sciendo

Predicting Financial Performance in the Romanian Transportation Sector: A Machine Learning Approach

Ana-Maria MARCU
Bucharest University of Economic Studies, Bucharest, Romania
anamaria.marcu@csie.ase.ro

Adrian DOMENTEANU
Bucharest University of Economic Studies, Bucharest, Romania
domenteanuadrian18@stud.ase.ro

Georgiana-Alina CRIȘAN
Bucharest University of Economic Studies, Bucharest, Romania
crisangeorgiana18@stud.ase.ro

Camelia DELCEA*
Bucharest University of Economic Studies, Bucharest, Romania
**Corresponding author, camelia.delcea@csie.ase.ro*

Abstract. *Since the beginning of the 21st century, Romania has experienced economic growth, driven by the liberalization of various industries. This shift has led to increased corporate profits, which in turn have contributed to a rise in Gross Domestic Product (GDP) and higher salaries for employees. Machine Learning and Artificial Intelligence have emerged as some of the most impactful technologies in recent years, due to their ability to analyze vast amounts of data and automate complex tasks. This research focuses on applying Machine Learning and Artificial Intelligence algorithms to analyze the most representative transportation companies in Romania. The study explores various financial indicators, including shareholders' equity, liabilities, number of employees, turnover, net profit, fixed assets, and current assets, as well as location-related data such as county, city, and date of establishment. A comprehensive data analysis approach has been implemented, beginning with data cleaning, followed by exploratory data analysis to identify patterns and correlations between variables through interactive visualizations. Furthermore, multiple Machine Learning algorithms have been developed to predict the net profit of these companies based on independent features, with model performance evaluated using specific metrics.*

Keywords: Transportation, Romania, Machine Learning, Artificial Intelligence, Finance, Python

Introduction

Romania's economic development has been substantial when comparing the period of 1989 to the present. In December 1989, Romania had approximately 8 million employees, with an average salary of $250, exports valued at $10 billion, and a Gross Domestic Product (GDP) of $54 billion. By 2008, the GDP had grown to approximately $137.2 billion, and this upward trend continued, reaching $243.32 billion in 2018—an increase of $33 billion compared to 2017(Lupu, 2018; Trading Economics, 2025). During the COVID-19 pandemic, Romania's economy, like the global

DOI: 10.2478/picbe-2025-0049

economy, was significantly affected, with periods of industrial sector shutdowns. By 2024, the average salary in Romania had risen to $1,865.79 (Roman, 2024).

In 2019, Romania ranked 12th in the European Union (EU) in terms of the total length of communal roads, which amounted to 85,525 kilometers. The country's motorway network has also expanded significantly over the years. In 1990, Romania had only 113 kilometers of motorways, increasing to 281 kilometers by 2007. By 2019, this figure had risen to 866 kilometers, and by the end of 2024, the total motorway length reached 1,214 kilometers, driven by substantial investments from the Romanian government in the transportation sector (Dumitrescu, 2024; Neagu et al., 2022). The increase in motorway length between 1990 and 2007 represents approximately 148%, while between 2007 and 2019, the growth reached 208%, one of the highest rates in the EU. Given Romania's strategic geographical position on the Black Sea, the transportation sector has the potential to benefit from further significant investments, facilitating the movement of goods and services across the region.

Similar with other domains, the transportation sector is very sensitive to external factors, and the best example is the COVID-19 pandemic that occurred in 2020. The number of tourists decreased by 52.16% from 2019 to 2020, and with 30% between 2021 and 2019, resulting in serious economic instability (Nancu et al., 2023).

The railway sector plays a crucial role in the Romanian economy, boasting the second-longest railway network in Europe after Poland. However, it remains highly inefficient due to a chronic lack of investment, making it unsustainable on its own. In 1990, Romania's railway system spanned 22,179 kilometers, but by 2007, this had declined to 20,668 kilometers. By 2019, the total railway length had further decreased to approximately 20,079 kilometers, placing Romania behind Poland, France, and Italy in terms of network size.

Electrified railway tracks are crucial for investment and efficient transport. In 1990, Romania had 3,680 kilometers of electrified railway, increasing to 3,959 in 2007 and 4,029 in 2019—a modest 9.48% growth over 30 years. In contrast, Turkey's network grew by 740%, Hungary 147%, Greece 72.77%, and Spain 55.61% (Haiduc et al., 2023). Although Romania's railway network meets demand, its correlation with infrastructure, income, and transported goods is negative due to poor rail quality, causing inefficiencies and financial losses.

In recent years, Romania has adopted numerous sustainable development principles, emphasizing both economic growth and environmental protection (Romanian Government, 2024). There is a strong correlation between GDP and the transportation sector, highlighting its crucial role in the country's economy. At the same time, Romania has one of the lowest per capita greenhouse gas emissions in the European Union (EU) and ranks among the top countries when comparing emissions levels relative to GDP size (Bocaneala et al., 2024).

Machine Learning (ML) domain incorporates methods used for automation, prediction which facilitates the decision-making process, observing patterns in data and correlation between features which were impossible to be identified by human (Surden, 2021, p. 8). The topic has grown exponentially in the last years, combined with statistics, computer science, artificial intelligence and data science (Jordan & Mitchell, 2015). There are multiple types of ML algorithms, such as supervised learning, unsupervised learning and reinforcement learning (Janiesch et al., 2021). The supervised learning investigates a dataset as input, together with labeled values for the target feature, predicting the output based on the existing data using training processes (Janiesch et al., 2021). Unsupervised learning focuses on predicting the target variables based on the input data which does not have a label or specification, and the algorithm should identify the patterns, while the reinforcement learning presents the system based on a specific goal that should be achieved.

A detailed list with the main steps that should be applied, together with the constraints is defined, minimizing the error and maximizing the outcome (Janiesch et al., 2021). ML techniques have substantial impact on numerous industries, completely redefining the companies all over the world, providing new possibilities of increasing the innovation among companies (Sharifani & Amini, 2023).

Methodology

The data that has been investigated is related to the transportation sector in Romania, referring to 100 most relevant firms, presenting the company name, address, city, county, date of establishment, net profit, fixed assets, turnover, current assets, liabilities, average number of employees, shareholders' equity and the analyzed year. The dataset was downloaded from ListaFirme, the biggest website for Romanian for financial information of companies (Lista Firme, 2025).

In order to conduct the analysis, Python programming language has been used, due to the variety of applications in which it has been used from almost every field. Python is one of the most used languages for data analysis, having an intuitive syntax, one of the biggest communities in the world and numerous libraries that facilitate the implementation of the code (Li, 2022). Python developed libraries which are used in interdisciplinary topics such as geography, computer science or biology (Chapman & Chang, 2000; Li, 2022). NumPy library is used for computing in linguistics, being able to store any type of data or structure. When a data analysis is performed, NumPy is frequently implemented, thanks to the high performance (Li, 2022). Pandas' module is able to read, store and set up data structures in Python, having high flexibility in processing Excel files (Li, 2022). Matplotlib is a library used especially for graphical representation of the data, such as scatter plots, pie, bar or line charts (Li, 2022). Scikit-learn module is very popular, storing numerous ML algorithms and metrics to compare the results of the models. The algorithms that have been implemented during the research are imported from scikit-learn library (Tran et al., 2022). Plotly Express represents a Python library which creates interactive graphical representation of the data, offering more than 30 functions for visualization (Plotly, 2025).

For conducting the analysis, a series of ML algorithms have been considered. In the following, a brief overview of these algorithms is provided.

Decision Tree is an ML algorithm used for classification and regression analysis. The regressor is implemented when a prediction is to scope the research, by using tree structure with nodes. It is one of the main algorithms that are used to predict a continuous feature, thanks to the versatility and accuracy of the model (Baladram, 2024a).

Random Forest represents an upgraded decision tree algorithm, with better understanding of the data, thanks to the possibility of combining multiple trees at the same time. It is one of the most used models by data scientists, since it has high efficiency (Baladram, 2024b).

Support Vector Regressor is part of Support Vector Machine models, mainly being used for regression analysis. The scope of the algorithm is to predict the target variable and to minimize the complexity of the problem, avoiding overfitting. Support Vector Regressor can be implemented also for nonlinear problems, having the scope of defining a balance among the accuracy and complexity (Rasifaghihi, 2023).

Neural Network represents a system that combines multiple elements that are able to investigate information in a dynamic manner, by taking into account the external factors, being inspired by the biological neural networks that exist in the human brain. The main elements in a neural network are nodes or neurons in biology, every node receives data as input and it's analyzing it. Each input that is part of the model has a certain weight, which is crucial to the neural network outcome (Fumo, 2017).

Ridge Regression derived from regression analysis, being very popular when the data has multicollinearity and overfitting. Multicollinearity appears when data is strongly correlated, leading to less accurate results. The Ridge Regression defines the connection between features that have been included in analysis (Gomede, 2024).

The data analysis is conducted as suggested in scientific literature. For example, a training dataset has been first considered. Training dataset contains the labeled information that is used by the algorithm to learn the patterns and connections between features, while the testing dataset evaluates the model after the training step and prediction are finalized, defining the accuracy of the model (Software, 2024).

Furthermore, feature importance has been taken into account. Feature importance consists in a ML technique which assigns a score to each metric that has been part of the data analysis process for a specific algorithm. The score is synonym with the importance of the feature, the higher the score is, the more relevant for the investigation is the variable, providing an overview on the prediction of the target variable. Thanks to the feature importance method, the models are more accurate, making more easy the understanding of the data and the interpretation of the results (Shin, 2024).

As a pre-processing step, encoding variables step has been considered, which focuses on converting the categorical features in numerical values, in order to be included in algorithms, since the majority of ML methods use numbers as input (Jarapala, 2023).

Various evaluation metrics have been used, as highlighted in scientific literature, briefly presented in the following. Mean Squared Error (MSE) calculates the mean of the squared differences among the real values and the predicted ones, while Mean Absolute Error (MAE) calculates the absolute difference between real values and predicted ones (Acharya, 2023). R^2 score represents a correlation metric, and it could have values between -1 and 1. Ideally, the values should be as close as possible to 1, which reflects a strong relationship among the features.

In the case in which it is observed that the data has a strong asymmetry, another step is needed in which a logarithm transformation is applied, which has the purpose of transforming the data into one with normal distribution, which will facilitate the application of statistical and ML methods (Zhao, 2019).

In terms of considered variables, a series of variables have been taken into account as described in the following. First, Shareholders' Equity is a financial metric of the companies that quantifies the strategic impact of collective negotiations, being strongly correlated with the change in stock prices (Becker & Olson, 1986). Current assets refer to raw materials, products that are work-in-progress, cash, while the liabilities represents the amount of money that should be paid to creditors and the accruals (Nobanee & Abraham, 2015). Fixed assets incorporate the equipment, machineries that are used in daily activities of the company in order to produce goods and services (Eickelkamp, 2015). Turnover is a key factor for managers since it measure the outcome of the companies, representing a metric that is taken into account when the strategy is defined (Walsh, 1989). Net profit calculates the financial performance of the firm by calculating the difference between total income and total expenses (Astuti, 2021).

Results

Due to data asymmetry, the information was logged in order to have a distribution as close as possible to the normal one. Figure 1 presents the Shareholders' Equity and Liabilities which have been logged, containing information for the five most frequent counties. When the Shareholders' Equity values increase, the Liabilities also increase, which shows the positive correlation between variables. For instance, for Arad county, on the bottom left bubble, there is a value for Logged Liabilities is 10.46 (34.891,55 RON) and the Logged Shareholders' Equity value is 10.29 (29.436,77 RON), while in the top right part, the Logged Shareholders' Equity value is 17.80 (53.757.835,97 RON), while for the Liabilities the value is 18.52 (110.441.883.56 RON).

Figure 1. Scatter Plot of Logged Shareholders' Equity and Liabilities

Source: Authors' own research.

Figure 2 explores the Logged Net Profit and Fixed Assets features for the transportation companies, reflecting a reduced correlation between variables. For Bucharest county, where the Logged Net Profit is small, having a value of 7.15 (1.274,10 RON), the Logged Fixed Assets are higher, with a value of 12.31 (221.903,97 RON). On the top right part of the graph, the values are 16.08 (9.626.208,60 RON) for Net Profit and 18.30 for Fixed Assets (88.631.687,64 RON).

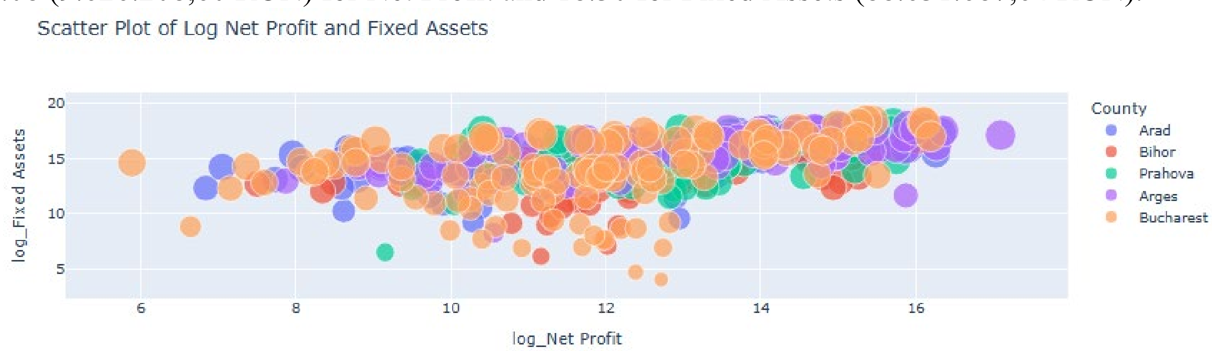

Figure 2. Scatter Plot of Logged Net Profit and Fixed Assets

Source: Authors' own research.

Figure 3 defines the collaboration between Logged Net Profit and Number of Employees. When the Net Profit increases, the Number of Employees slightly increases, which reflects the poor correlation between variables. For instance, for Arges county, on the right side there are bubbles with very high profits, 15.87 (7.802.853.04 RON) and 1 employee, but there are also, 16.35 (12.609.991.07 RON) and 331 employees.

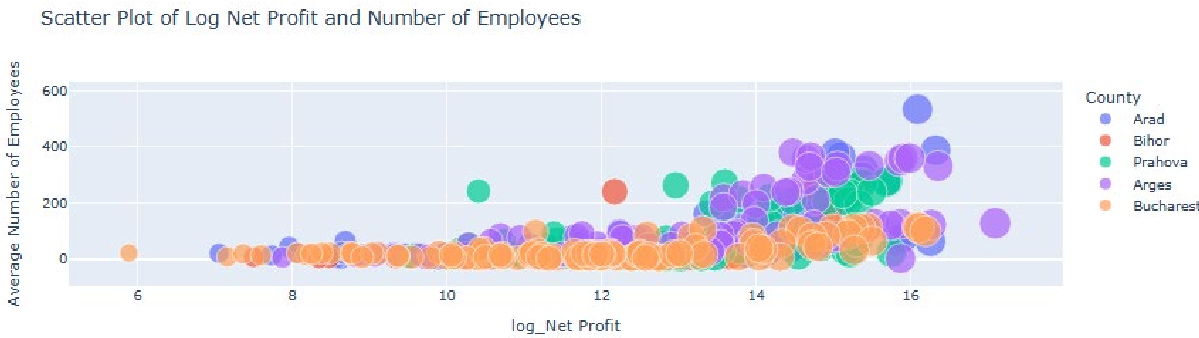

Figure 3. Scatter Plot of Logged Net Profit and Number of Employees

Source: Authors' own research.

Figure 4 illustrates the distribution of Logged Net Profit feature using boxplot graph. The minimum value, which is an outlier, is 2.48 or 11.94 RON. The first line in the boxplot is equivalent with the minimum value that is not an outlier having a value of 6.24 or 512,85 RON. The second line is the first quartile, which means that 25% of the values are below 10.57 or 38.949,67 RON. The second quartile is the median value, showing that half of the profits are below 12.12 or 183.505,51 RON, while the third quartile shows that 75% of the values are below 13.48 or 714.972,95 RON. The maximum value is 17.09 or 26.429.728,12 RON.

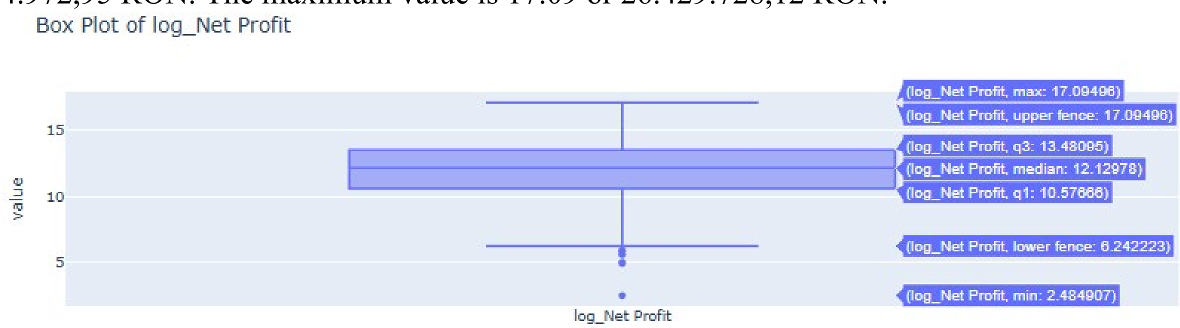

Figure 4. Box Plot of Logged Net Profit

Source: Authors' own research.

Figure 5 explored the Logged Turnover distribution. The lowest value that was registered is 10.34 or 30.946.03,03 RON, which represents an outlier, while the minimum value which is not categorized as an outlier is 11.99 or 161.135,35 RON, while the first quartile value points out that 25% of the company's turnover is lower than 14.82 or 2.730.512,83 RON. The median value is 15.69 or 6.517.490,71 RON, while the third quartile value is 16.71, which means that only 25% of the values are higher than 18.074.271,12 RON. The maximum value is 19.53 or 303.229.347,62 RON. The top outlier is 19.91 which is equivalent to 443.407.602,09 RON.

DOI: 10.2478/picbe-2025-0049, pp. 611-624, ISSN 2558-9652 |
Proceedings of the 19th International Conference on Business Excellence 2025

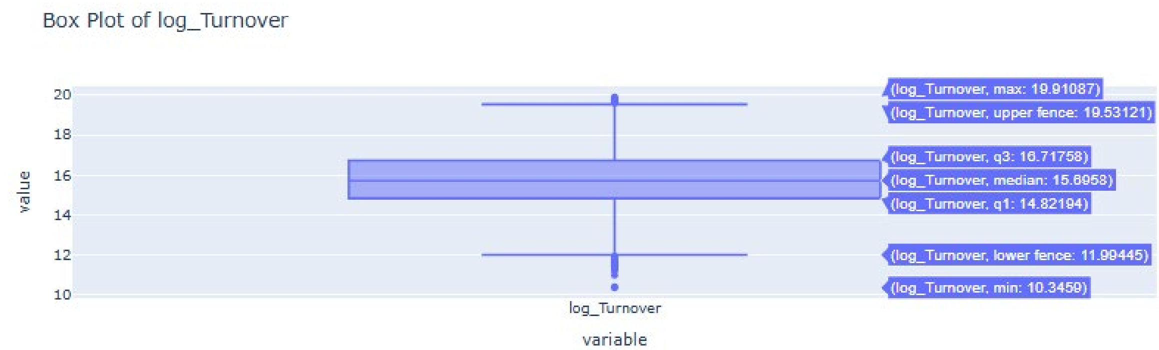

Figure 5. Box Plot of Logged Turnover

Source: Authors' own research.

Figure 6 focuses on the Logged Shareholders' Equity. The smallest outlier that has been discovered has a value of 6.54 or 692,28 RON, while the lowest value which is not considered outlier is 9.42 or 12.332,58 RON. The first quartile has a value of 12.93 which is equivalent to 412.503,51 RON, and the median value or second quartile has a value of 14.26 which is 1.559.693,67 RON after applying the exponential function. The third quartile value is 15.30 or 4412711,89 RON, while the maximum value is 18.12, which is equivalent to 74.031.408,47 RON.

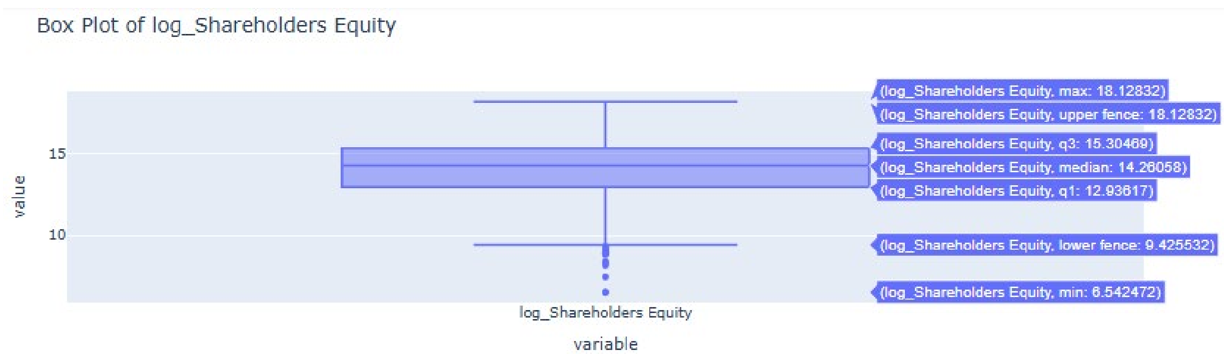

Figure 6. Box Plot of Logged Shareholders' Equity

Source: Authors' own research.

Figure 7 details the distribution of Logged Fixed Assets. There is a left-handed asymmetry, since the majority of the values are on the right part of the graph. The most frequent values are between 14 (1.202.604,28 RON) and 14.5 (1.982.759,26 RON), with a frequency of 202, while the frequency between 14.5 (1.982.759,26 RON) and 15 (3.269.017,37 RON) is 201. There are also values between 4 (54,59 RON) and 10 (22.026,46 RON) but with a very small impact.

DOI: 10.2478/picbe-2025-0049, pp. 611-624, ISSN 2558-9652 |
Proceedings of the 19th International Conference on Business Excellence 2025

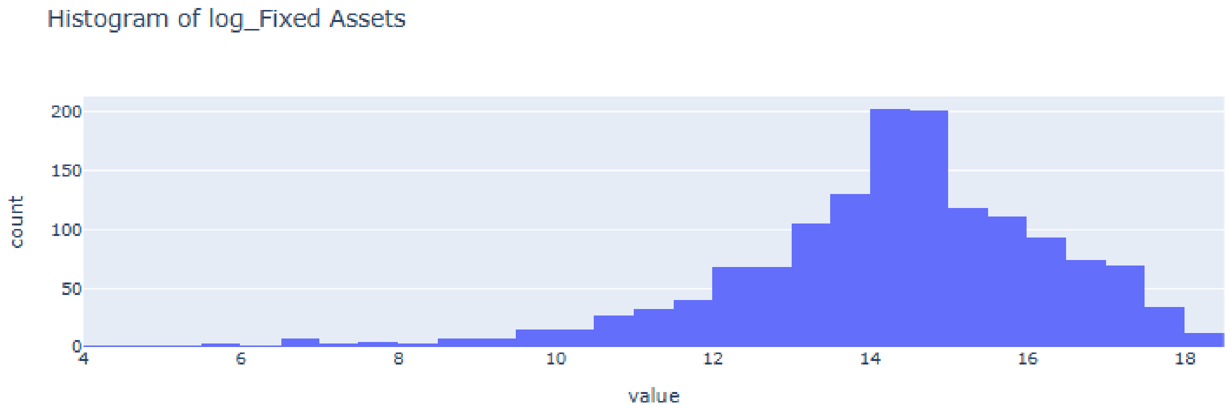

Figure 7. Histogram of Logged Fixed Assets

Source: Authors' own research.

Figure 8 explores the distribution of Logged Current Assets. There is a very small left-handed asymmetry, the values are majority in the right part of the graph, but very close to the mean. The most frequent values are between 14.2 (1.468.864,18 RON) and 14.39 (1.776.223,43 RON) with a frequency of 101, and 82 appearances between 15.2 (3.992.786,83 RON) and 15.39 (4.828.275,87 RON). Between 4 (54,59 RON) and 13 (442.413,39 RON) there are very few appearances, similar between 17 (24.154.952,75 RON) and 18 (65.659.969,13 RON).

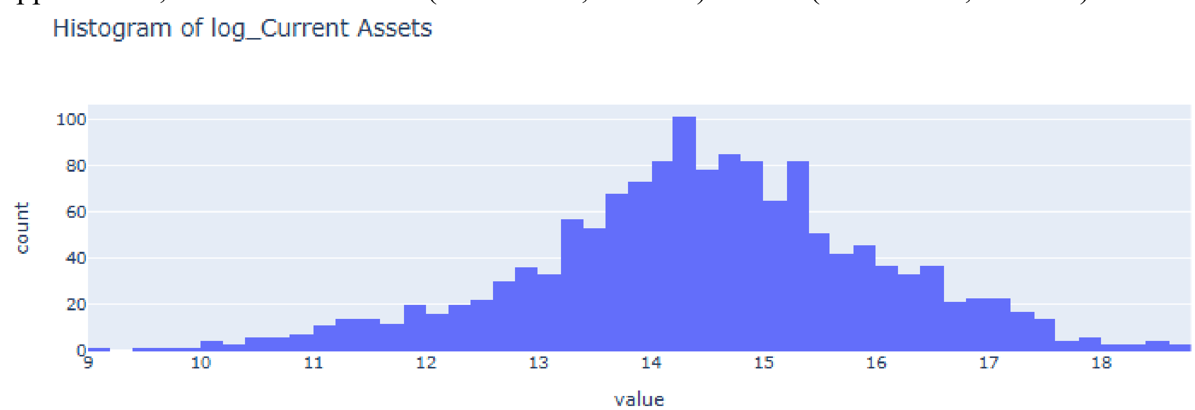

Figure 8. Histogram of Logged Current Assets

Source: Authors' own research.

Figure 9 incorporates the distribution of Logged Liabilities, the majority of values being between 14 and 15. The highest frequency, with 84 appearances, is between 14.2 and 14.39, while the next most frequent values are between 14.40 (1.794.074,77 RON) and 14.59 (2.169.484,19 RON) with 79 occurrences and between 15.00 (3.269.017,37 RON) and 15.19 (3.953.057,94 RON) with 78 occurrences.

DOI: 10.2478/picbe-2025-0049, pp. 611-624, ISSN 2558-9652 |
Proceedings of the 19th International Conference on Business Excellence 2025

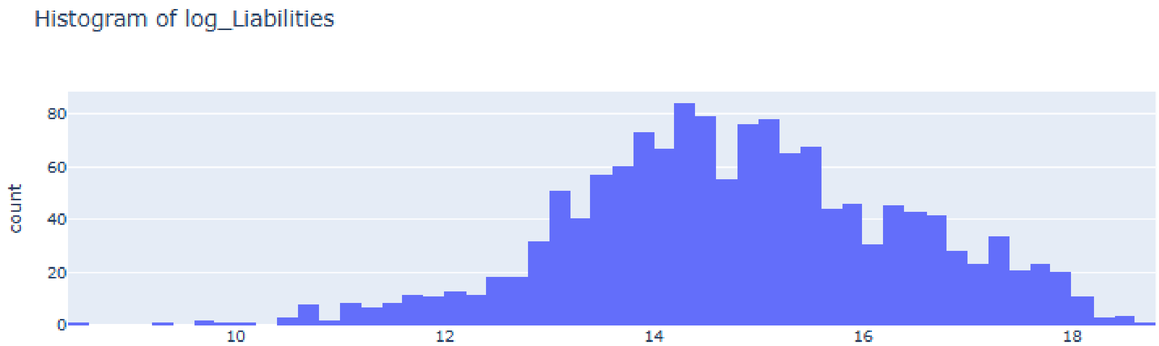

Figure 9. Histogram of Logged Liabilities

Source: Authors' own research.

In order to perform the ML prediction, the model has been divided into two subgroups, one for training the algorithm and one for testing. The size of the groups is 80% for training and 20% for testing. The target variable in the analyzed scenario is the log Net Profit, while the independent variables that will predict the results are Year, Average Number of Employees, City_encoded, County_encoded, log_Fixed Assets, log_Turnover, log_Current Assets, log_Liabilities, log_Shareholders Equity.

Table 1 incorporates the metrics that have been used in order to measure and classify the models based on the accuracy, using MSE, MAE, R^2 score. Since the models could overfit, a test has been performed, in order to assure the quality of the results. If the Training MSE value is higher than Testing MSE, then it's considered overfitting, else there is no overfitting.

For the Ridge Regression model, the MSE value is 1.94 for training and 1.96 for testing. The MAE value is 1.07, which is relatively small compared with MSE, but high compared with the rest of the models, while the R^2 Score for testing is 0.54, while for training is 0.55.

The Neural Network model has a training error of 1.19, while for testing the error is higher, 1.67, so there is no overfitting. The MAE is 0.96, with a R^2 Score of 0.61 for training, and for testing, the value is 0.73.

The decision tree algorithm has an error for training dataset of 1.31, while for testing is 1.98, resulting in no overfitting. The MAE error is 1.05, while the R^2 Score is 0.53 for testing and 0.69 for training.

Random Forest is the algorithm with the smallest error and highest R^2 Score. The training MSE error is only 0.86, with a testing MSE error of 1.45, while the MAE value is 0.92. The R^2 Score metric score is 0.67 for testing and for training is 0.80.

The last algorithm that has been tested is Support Vector Regressor, which does not overfit, since the training error (1.62) is smaller than the testing error (1.63), having the closest errors between testing and training among all five algorithms. The MAE value is 0.92, while the R^2 Score is 0.61 for testing and 0.62 for training. In this scenario, a standing for algorithms based on errors and R^2 Score has been performed:

1. Random Forest (most reduced values for all error metrics and highest R^2 Scores)
2. Neural Network (second smallest values for all error features and second highest R^2 Scores)
3. Support Vector Regressor (third smallest values for all error features and third highest R^2 Scores)

4. Decision Tree (Training MSE, Testing MSE and MAE values are smaller compared to Ridge Regression, but R^2 Score in testing is smaller than Ridge Regression)
5. Ridge Regression (R^2 Score on testing is higher than Decision Tree, but due to the error metrics and R^2 Score on training, is classified in last place)

Table 1. ML algorithms metrics

Model	Training MSE	Testing MSE	MAE	R^2 Score Training	R^2 Score Testing	Overfitting
Ridge Regression	1.94	1.96	1.07	0.55	0.54	No
Neural Network	1.19	1.67	0.96	0.73	0.61	No
Decision Tree	1.31	1.98	1.05	0.69	0.53	No
Random Forest	0.86	1.45	0.92	0.80	0.67	No
Support Vector Regressor	1.62	1.63	0.92	0.62	0.61	No

Source: Authors' own research.

Figure 10 presents the distribution of importances for the features that have been included in ML algorithms. The (A) figure shows the most significant features for the Decision Tree model. In first place is log_Shareholders' Equity metric with an importance of 0.575, followed by log_Current Assets with a value of 0.206, while in third place is log_Turnover with only 0.082. The city of the company and county have almost the same importance, with 0.044 and 0.043 respectively. Log_Liabilities, log_Fixed Assets and Year have the least importance, having 0.029, 0.012, and 0.01, while the Average Number of Employees does not have impact. In figure (B) are described the features importance based on Random Forest algorithm. In the first place is log_Shareholders Equity with a value of 0.24, followed by log_Current Assets with 0.205 and log_Turnover with 0.187. In fourth place is the Average Number of Employees with a value of 0.131, while log_Liabilities have 0.068, log_Fixed Assets have 0.056, County have 0.043, City have 0.04, and Year have 0.03. The figure (C) presents the results of Support Vector Regressor algorithm, which has in first place of feature importances the log_Shareholders' Equity with a value of 0.35, with log_Turnover in second place with 0.279, log_Current Assets with a value of 0.145, while the rest of the variables have a smaller impact, but are still worthy to be mentioned: Year with a value of 0.05, log_Liabilities with 0.042, log_Fixed Assets with 0.036, County with 0.033, Average Number of Employees with 0.03 and City with 0.025. The fourth figure, (D) includes the output of Neural Network model. The most representative feature is log_Turnover with a value of 0.834, followed by log_Liabilities with a value of 0.613, log_Current Assets with 0.578, log_Shareholders Equity which has a contribution of 0.316, Average Number of Employees with 0.253, log_Fixed Assets with 0.209. Year, County and City have the least impact, having a contribution of 0.095, 0.074 and 0.05. The last figure, (E) contains the results of Ridge Regression feature importance, with log_Turnover in first place, with a contribution of 0.804, followed by log_Shareholders Equity, with a contribution of 0.71. In third place is log_Liabilities which has a value of 0.46, log_Fixed Assets has a value of 0.43. The remaining variables have a reduced impact but are relevant for the scope of the research: Year has a contribution of 0.091, County 0.051, Average Number of Employees 0.048, City 0.036, while log_Fixed Assets has the most reduced impact, with a value of only 0.009.

Figure 10 Features importance for Machine Learning algorithms: (A) Decision Tree, (B) Random Forest, (C) Support Vector Regressor, (D) Neural Network, (E) Ridge Regression

Source: Authors' own research.

Table 2 presents the first five predictions of all five ML algorithms and the real values. The Ridge Regression predictions have a significant error of 1.41 (the division between real values and predicted values have been calculated, followed by the average of the errors). The same process was applied to all ML algorithms, resulting in an error of 0.48 for Neural Network model, 2.90 for Decision Tree model, 0.87 for Random Forest algorithm, and 0.71 for Support Vector Regressor algorithm.

Table 2 Prediction of Machine Learning algorithms compared to real values for the first five values

Real Values	Ridge Regression prediction	Neural Network prediction	Decision Tree prediction	Random Forest prediction	Support Vector Regressor prediction
2.669.637,00	3.774.540,54	3.975.014,89	4.122.137,73	2.453.559,74	1.866.787,80
3.291,00	72.697,25	20.598,72	2.808,04	16.468,86	96.123,91
288.497,00	51.930,66	61.241,62	85.141,71	54.593,51	132.575,68
55.914,00	36.177,98	46.729,04	75.320,96	48.502,86	57.860,70
2.285.582,00	1.291.206,53	933.421,34	523.914,98	9.66.531,59	1.403.698,61

Source: Authors' own research.

Conclusion

First, it should be mentioned that the research faced several limitations which might have affected the research outcomes, such as the fact that the data is taken from ListaFirme, which includes only annual financial details for some transportation companies or that, while Python is widely used for data analysis, it has library and model limitations.

The transportation sector has grown since the 21st century, reflected in financial data. Most metrics showed above-average values, indicating positive firm evolution. Feature importance varied across models in predicting log_Net_Profit, with log Shareholders' Equity and log_Current_Assets having the most impact. The Random Forest model achieved the highest accuracy and lowest error across datasets.

ML effectively predicts transportation firms' financial performance, ranking features by importance and assessing accuracy. Future research will explore correlations between financial data and macroeconomic factors like GDP, highway networks, railway electrification, inflation, and unemployment to better understand sector dynamics.

Acknowledgement

This work was funded by the EU's NextGenerationEU instrument through the National Recovery and Resilience Plan of Romania - Pillar III-C9-I8, managed by the Ministry of Research, Innovation and Digitization, within the project entitled „Place-based Economic Policy in EU's Periphery – fundamental research in collaborative development and local resilience. Projections for Romania and Moldova (PEPER)", contract no. 760045/23.05.2023, code CF 275/30.11.2022. This paper was co-financed by The Bucharest University of Economic Studies during the PhD program.

References

Acharya, N. (2023, August 30). *Choosing Between Mean Squared Error (MSE) and Mean Absolute Error (MAE) in Regression: A Deep Dive*. Medium. https://medium.com/@nirajan.acharya777/choosing-between-mean-squared-error-mse-and-mean-absolute-error-mae-in-regression-a-deep-dive-c16b4eeee603

Astuti, W. (2021). A Literature Review of Net Profit Margin. *Social Science Studies*, *1*(2), Article 2. https://doi.org/10.47153/sss12.2262021

Baladram, S. (2024a, November 30). *Decision Tree Regressor, Explained: A Visual Guide with Code Examples*. Towards Data Science. https://medium.com/towards-data-science/decision-tree-regressor-explained-a-visual-guide-with-code-examples-fbd2836c3bef

Baladram, S. (2024b, November 30). *Random Forest, Explained: A Visual Guide with Code Examples*. Towards Data Science. https://medium.com/towards-data-science/random-forest-explained-a-visual-guide-with-code-examples-9f736a6e1b3c

Becker, B. E., & Olson, C. A. (1986). The Impact of Strikes on Shareholder Equity. *ILR Review*, *39*(3), 425–438. https://doi.org/10.1177/001979398603900309

Bocaneala, A. M., Manole, D. S., Gherasim, E. A., Motorga, B., & Iliescu, L. C. (2024). European Funding for Sustainable Transport Systems—Influencing Factor of Regional Economic Development in Romania. *Economies*, *12*(3), Article 3. https://doi.org/10.3390/economies12030051

Chapman, B., & Chang, J. (2000). Biopython: Python tools for computational biology. *SIGBIO Newsl.*, *20*(2), 15–19. https://doi.org/10.1145/360262.360268

Dumitrescu, R. (2024, December 16). *Romania Insider*. Romania Insider. https://www.romania-insider.com/romania-opens-highway-segment-sibiu-pitesti-december-2024

Eickelkamp, T. (2015). Significance of fixed assets in life cycle assessments. *Journal of Cleaner Production*, *101*, 97–108. https://doi.org/10.1016/j.jclepro.2015.03.075

Fumo, J. (2017, October 22). A Gentle Introduction To Neural Networks Series—Part 1. *Towards Data Science*. https://medium.com/towards-data-science/a-gentle-introduction-to-neural-networks-series-part-1-2b90b87795bc

Gomede, E. (2024, January 2). *Understanding Ridge Regression in Statistical Analysis*. The Modern Scientist. https://medium.com/the-modern-scientist/understanding-ridge-regression-in-statistical-analysis-de42947ed18b

Haiduc, A., Neagu, O., & Nicoara, S. (2023). The Evolution Of Railway Transport Infrastructure in Romania After 1990. *THE ANNALS OF THE UNIVERSITY OF ORADEA. ECONOMIC SCIENCES*, *32*, 33–47. https://doi.org/10.47535/1991AUOES32(2)003

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685–695. https://doi.org/10.1007/s12525-021-00475-2

Jarapala, K. N. (2023, March 27). *Categorical Data Encoding Techniques*. AI Skunks. https://medium.com/aiskunks/categorical-data-encoding-techniques-d6296697a40f

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aaa8415

Li, Y. (2022). Python Data Analysis and Attribute Information Extraction Method Based on Intelligent Decision System. *Mobile Information Systems*, *2022*(1), 2495166. https://doi.org/10.1155/2022/2495166

Lista Firme. (2025). *Romanian database website for companies*. Romanian database website for companies. https://www.listafirme.ro/

Lupu, N. (2018). *The European Capital in the 20th Century—Between Agony and Ecstasy The Place of Romania in the European Economic and Political Context—ProQuest*. *21*(1), 5–23.

Nancu, D., Mieila, M., Manole, A. M., & Isbasoiu, G. D. (2023). The Impact of COVID-19 on Tourist Destinations in Romania. *Proceedings of the International Conference on Business Excellence*, *17*(1), 549–560. https://doi.org/10.2478/picbe-2023-0052

Neagu, O., Haiduc, A., & Nicoară, S. (2022, October 31). The Evolution of Road Transport Infrastructure In Romanian After 1990. Economic Implications. *The Annals of the University of Oradea. Economic Sciences*. https://anale.steconomiceuoradea.ro/

en/2022/10/31/the-evolution-of-road-transport-infrastructure-in-romania-after-1990-economic-implications/

Nobanee, H., & Abraham, J. (2015). Current assets management of small enterprises. *Journal of Economic Studies*, *42*(4), 549–560. https://doi.org/10.1108/JES-02-2013-0028

Plotly. (2025). *Plotly express in Python*. https://plotly.com/python/plotly-express/

Rasifaghihi, N. (2023, November 14). *Support Vector Regressor—Theory and Coding Exercise in Python*. Medium. https://medium.com/@niousha.rf/support-vector-regressor-theory-and-coding-exercise-in-python-ca6a7dfda927

Roman, M. (2024, August 23). *Average Salary in Romania*. Timeular. https://timeular.com/average-salary/romania/

Romanian Government. (2024). *Romania Sustainable Development Strategy*. ESDN. https://www.esdn.eu/country-profiles/detail?tx_countryprofile_countryprofile%5Baction%5D=show&tx_countryprofile_countryprofile%5Bcontroller%5D=Country&tx_countryprofile_countryprofile%5Bcountry%5D=23&cHash=16646567f4a0bcb4238f9b312843a36b

Sharifani, K., & Amini, M. (2023). *Machine Learning and Deep Learning: A Review of Methods and Applications*. https://papers.ssrn.com/abstract=4458723

Shin, T. (2024). *Understanding Feature Importance in Machine Learning*. Built In. https://builtin.com/data-science/feature-importance

Software, E. (2024, September 26). *Understanding Training, Validation, and Testing Data in ML*. Medium. https://medium.com/@eastgate/understanding-training-validation-and-testing-data-in-ml-c297e53d6096

Surden, H. (2021). *Chapter 8: Machine learning and law: An overview*. https://www.elgaronline.com/edcollchap/edcoll/9781788972819/9781788972819.00014.xml

Trading Economics. (2025). *Romania GDP Evolution*. https://tradingeconomics.com/romania/gdp

Tran, M.-K., Panchal, S., Chauhan, V., Brahmbhatt, N., Mevawalla, A., Fraser, R., & Fowler, M. (2022). Python-based scikit-learn machine learning models for thermal and electrical performance prediction of high-capacity lithium-ion battery. *International Journal of Energy Research*, *46*(2), 786–794. https://doi.org/10.1002/er.7202

Walsh, J. P. (1989). Doing a deal: Merger and acquisition negotiations and their impact upon target company top management turnover. *Strategic Management Journal*, *10*(4), 307–322. https://doi.org/10.1002/smj.4250100402

Zhao, K. (2019). *The Magic of Log*. Medium. https://medium.com/@kaizhao44/the-magic-of-log-b3e69acd314f