.577938
2	KNN	0.777143	0.545679	0.690063	0.507059
3	MLP	0.771429	0.465201	0.495402	0.459883
4	DT	0.731429	0.524763	0.523265	0.526917
5	LogR	0.605714	0.434786	0.44883	0.428604

In the classification case, RF shows the best values for all four metrics. The accuracy value (0.82) is very good for a multinomial classification model, and, by their values, the other three metrics (*FM*, *Precision*, and *Recall*) lead to the same conclusion: RF is the best performing of the analysed models. The probability that a company is correctly classified in the profitability class is 82%, and the *precision* value of 0.88 indicates a relatively small number of FP predictions, which is a positive aspect. Even if the *recall* value is the best among the existing ones (0.57), it still expresses a rather high tendency of RF to generate FN predictions.

With an accuracy of 0.77 (77%), the KNN demonstrates good effectiveness in classification problems, ranking second, with FM and *precision* metrics justifying this position. However, even in this case, the *recall* value is quite low, even lower than the fourth ranked model (DT).

MLP is only third with an accuracy very close to that of KNN, but with the other metrics below it.

DT and LogR provide reasonable values for accuracy (0.73 and 0.60, respectively), but are notable for low values for the other metrics, so they cast some doubt on their effectiveness in classifications of the type examined in our research.

5. Conclusions

For any company, performance prediction is an important issue and forms the basis for determining future business activity. The use of ML in economics and, especially, in finance allows both researchers and practitioners to gain new insights into financial data (Husmann et al., 2022), which is also very useful in managerial decision-making.

Our study demonstrates the usefulness of regression and classification ML algorithms in predicting the economic and financial performance of Romanian companies that prepare IFRS financial statements. Following the evaluation of the metrics for the ML algorithms related to the two methods, the best results were obtained by the MLP model (R^2 =0.856), for regression, and the RF model (*accuracy=0.*828), for classifications. Analysing the research results, the RF and MLP algorithms proved to be effective tools to predict performance, while demonstrating their superiority over other ML algorithms. In this sense, our research results are in agreement with those obtained by Fatih ECER (Ecer, 2013a) and Gregova et al. (Gregova et al., 2020).

Following our research, it appears that both regression and classification, as supervised machine learning methods, can be successfully used to predict the economic and financial performance of Romanian companies, using, as main predictors, indicators from financial statements prepared in accordance with IFRS. Researching the literature, we found that most studies on predicting economic and financial performance use as predictors financial ratios, and less often gross values taken directly from financial statements (balance sheet, profit and loss account, etc.).

From this point of view, our study is a plus in the literature, as it does not use, as independent variables, such financial ratios.

As this study was conducted on a relatively limited dataset, for future studies, we plan to extend the research to other companies in Romania (which, for the time being, do not report under IFRS) or even to companies from other countries.

REFERENCES

[1] **Alaka, H.A., Oyedele, L.O., Owolabi, H.A., Kumar, V., Ajayi, S. O., Akinade, O.O., Bilal, M. (2018),** ***Systematic review of bankruptcy prediction models: Towards a framework for tool selection.*** *Expert Systems with Applications*, 94, 164–184, https://doi.org/10.1016/j.eswa.2017.10.040;

[2] **Athey, S., Imbens, G.W. (2019).** ***Machine Learning Methods That Economists Should Know About.*** *Annual Review of Economics*, 11(1), 685–725, https://doi.org/10.1146/annurev-economics-080217-053433;

[3] **Brown, S. (2021), Ideas made to matter - artificial intelligence.** ***Machine learning, explained,*** April 2021. *MIT MANAGEMENT SLOAN SCHOOL*, https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained;

[4] **Budak, A., Sarvari, P.A. (2021),** ***Profit margin prediction in sustainable road freight transportation using machine learning.*** *Journal of Cleaner Production*, 314, 127990, https://doi.org/10.1016/j.jclepro.2021.127990;

[5] **Bussmann, N., Giudici, P., Marinelli, D., Papenbrock, J. (2021),** ***Explainable Machine Learning in Credit Risk Management.*** *Computational Economics*, 57(1), 203–216, https://doi.org/10.1007/s10614-020-10042-0;

[6] **Cano-Ortiz, S., Pascual-Muñoz, P., Castro-Fresno, D. (2022)**, ***Machine learning algorithms for monitoring pavement performance.*** *Automation in Construction*, 139, 104309, https://doi.org/10.1016/j.autcon.2022.104309;

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

[8] **Clement, C. (2020),** ***Machine learning in bankruptcy prediction - a review.*** *Journal of Public Administration, Finance and Law*, 17, article number: 17;

[9] **Delen, D., Kuzey, C., Uyar, A. (2013),** ***Measuring firm performance using financial ratios: A decision tree approach***. *Expert Systems with Applications*, 40(10), 3970–3983, https://doi.org/10.1016/j.eswa.2013.01.012;

[10] **Dia, I., Ahvar, E., Lee, G.M. (2022),** ***Performance Evaluation of Machine Learning and Neural Network-Based Algorithms for Predicting Segment Availability in AIoT-Based Smart Parking***. *Network*, 2(2), article number: 2, https://doi.org/10.3390/network2020015;

[11] **Ecer, F. (2013a),** ***Artificial neural networks in predicting financial performance: An application for Turkey's top 500 companies.*** *Economic Computation and Economic Cybernetics Studies and Research*, http://www.ecocyb.ase.ro/20132/Fatih%20Ecer%20(T).pdf;

[12] **Ecer, F. (2013b),** ***Comparing the Bank Failure Prediction Performance of Neural Networks and Support Vector Machines: The Turkish Case.*** *Economic Research-Ekonomska Istraživanja*, 26(3), 81–98, https://doi.org/10.1080/1331677X.2013.11517623;

[13] **Gregova, E., Valaskova, K., Adamko, P., Tumpach, M., Jaros, J. (2020),** ***Predicting Financial Distress of Slovak Enterprises: Comparison of Selected Traditional and Learning Algorithms Methods.*** *Sustainability*, 12(10), 3954, https://doi.org/10.3390/su12103954;

[14] **Hoang, D., Wiegratz, K. (2022),** ***Machine Learning Methods in Finance: Recent Applications and Prospects.*** *SSRN Electronic Journal*;

[15] **Hossin, M., Sulaiman, M.N. (2015),** ***A Review on Evaluation Metrics for Data Classification Evaluations.*** *International Journal of Data Mining & Knowledge Management Process*, 5(2), article numer: 2, https://doi.org/10.5121/ijdkp.2015.5201;

[16] **Husmann, S., Shivarova, A., Steinert, R. (2022),** ***Company classification using machine learning.*** *Expert Systems with Applications*, 195, 116598, https://doi.org/10.1016/j.eswa.2022.116598;

[17] **Krusinskas, R., Lakstutiene, A., Stankeviciene, J. (2022),** ***The research of reliability of bankruptcy prediction models in Lithuanian companies.*** *Transformations in Business & Economics*, 13(32), http://www.transformations.knf.vu.lt/32/article/ther;

[18] **Lam, M. (2004),** ***Neural network techniques for financial performance prediction: Integrating fundamental and technical analysis.*** *Decision Support Systems*, 37(4), 567–581, https://doi.org/10.1016/S0167-9236(03)00088-5;

[19] **Li, H., Sun, J. (2009),** ***Predicting business failure using multiple case-based reasoning combined with support vector machine.*** *Expert Systems with Applications*, 36(6), article number: 6, https://doi.org/10.1016/j.eswa.2009.01.013;

[20] **Manogna, R.L., Mishra, A.K. (2021),** ***Measuring financial performance of Indian manufacturing firms: Application of decision tree algorithms.*** *Measuring Business Excellence*, https://doi.org/10.1108/MBE-05-2020-0073;

[21] **Özlem, Ş., Tan, O.F. (2022),** ***Predicting cash holdings using supervised machine learning algorithms.*** *Financial Innovation*, 8(1), 44, https://doi.org/10.1186/s40854-022-00351-8;

[22] **Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D. (2011),** S***cikit-learn: Machine Learning in Python.*** *Machine learning in Python*, 6;

[23] **Qi, E., Deng, M. (2019),** ***R&D Investment Enhance the Financial Performance of Company Driven by Big Data Computing and Analysis.*** *Computer Systems Science and Engineering*, 34(4), article number: 4, https://doi.org/10.32604/csse.2019.34.237;

[24] **Wei, X., Chen, W., Li, X. (2021),** ***Exploring the financial indicators to improve the pattern recognition of economic data based on machine learning.*** *Neural Computing and Applications*, 33(2), 723–737, https://doi.org/10.1007/s00521-020-05094-0;

[25] **Yeh, C.-C., Chi, D.-J., Hsu, M.-F. (2010),** ***A hybrid approach of DEA, rough set and support vector machines for business failure prediction.*** *Expert Systems with Applications*, 37(2), 1535–1541, https://doi.org/10.1016/j.eswa.2009.06.088;

[26] **Young, Z., Steele, R. (2022),** ***Empirical evaluation of performance degradation of machine learning-based predictive models – A case study in healthcare information systems.*** *International Journal of Information Management Data Insights*, 2(1), article number: 1. https://doi.org/10.1016/j.jjimei.2022.100070.