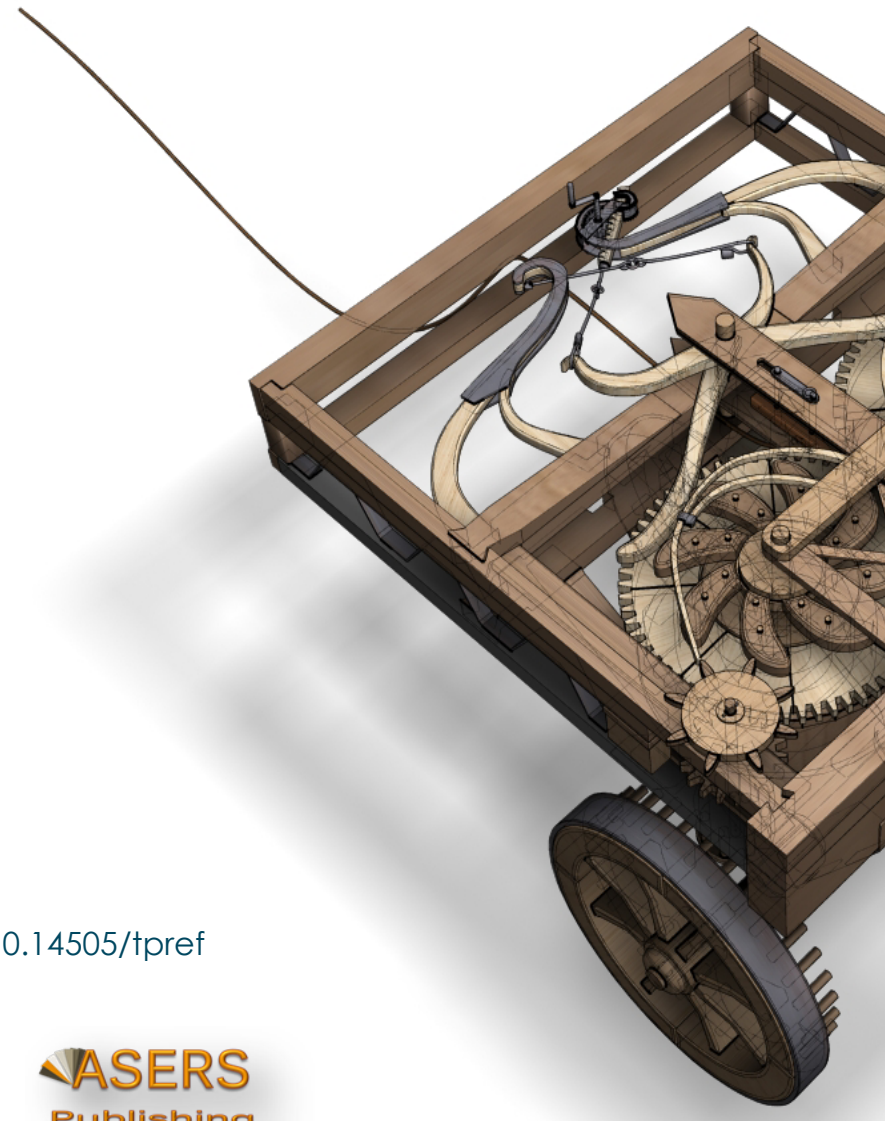
Theoretical and Practical Research in Economic Fields



Quarterly
Volume XV
Issue 3(31)
Fall 2024

ISSN: 2068 – 7710
Journal DOI: https://doi.org/10.14505/tpref

ASERS
Publishing

Volume XV
Issue 3(31)
Fall 2024

ASERS Publishing
ISSN 2068 – 7710
Journal's Issue DOI:
https://doi.org/10.14505/tpref.v15.3(31).00

Table of Contents

Volume XV
Issue 3(31)
Fall 2024

ASERS Publishing
http://www.asers.eu/asers-publishing
ISSN 2068 – 7710
Journal's Issue DOI:
https://doi.org/10.14505/tpref.v15.3(31).00

Call for Papers
Winter Issue
Theoretical and Practical Research in Economic Fields

Many economists today are concerned by the proliferation of journals and the concomitant labyrinth of research to be conquered in order to reach the specific information they require. To combat this tendency, **Theoretical and Practical Research in Economic Fields** has been conceived and designed outside the realm of the traditional economics journal. It consists of concise communications that provide a means of rapid and efficient dissemination of new results, models, and methods in all fields of economic research.

Theoretical and Practical Research in Economic Fields publishes original articles in all branches of economics – theoretical and practical, abstract, and applied, providing wide-ranging coverage across the subject area.

Journal promotes research that aim at the unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems and that are penetrated by constructive and rigorous thinking. It explores a unique range of topics from the frontier of theoretical developments in many new and important areas, to research on current and applied economic problems, to methodologically innovative, theoretical, and applied studies in economics. The interaction between practical work and economic policy is an important feature of the journal.

Theoretical and Practical Research in Economic Fields is indexed in SCOPUS, RePEC, ProQuest, Cabell Directories and CEEOL databases.

The primary aim of the Journal has been and remains the provision of a forum for the dissemination of a variety of international issues, practical research, and other matters of interest to researchers and practitioners in a diversity of subject areas linked to the broad theme of economic sciences.

At the same time, the journal encourages the interdisciplinary approach within the economic sciences, this being a challenge for all researchers.

The advisory board of the journal includes distinguished scholars who have fruitfully straddled disciplinary boundaries in their academic research.

All the papers will be first considered by the Editors for general relevance, originality, and significance. If accepted for review, papers will then be subject to double blind peer review.

Deadline for submission of proposals:	10th November 2024
Expected publication date:	December 2024
Website:	http://journals.aserspublishing.eu/tpref
E-mail:	tpref@aserspublishing.eu

To prepare your paper for submission, please see full author guidelines in the following file: https://journals.aserspublishing.eu/tpref/Template_for_Authors_TPREF_2024.docx on our site.

DOI: https://doi.org/10.14505/tpref.v15.3(31).01

Exploring Profitability in Albanian Banks through Decision Tree Analysis

Olsi XHOXHI
Department of Economics, Faculty of Economy
"Ismail Qemali" University of Vlorë, Albania
ORCID: 0009-0002-1138-1331
olsi.xhoxhi@univlora.edu.al

Zamira SINAJ
Department of Business, Faculty of Economy
"Ismail Qemali" University of Vlorë, Albania
ORCID: 0000-0003-2231-6842
zamira.sinaj@univlora.edu.al

Liridon ISMAILI
Faculty of Natural Sciences and Mathematics
University of Tetova, North Macedonia
liridon.ismaili@unite.edu.mk

Article info: Received 3 May 2024; Received in revised form 28 May 2024; Accepted for publication 23 July 2024; Published 30 September 2024.

Abstract: This study intends to predict Return on Assets (ROA) and assess the significance of several key dependent variables, including Profit per Outlet, Profit per Employee, Natural Logarithm of Assets (ln(assets)), and Loan to Deposits Ratio. In this respect, decision tree regression is employed as the major analytical tool. Within the scope of quarterly data sourced from some leading Albanian banks, such as American Bank of Investments, Banka Kombëtare Tregtare, Credins Bank, Fibank, Intesa Sanpaolo Bank, ProCredit Bank, Raiffeisen Bank, Tirana Bank, and Union Bank, data spanning from January 2016 to December 2022 were taken and used in this analysis. Data collection was done through the website of the Albanian Association of Banks. The following section presents all the relevant information on the selected model, mean squared error (MSE) for training data at 0.5988 and for test data at 0.8912, mean absolute error (MAE) of 0.4764 for the training data and 0.6257 for the test data, and R-squared (R^2) at 0.7706 for training data and 0.4244 for test data. All the findings clearly show that decision tree regression is very effective in the prediction of ROA and also helps in elucidating the relative importance of the selected dependent variables within the context of Albanian banking institutions.

Keywords: Albanian banks; feature importance; machine learning; return on assets.

JEL Classification: C87; E44; G21; G28.

Introduction

The financial health of the banking sector is an important sign of economic progress; it reflects not only the health of financial systems but also more general global economic trends. As concerns on predicting profitability rise for investors, central banks, and other stakeholders over time, since banks come to be regarded as the platform for more capital allocation and growth, predicting bank profitability has become a matter of great concern among all stakeholders. While revenue generation and risk management are central to profitability, the array of macroeconomic conditions, regulatory policies, and innovative technological advancements among others all carry their own subtleties. An important profitability statement is the forecasting of bank profitability that may help in strategic decision-making, investment evaluation, regulatory oversight, and policy formulation. If one examines the banking landscape and even beyond with technological innovations and regulatory changes, forecasting profitability becomes quite important to navigate uncertainties and seize opportunities. The originality of this research lies in the fact that it applied these innovative decision tree methodologies to the Albanian banking sector, which is relatively unexplored in the literature.

This research makes a more detailed contribution to the literature with respect to targeted statistical methods in analyzing bank profitability through the application of decision tree analysis in discovering intricate patterns and relationships that might have been overlooked by conventional methods applied earlier in previous studies. Of importance to this work, therefore, is the fact that if successful, it will help both policymakers and banks to obtain more accurate and actionable insights for making informed decisions aimed at improving profitability and sustainability in one of the world's most dynamic financial environments.

1. Literature Review

Bank profitability has been a subject of extensive scholarly interest, with many researchers focusing on the different aspects of this complicated issue. The present study uses a cross-sectional framework, combining the Operational Model and the MLR techniques, to create a flexible, well-specified regression model able to predict and explain bank profitability at high accuracy. Numerous studies have made efforts towards deriving the determinants of bank profitability, using various methods and frameworks.

Based on the results of Elamir (2020), the investment decision-making issue is solved through prediction models with machine learning methods such as linear regression, regression tree, pruned regression tree, conditional inference tree, and cubist regression. The study was conducted on eight MENA countries for the analysis of sixty-three banks, comprising 630 observations of data. The dataset was divided into training (2008-2017) and testing (2018) sets for model development and validation, respectively. Evaluation metrics used for model assessment are Root Mean Square Error, R-squared, and Mean Absolute Error. It is noted that out of the many models tested, the best one was the cubist regression one and it outperformed all others with a tremendous margin. This cubist regression has attained a 96% R-squared for training data, providing very good predictive performance compared to other models. It is interesting that total assets, bank book value, and total liability have emerged as the major determinants of the earnings per share, therefore pointing the way for bank managers to strengthen financial market stability. The research thus highlights the need for the application of multiple techniques to face multidimensional data, leaving a scope for future research involving further use of financial and non-financial data sources.

Wei *et al.* (2021) did research work on the prediction of various features for the measurement of performance in commercial banks with boosting regression trees. It is different from the previous studies focusing on singular characteristics and involved a more comprehensive view by exploring multidimensional attributes. They introduced an adaptively reduced step size gradient boosting regression tree algorithm well-suited for the bank data and performed better than existing classification methods, like BIRCH. Data from 30 provinces in China was used to conduct a hierarchical cluster analysis, which identified four distinct groups of banks. Particularly, they found that NPLR has a relatively higher coefficient than many other performance evaluation items, indicating that NPLR itself has a better ability of discriminating between good and poor performance.

Atiku and Obagbuwa's (2021) takes an informative look at the performance contribution in the competitive banking industry. They are interested in strategic human resource management by aligning organization’s goals with HR strategies. They employed machine learning algorithms to predict bank performance using data collected from 305 respondents via survey-based quantitative data. The results showed how skills, attitude, and behavior of employees are important for better organizational performance, with accuracies between 74% and 81% when using machine learning models. Employee attitude ranked supreme above other factors. Therefore, the research shows that it is essential to focus on improving employee attitudes if one is to enhance performance within Nigeria’s banking sector.

The Hamzeh F. Assous (2022) study tried to study the ratio in traditional and Islamic banks' efficiency in Saudi Arabia. Using data from 2014 to 2018, from 11 Saudi banks, including 4 Islamic ones, which are annual financial reports, research focuses on the different financial ratios regarding profitability, management practices, assets, loans, and capital adequacy. In the beginning, the relationships between these financial measures and the efficiency of the banks are assessed by means of regression analysis. Then, machine learning prediction models like Support Vector Machine (SVM), CHAID, linear regression, and neural networks are used, where feature selection techniques are applied. The study establishes that profitability, liquidity, and managerial practices significantly affect the efficiency of both types of banks. The CHAID model was hence identified as the best prediction model, and capital adequacy ratios, specifically CAR total and CAR tier 1, emerge as the most important predictors of cost efficiency. The results can be significant for academia, investors, and policymakers because they enable the development of early warning systems to anticipate financial distress scenarios.

Nguyen Minh Tuan's (2022) puts forward that the critical role played by financial deposits is most important in the course of global health crises and economic downturns. Banks still wonder how to differentiate their

customers, and some even go so far as to undergo a significant transformation in order to attract and retain customers while others try to hold onto existing clientele. This calls for a more innovative approach rather than just enthusiasm. In responding to this challenge, it requires more than just the use of basic tools; it needs advanced tools like deep neural network models. By using these approaches, Nguyen will reveal a new methodology through the models like LSTM, GRU, BiLSTM, BiGRU, and SimpleRNN to extend analysis of customer descriptions. Many prior research efforts have indeed concentrated more on the direction of algorithmic models, although Nguyen's research only praises the contribution of deep machine learning. The outcomes present GRU as attaining the highest accuracy of 90.08% at the 50th epoch, closely following BiLSTM with 90.05%. These results are worth for banks to validate the fit of customers towards deposit products within a highly competitive financial sector, with the aim of making wise choices.

According to the study by Hong Hanh Le and Jean-Laurent Viviani (2018), in order to predict the failure of banks, both traditional statistical methods and machine learning techniques were compared. In analyzing the results, they used data of around 3000 US banks for a time frame of five years. Within the provided period, they noted that out of 3000, 1438 banks failed while 1562 survived. Among key aspects such as the quality of loans and profitability, 31 financial ratios were introduced using methods such as discriminant analysis, logistic regression, artificial neural networks, support vector machines, and k-nearest neighbors. The results indicated that artificial neural networks and k-nearest neighbors had the most accuracy in predicting bank failures.

In their work, Appiahene *et al.* (2020) used Data Envelopment Analysis (DEA) with three machine learning models in assessing and forecasting the efficiency of bank branches in Ghana. To this end, they analyzed the efficiency of 444 bank branches at both deposit and investment stages using the criteria for the model application. Efficient or inefficient classification of these banks was based on an 80% cutoff point. Three machine learning algorithms were then used to predict efficiency; 70% of the bank branches were randomly selected for model training and validation. The decision tree (DT) model, specifically using the C5.0 algorithm, performed exceptionally, with the potential for practical application within Ghanaian banking. However, it stresses the need for further improvement in deposit collection and investment efficiency. Appiahene *et al.* suggest in their work that the best way to measure the efficiency of a bank is not simply by using the overall performance metric but by using a comprehensive approach. The study also concludes that among the existing models, decision tree model stands out as the most effective predictive tool. Future studies may need to explore other variables and algorithms to further improve the accuracy in prediction.

Gržeta *et al.* (2023) investigate the effects of the Basel III regulatory framework, which operates at a global level and is incorporated into the law of all member states of the European Union, on bank efficiency and profitability. Their study has particularly tried to answer the following question: How is the performance of banks affected by the introduction of Basel II and Basel III? Bank performance variables that have been checked in this research relating to bank performance are bank size and bank-specific and macroeconomic factors. The two-step empirical approach was based on an analysis of data collected for 433 European commercial banks for the period from 2006 to 2015. In the first step, the relative efficiency is estimated by applying the non-parametric data envelopment analysis. In the second step, it applies the generalized method of moments in order to examine the impact of various bank-specific variables, along with regulatory changes and macroeconomic variables, upon bank performance in terms of both profitability and efficiency measures. The findings indicate that the effect of regulation is very significantly related to bank size. For large and medium banks, the increased effect of new regulation increases efficiency and profitability. Not so with the smallest banks, as added administrative and regulatory challenges bring down their performance. Well, the findings indicate that big banks easily adapted to the changing environment brought about by regulation, while for the smaller ones, such added pressure from the new framework was unbearable. Therefore, it is postulated in this study that a one-size-fits-all approach to regulation is not ideal and thus proposes a differentiated regulatory framework in pursuit of a level playing field and systemic risk mitigation.

Prenaj *et al.* (2024) set up a research study that evaluated the various drivers of commercial banking profitability for Western Balkans countries during the period 2010–2020. They took the return on assets and the return on equity as the primary indicators of bank performance, and these were further evaluated in terms of their determinants: non-performing loans, the number of banks, growth in real GDP, the unemployment rate, and consumer price inflation. Their output showed that real GDP growth, the rate of unemployment, and inflation had positive impacts on the performance of banks, whereas non-performing loans and the number of banks had negative effects. The banking system in Kosovo also had an outstanding ROE performance, almost three times higher than the regional average, with the highest percentage of unemployment. Further, Montenegro was weak

at the beginning but showed year-by-year ROE improvements. The findings of this research study have very important implications for the policy framers and regulators, especially in the Kosovo case.

Abu Khalaf and Awad (2024) assess how the liquidity risk impacts the profitability of banks in the Middle East and North Africa region. Utilizing data compiled using Refinitiv Eikon for the 11-year period from 2012 to 2022, this research paper takes the dependent variable to be Return on Equity and the independent variable to be liquidity risk, with controls for bank size, loan quality, inflation, GDP, income diversification, operational efficiency, capital adequacy, and growth, in the context of 71 banks in MENA. In this case, the investigation is conducted using OLS and panel regression estimates, both fixed and random effects, to assess the impact of liquidity risk on the banks' profitability. As the results demonstrated, the bank size, the operational efficiency, and the extent of non-performing loans all negatively affect the level of profitability, indicating that possibly larger banks have higher operating costs and hence reduced levels of profitability, for the MENA region. On the other hand, liquidity risk, capital adequacy, income diversification, and growth positively influence ROE, thus providing support to the risk-reward theory that banks with better growth opportunities, higher capital adequacy, diversified income sources, and higher liquidity risk experience greater profitability. The GMM is used to confirm the robustness of these results.

The panel data regression model developed for Jigeer and Koroleva (2023) research contains variables that measured and compared various internal and external factors affecting the commercial banks' profitability in China's cities. The study pinpoints 16 listed city commercial banks, and these employ unbalanced data in the period between 2008 and 2020. Use various panel data estimation methodologies like fixed effects, random effects, and comparing them with pooled OLS to find the best-fitted model via statistical hypothesis testing. The results of the study, in this respect, indicated that the said factors- bank size, capital adequacy, credit quality, and operational efficiency- and external factors, in the form of provincial GDP, and inflation, significantly impact profitability, while liquidity does not. The current study contributes to the existing literature on the identification of key determinants of the situation related to Chinese bank profitability in the creation of practical insights for banking management, regulators, as well as municipal and state authorities.

In their study, Doğan and Yildiz (2023) tried to find both internal and external factors which affect banks' profitability in Turkey. The data was derived from 23 public, private, and foreign banks from 2007 to 2020 for the analysis. ROE and ROA indicators are applied for profitability measures, and Dynamic GMM and Fixed Effect Model methods used in trying to enhance model reliability. The results show a positive and significant relationship between inflation and GDP growth rates with both ROA and ROE. In addition, the GMM results depict a positive relationship of ROA and ROE with their respective 1-year and 2-year lags that evidence consistency in profitability in the Turkish banking sector.

O'Connell (2022) examined the impact of bank-specific, industry-specific, and macroeconomic factors on the profitability of the UK banks using an extended analysis period and pioneering econometric methods. This paper emphasizes that good capital ratios along with increased shares of short-term liquid assets determine bank profitability. It is, in addition, revealed that banks with efficient funding strategies and increased labor productivity tend to perform profitably - while operating expenses negatively influence it slightly. They could not establish any economies of scale or did any evidence back structure conduct performance (SCP) hypothesis. This implies that, all matters being constant, concentration does not influence profitability. Macroeconomic variables such as pair, also are essential variables, that add up to bank performance. The study suggests that macro-UK banks thrive well in a high interest environment together with high loan growth; though of late, the UK economy has been shaky because of Brexit and COVID-19, which has seen reduced rates of loan growth. Moreover, the recent growth in the economy of the UK and rising expectations for interest rates might eventually increase the profitability risks of the banks.

Amaral (2024) examined the profitability of Portuguese and Spanish banks for the period 2014 - 2019, which has remained low despite partial recovery in recent years. The research assesses the factors impacting bank profitability by using two econometric models with return on assets and return on equity as dependent variables, and seven independent variables: size, loans, credit risk, and solvency. The findings of the study established that credit risk had the most negative impact on profitability in both banking systems, while solvency positively influenced profitability in the Portuguese banking sector. Moreover, operating efficiency has a negative effect on profitability in Portuguese banks, hence revealing the influence of internal factors on bank performance. Although the study offers valuable insights for bank decision-makers and regulators, it puts forward a self-constraining view with respect to limitations owing to small sample size and the relatively short period of time in the analysis. Further studies should expand samples and periods of analysis and further test additional financial and non-financial indicators that affect bank profitability.

By using a sample of 125 local Italian banks for the period 2006-2018, Cocozza and Curcio (2024) investigated drivers of net interest margin and interest rate risk exposure. They stress two different sub-periods, 2006-2011 and 2012-2018, underlining in particular the impact of unconventional monetary policies by the ECB. In this regard, the study finds that net interest margin increases with the intensity of maturity transformation-particularly from 2012 to 2018-and with exposure to interest rate risk. Although several studies have analyzed the effect of interest rate risk on net interest income, this paper innovates by considering explicitly three sources of this risk: loan and deposit activities, securities portfolios, and derivatives positions. Correspondingly, their estimations indicate that maturity transformation enhances the interest rate risk, in particular during the latter period. More importantly, while ECB funding initially tended to increase interest rate risk exposure between 2006 and 2011, afterwards it decreased the interest rate risk exposure, suggesting that the long-term refinancing operations by the ECB improved funding stability and the capability of banks to manage interest rate shocks, in contrast to deposits held at the ECB.

2. Data and Methodology

This study will utilize quarterly data from Albanian banks including American Bank of Investments, Banka Kombëtare Tregtare, Credins Bank, Fibank, Intesa Sanpaolo Bank, ProCredit Bank, Raiffeisen Bank, Tirana Bank, and Union Bank. The dataset spans from January 2016 to December 2022, comprising a total of 252 data points. The data were obtained from the official website of the Albanian Association of Banks.

The dependent variable being analyzed is Return on Assets (ROA), which is a financial metric used to evaluate a bank's efficiency in generating profits from its assets. The independent variables being considered in the study are:

1. Profit per Outlet: This variable represents the profitability of each branch or outlet operated by the bank. It measures the amount of profit generated by each outlet individually, providing insight into the performance of different operational units.
2. Profit per Employee: This variable denotes the profitability of the bank per employee. It assesses the efficiency of human resource utilization in generating profits and can indicate the productivity of the workforce.
3. Natural Logarithm of Assets (ln(assets)): This variable involves taking the natural logarithm of the total assets of the bank. Transforming the assets in this way can help in dealing with the potential skewness or variability in asset sizes across different banks, making comparisons more meaningful.
4. Loan to Deposits Ratio: This variable represents the ratio of loans held by the bank to its total deposits. It is a measure of the bank's liquidity risk and its reliance on borrowed funds compared to deposits. A higher ratio typically indicates higher risk but also potentially higher returns.

These independent variables are being examined to understand their relationship with the dependent variable, ROA. The study aims to determine how variations in Profit per Outlet, Profit per Employee, ln(assets), and Loan to Deposits Ratio affect the bank's Return on Assets. By analyzing these relationships, we can gain insights into the factors influencing the financial performance and efficiency of the bank.

As a model in this study we will use decision tree regression which is a machine learning algorithm used for regression tasks, where the goal is to predict a continuous target variable. The process of building a decision tree regression model involves recursively partitioning the feature space into regions (represented by the leaf nodes) using decision rules based on feature thresholds. The predicted value at each leaf node is calculated as an aggregation of the target variable values of the data points falling into that region. The algorithm aims to minimize a chosen splitting criterion (MSE) at each step. Below is a breakdown of the components and mathematical formulas involved in decision tree regression:

1. Splitting Criterion: Let Q(X,y) represent the splitting criterion function, where X is the feature matrix and y is the target variable vector. For example, Mean Squared Error (MSE) can be represented as:

$$Q(X,y)=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2$$

where n is the number of data points, y_i is the actual target value, and $\hat{y}_i$ is the predicted target value.

2. Decision Rule: At each node, the decision rule is based on selecting a feature j and a threshold value t that minimizes the splitting criterion:

$$j,t=\arg\min_{j,t}\left[Q(X,y)-\frac{N_{left}}{N}Q\left(X_{left},y_{left}\right)-\frac{N_{right}}{N}Q\left(X_{right},y_{right}\right)\right]$$

where N_{left} and N_{right} are the number of data points sent to the left and right child nodes after the split, respectively.

3. Leaf Node Prediction: The predicted value at each leaf node is the mean (or median) of the target variable values of the data points assigned to that leaf node.

$$\hat{y}_{leaf} = \frac{1}{N_{leaf}} \sum_{i \in leaf} y_i$$

where N_{leaf} is the number of data points in the leaf node.

4. Model Prediction: To make a prediction for a new data point x, the algorithm traverses the tree from the root node down to a leaf node based on the feature values of x. The predicted value for regression is then the value associated with the leaf node reached.

3. Research Results

In this section, we will explore the descriptive statistics of the variables: Loan/deposits (%), ln(assets), Profit per Employee, and Profit per Outlet. This analysis will include measures such as the mean, standard deviation, minimum, 25th percentile, median (50th percentile), 75th percentile, and maximum values, which will be presented in Table 1 to provide a comprehensive understanding of the data distribution and central tendencies.

Table 1. Descriptive Statistics

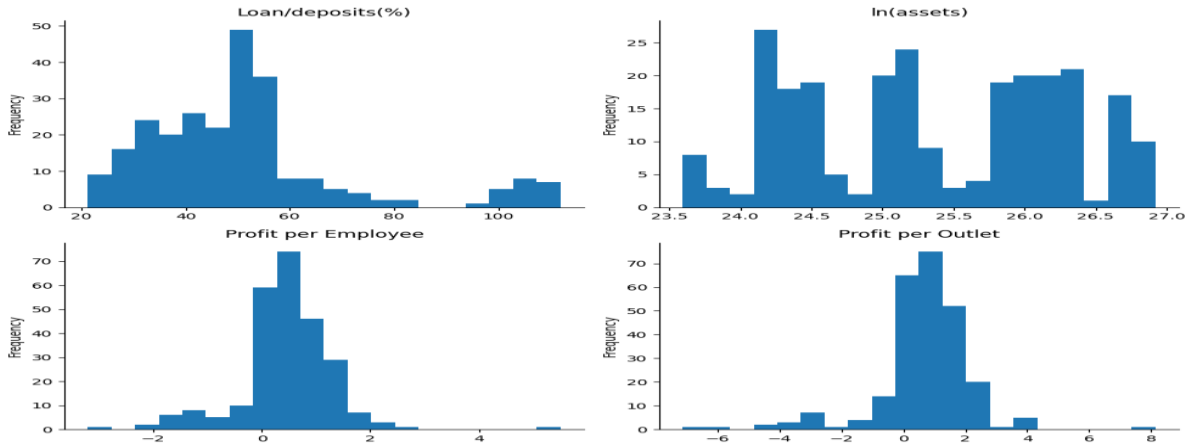
index	Loan/deposits(%)	ln(assets)	Profit per Employee	Profit per Outlet
mean	51,26	25,34	458928,45	6667054,21
std	20,12	0,92	862348,66	15069761,89
min	21,26	23,59	-3206685,08	-71188408,80
25%	38,55	24,45	108718,03	1681646,55
50%	49,29	25,24	458449,99	7252827,55
75%	55,73	26,15	880625,89	13916463,57
max	111,90	26,92	5502258,34	81311150,97

Source: Author's calculations

The findings suggest that, on average, the loan-to-deposits ratio stands at 51.26%, with a standard deviation of 20.12%. The natural logarithm of assets averages at 25.34, with a narrow standard deviation of 0.92. Profit per employee is reported at an average of $458,928.45, but with a considerable standard deviation of $862,348.66, indicating significant variability across institutions. Profit per outlet follows a similar pattern, with an average of $6,667,054.21 and a substantial standard deviation of $15,069,761.89. The distribution of data shows wide ranges, from negative profits to high positive values, indicating diverse performance levels within the sample. The median values provide a more robust measure of central tendency, showing a loan-to-deposits ratio of 49.29%, ln(assets) of 25.24, profit per employee of $458,449.99, and profit per outlet of $7,252,827.55. These statistics illustrate the dispersion and central tendencies of key financial indicators in the dataset, offering insights into the financial performance and operational efficiency of the institutions surveyed.

The histograms depicted in Figure 1 illustrate the distribution of independent variables within the dataset. Each histogram encapsulates the frequency distribution of a distinct variable, showcasing the range and spread of values it encompasses. These visual representations provide valuable insights into the underlying data patterns and characteristics, aiding in understanding the variability and tendencies within the variables.

Figure 1. Histograms of independent variables



Source: Author's calculations

In Table 2, the decision tree regression model demonstrates moderate predictive performance, as indicated by the evaluation metrics. The Mean Squared Error (MSE) measures the average squared difference between the actual and predicted values, showcasing a value of 0.5988 for the training data and 0.8912 for the test data. Additionally, the Mean Absolute Error (MAE) presents a similar trend, with values of 0.4764 and 0.6257 for training and test data, respectively. These metrics provide insight into the magnitude of errors in the model's predictions, with lower values indicating better accuracy. The R-squared (R^2) coefficient, representing the proportion of the variance in the dependent variable that is predictable from the independent variables, suggests a reasonably good fit for the training data (0.7706), while the test data shows a lower but still significant R^2 value of 0.4244. This indicates that the model explains approximately 77.06% of the variance in the training data and 42.44% in the test data, respectively, implying that a substantial portion of the variance is captured by the model. Overall, while the model performs relatively well in explaining the variability in the data, caution should be exercised to ensure it generalizes effectively to unseen data.

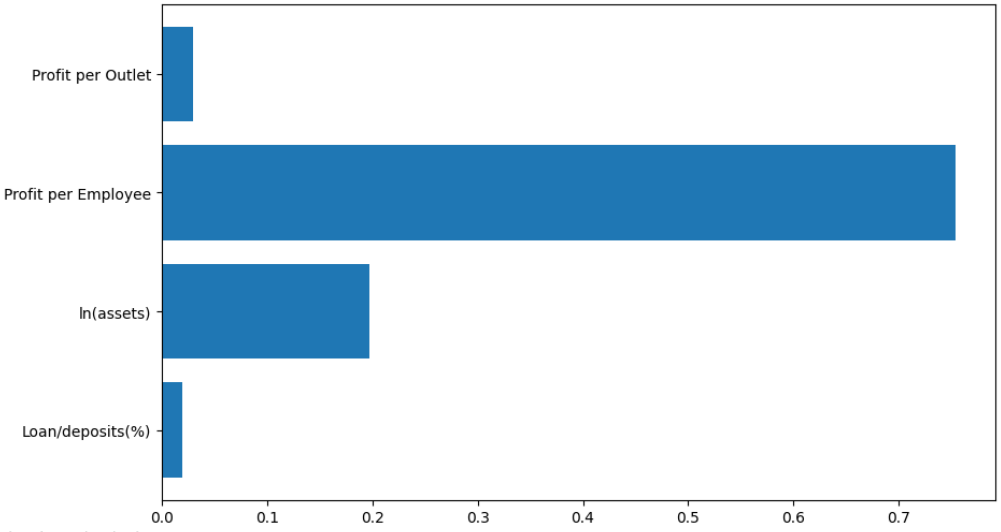
Table 2. Evaluation Metrics

Mean Squared Error (MSE) for Training Data	0.5988
Mean Squared Error (MSE) for Test Data	0.8912
Mean Absolute Error (MAE) for Training Data	0.4764
Mean Absolute Error (MAE) for Test Data	0.6257
R-squared (R^2) for Training Data	0.7706
R-squared (R^2) for Test Data	0.4244

Source: Author's calculations

In Figure 2, we present the feature importance analysis of the model, illustrating the significance of various input features.

Figure 2. Feature importance

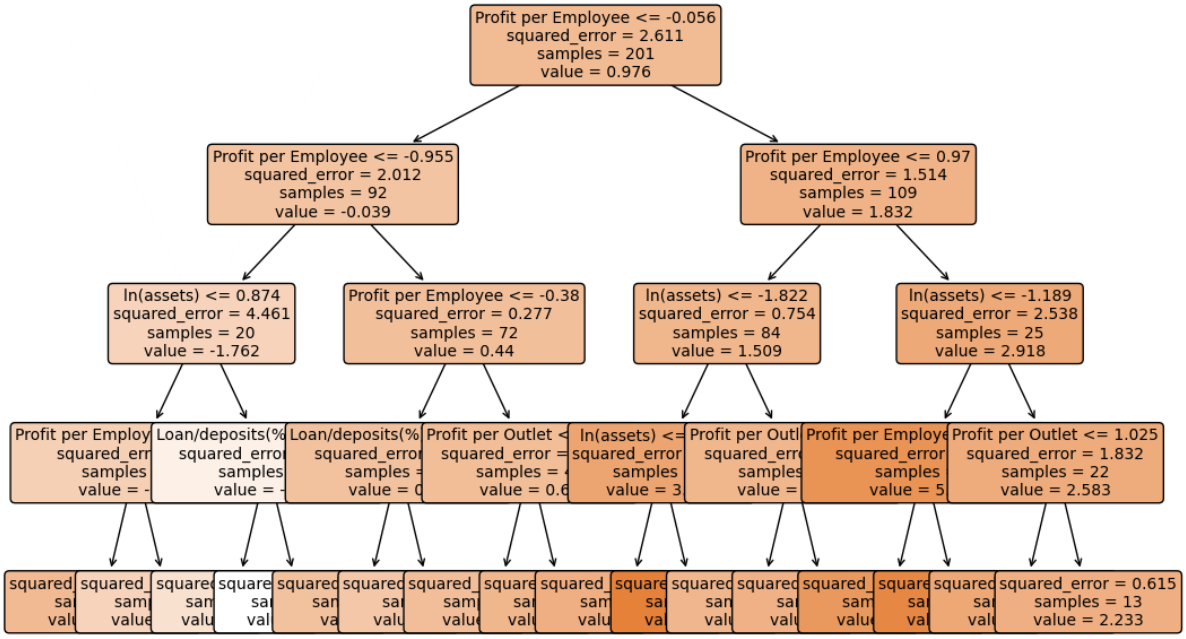


Source: Author's calculations

In analyzing the datas, four key features were assessed, each assigned an importance score indicating its significance in the evaluation. The first feature, measuring the ratio of loans to deposits, held a minimal importance score of 0.0196, suggesting it had a relatively minor impact on the overall assessment. Conversely, the second feature, the natural logarithm of assets, possessed a significantly higher importance score of 0.1968, indicating its greater relevance in the analysis. The third feature, profit per employee, stood out with a substantial importance score of 0.7637, implying its significant influence on the evaluation metrics. Lastly, the fourth feature, profit per outlet, had a relatively low importance score of 0.0199, suggesting it contributed minimally to the overall assessment compared to the other features. These importance scores provide insights into the relative significance of each feature in the financial analysis.

Meanwhile, Figure 3 showcases the decision tree visualization, offering insights into the decision-making process of the model.

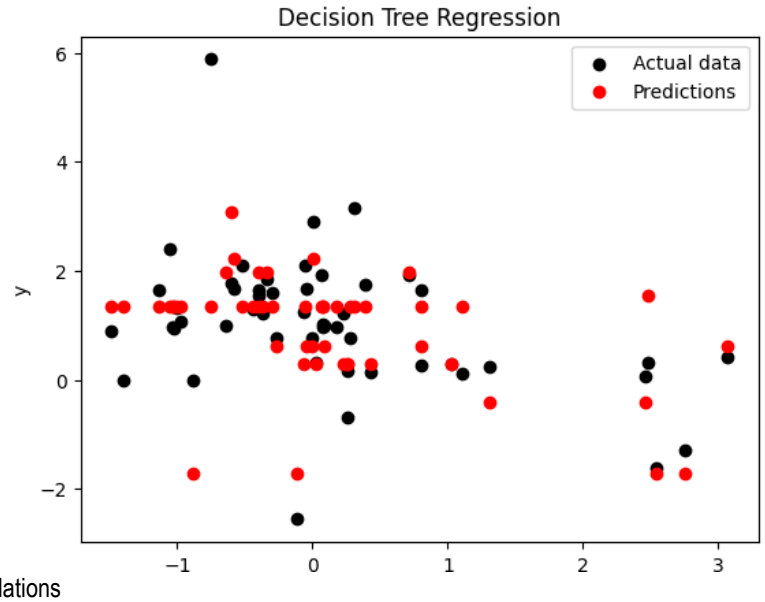
Figure 3. Decision tree visualization



Source: Author's calculations

In the figure provided below, we present the predictions made for the test data set. Through rigorous analysis and the utilization of advanced predictive models, we have endeavored to anticipate the outcomes of the test data with precision and accuracy. Each data point in the figure corresponds to a prediction generated by our model, based on the patterns and trends identified in the training data. Our objective is to assess the performance of the model in extrapolating from known data to predict the behavior of unseen data points. By comparing these predictions to the actual outcomes of the test data, we aim to evaluate the efficacy and reliability of our predictive methodologies. This figure serves as a visual representation of our predictive insights and offers valuable insights into the performance of our model in real-world scenarios.

Figure 4. Decision tree predictions



Source: Author's calculations

Conclusion

In conclusion, this study utilizes decision tree regression to predict Return on Assets (ROA) and assesses the significance of key dependent variables within the Albanian banking sector. Utilizing quarterly data from January 2016 to December 2022 sourced from prominent Albanian banks, the analysis encompasses a total of 252 data

points. The results demonstrate promising predictive accuracy, as evidenced by Mean Squared Error (MSE) and Mean Absolute Error (MAE) values, along with R-squared (R^2) coefficients. The R-squared (R^2) for the training data is 0.7706, indicating a strong fit of the model to the training data. For the test data, the R-squared (R^2) is 0.4244, suggesting that the model performs reasonably well on unseen data, though with a lower level of explanatory power compared to the training data.

Notably, decision tree regression proves effective in elucidating the relative importance of variables such as Profit per Outlet, Profit per Employee, Natural Logarithm of Assets (ln(assets)), and Loan to Deposits Ratio in determining ROA. In analyzing the datas, four key features were assessed, each assigned an importance score indicating its significance in the evaluation. The first feature, measuring the ratio of loans to deposits, held a minimal importance score of 0.0196, suggesting it had a relatively minor impact on the overall assessment. Conversely, the second feature, the natural logarithm of assets, possessed a significantly higher importance score of 0.1968, indicating its greater relevance in the analysis. The third feature, profit per employee, stood out with a substantial importance score of 0.7637, implying its significant influence on the evaluation metrics. Lastly, the fourth feature, profit per outlet, had a relatively low importance score of 0.0199, suggesting it contributed minimally to the overall assessment compared to the other features.

These findings are of paramount importance to bank shareholders, providing insights that can inform strategic decision-making and optimize returns on investment. Moreover, this study suggests that future research should consider incorporating additional dependent variables to further enhance the predictive power and explanatory capability of models in assessing banking performance. Such endeavors hold the potential to deepen our understanding of the multifaceted factors influencing ROA and contribute to more comprehensive risk management and value creation strategies within the banking industry.

Credit Authorship Contribution Statement

Authors have contributed equally to this research.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of Use of Generative AI and AI-Assisted Technologies

The authors declare that they have not used generative AI and AI-assisted technologies during the preparation of this work.

References

[1] A. H. Elamir, Elsayed (2020). Modeling and Predicting Earnings per Share via Regression Tree Approaches in Banking Sector: Middle East and North African Countries Case. *Investment Management and Financial Innovations*, 17 (2): 51–68. DOI: https://doi.org/10.21511/imfi.17(2).2020.05

[2] Abu Khalaf, Bashar and Antoine B. Awad (2024). Exploring the Bearing of Liquidity Risk in the Middle East and North Africa (MENA) Banks. *Cogent Economics and Finance*, 12 (1). DOI:https://doi.org/10.1080/23322039.2024.2330840.

[3] Amaral, Marco (2024). Profitability of Commercial Banks in Portugal and Spain: A Panel Data Analysis Model." *Revista de Métodos Cuantitativos Para La Economía Y La Empresa*, (March) 1–18. DOI:https://doi.org/10.46661/rev.metodoscuant.econ.empresa.8172

[4] Appiahene, Peter, Yaw Marfo Missah, and Ussiph Najim (2020). Predicting Bank Operational Efficiency Using Machine Learning Algorithm: Comparative Study of Decision Tree, Random Forest, and Neural Networks. *Advances in Fuzzy Systems,* 2020 (July): 1–12. DOI: https://doi.org/10.1155/2020/8581202

[5] Assous, Hamzeh F. (2022). Prediction of Banks Efficiency Using Feature Selection Method: Comparison between Selected Machine Learning Models. *Complexity* 2022 (April): 1–15. DOI:https://doi.org/10.1155/2022/3374489

[6] Atiku, Sulaiman O., and Ibidun C. Obagbuwa (2021). Machine Learning Classification Techniques for Detecting the Impact of Human Resources Outcomes on Commercial Banks Performance. *Applied Computational Intelligence and Soft Computing* 2021 (September): 1–16. DOI:https://doi.org/10.1155/2021/7747907

[7] Cocozza, Rosa, and Domenico, Curcio (2024). Interest rates, profitability and risk: Evidence from local Italian banks over the years 2006-2018. *Risk Management Magazine* 19 (2): 23-42. DOI:https://dx.doi.org/10.47473/2016rrm

[8] Doğan, Mesut, and Feyyaz Yildiz (2023). Testing the Factors That Determine the Profitability of Banks with a Dynamic Approach: Evidence from Turkey. *Journal of Central Banking Theory and Practice*, 12 (1): 225–48. DOI: https://doi.org/10.2478/jcbtp-2023-0010

[9] Gržeta, Ivan, Saša Žiković, and Ivana Tomas, Žiković (2023). Size Matters: Analyzing Bank Profitability and Efficiency under the Basel III Framework. *Financial Innovation,* 9 (1). DOI: https://doi.org/10.1186/s40854-022-00412-y

[10] Jigeer, Shawuya, and Ekaterina, Koroleva (2023). The Determinants of Profitability in the City Commercial Banks: Case of China. *Risks*, 11 (3): 53. DOI: https://doi.org/10.3390/risks11030053

[11] Le, Hong Hanh, and Jean-Laurent, Viviani (2018). Predicting Bank Failure: An Improvement by Implementing a Machine-Learning Approach to Classical Financial Ratios. *Research in International Business and Finance*, 44 (April): 16–25. DOI: https://doi.org/10.1016/j.ribaf.2017.07.104

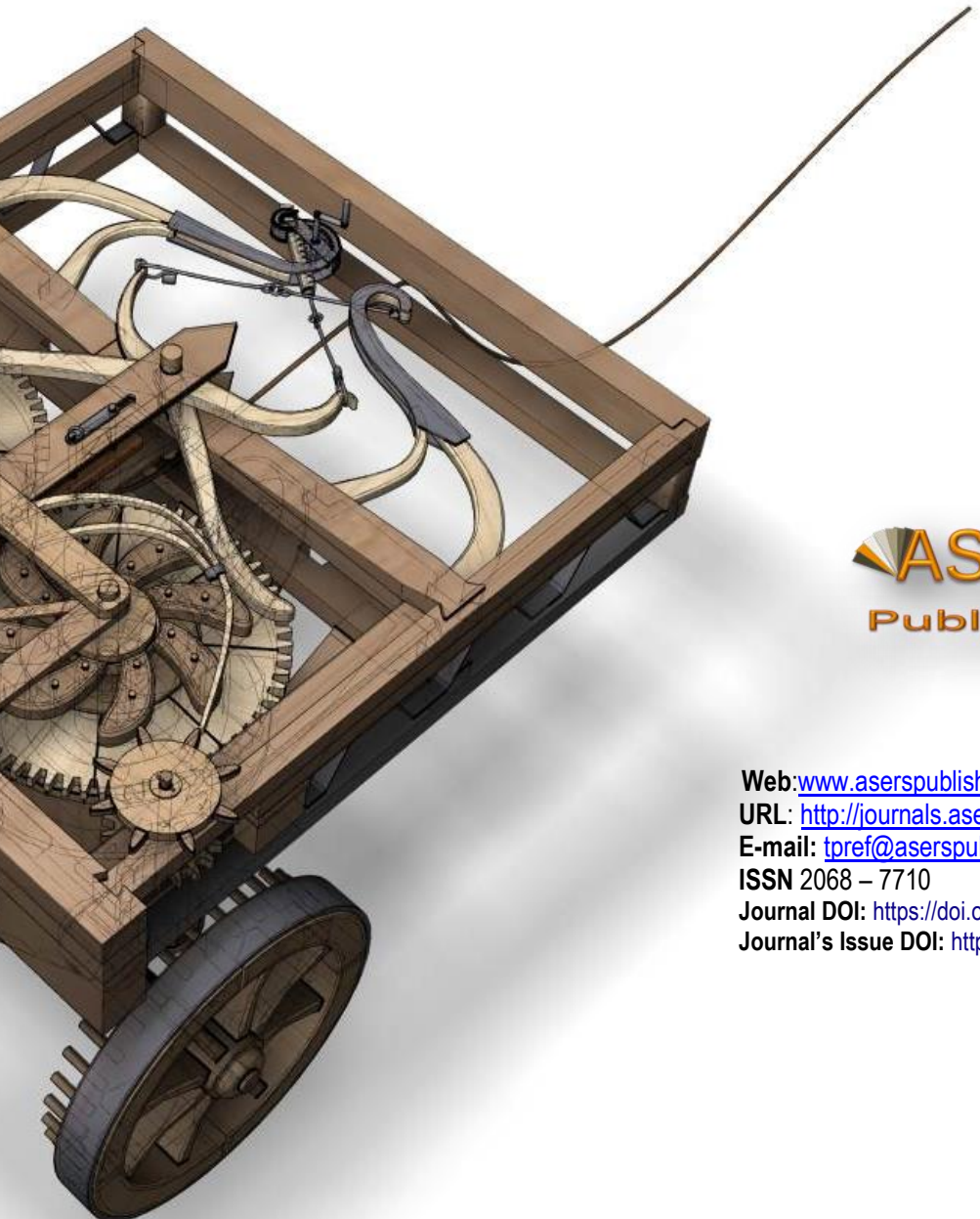
[12] Nguyen, Tran Tuan (2022). Machine Learning Performance on Predicting Banking Term Deposit. In Proceedings of the 24th International Conference on Enterprise Information Systems: 267-272. DOI:https://doi.org/10.5220/0011096600003179

[13] O'Connell, Michael (2022). Bank-Specific, Industry-Specific and Macroeconomic Determinants of Bank Profitability: Evidence from the UK. *Studies in Economics and Finance*, 40 (1). DOI:https://doi.org/10.1108/sef-10-2021-0413

[14] Prenaj, Vlora, Iliriana Miftari and Sokol, Berisha (2024). The Determinants of Profitability of Western Balkan Countries Commercial Banks: A Panel Data Approach. *Economic Alternatives,* 30 (2): 243–58. DOI:https://doi.org/10.37075/ea.2024.2.02

[15] Wei, Jun, Tao Ye, and Zhe Zhang (2021). A Machine Learning Approach to Evaluate the Performance of Rural Bank. *Complexity* 2021 (January): 1–10. DOI: https://doi.org/10.1155/2021/6649605

ASERS



Web:www.aserspublishing.eu
URL: http://journals.aserspublishing.eu/tpref
E-mail: tpref@aserspublishing.eu
ISSN 2068 – 7710
Journal DOI: https://doi.org/10.14505/tpref
Journal's Issue DOI: https://doi.org/10.14505/tpref.v15.3(31).00