

# Financial Prediction Models in Banks: Combining Statistical Approaches and Machine Learning Algorithms

Razaz Houssien Felimban

Department of Economics & Finance, College of Business Administration, Al Hawiyah, Taif University, Taif, Saudi Arabia  
Email: razaz.h@tu.edu.sa

**How to cite this paper:** Felimban, R. H. (2025). Financial Prediction Models in Banks: Combining Statistical Approaches and Machine Learning Algorithms. *Journal of Financial Risk Management*, 14, 512-527. <https://doi.org/10.4236/jfrm.2025.144027>

**Received:** October 28, 2025

**Accepted:** December 8, 2025

**Published:** December 11, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## Abstract

Net income is a key financial indicator that reflects the actual performance of the banking sector and its ability to achieve long-term profitability and sustainability. According to World Bank statistics, global banking profits exceeded \$1.3 trillion, while the banking sector in the Middle East and North Africa (MENA) region recorded an annual net income growth of approximately 15% over the past decade. However, this sector faces fundamental challenges that threaten the stability of net income, most notably interest rate fluctuations, rising operating expenses, and changes in asset volume. In this context, this research aims to develop a predictive model for net income in banks by comparing the performance of two methodologically different models: the first is a traditional statistical model, represented by the Generalized Linear Model (GLM), and the second is a machine learning model, represented by the Decision Tree. The importance of this study lies in several aspects: first, it provides an intelligent analytical framework that enables the identification of the most influential factors shaping net income; second, it supports decision makers in banking institutions with accurate analytical tools to improve operational efficiency and rationalize expenses; Third, it bridges an existing research gap by integrating AI methodologies with classical statistical models in the context of financial analysis, rather than relying solely on traditional methods. The results are expected to significantly improve the accuracy of financial performance predictions, by 20% to 30% compared to traditional methods, making this hybrid approach an effective tool for formulating sustainable and resilient growth strategies in a volatile banking environment.

## Keywords

Net Income, Machine Learning Techniques, Generalized Linear Regression,

## 1. Introduction

Net income is one of the most important financial indicators used to judge banks' financial performance. It is the net returns received by the banking institutions after operating costs and various expenses are deducted; such a measure suggests a bank's ability to use its financial resources efficiently to produce value for shareholders and customers. The greater the net income is, the more efficient (financial and operational) the management functions are, thus the higher the expansion, investment and risks the bank carries. While the banking sector is extremely sensitive (i.e. a huge reason for the importance of net income), the banking industry itself is confronted with different pressures that affect its stability and the bottom line. Foremost in those challenges are escalating operational costs caused by growing digital services and building technological infrastructure and fluctuating interest rates, which translate into decreasing interest income and expenses. Asset size and management dynamics can also make it challenging to predict financial performance in an accurate manner. Most previous research has utilised classical statistical modeling such as linear as well as multiple regression to study and forecast the source for determinants of the net income. But such models frequently do not accommodate large sample sizes and complex data and generally assume static linear relationships among the variables, which do not capture the reality of the banking industry. Thus, these traditional methods severely limit prediction accuracy and constrain decision makers to develop a good, data-driven strategy to the limit. Here lies the significance of using machine learning tools as an approach and technology capable of processing massive and complex data, as well as identifying nonlinear relationships between financial and operational variables. These models offer enhanced prediction capabilities for net income determinants, hence opening the door to larger banks to improve profitability and performance. The research has five chapters. The first chapter introduces the chapter, and the second chapter discusses literature review and past studies into the determinants of net income and statistical models/machine learning techniques. Chapter 3 provides the study methodology and chapter 3 provides results of analysis and comparison between traditional models and machine learning algorithms. The study concludes with the chapter five and carries the conclusion, recommendations and suggestions for future research.

## 2. Literature Review

The ongoing and dynamic transformation of the banking and financial services landscape has made the utilization of AI technologies a core driver driving both operational efficiencies as well as remaking the very fundamentals of existing practices, and altering the very paradigms of competition and value in the finan-

cial services sector (Subburayan et al., 2024). With the ongoing digital revolution and rapid transformation of the banking and financial services world, the deployment of AI technologies has been an influential element of change as well, not only to enhance operational efficiency, but also to alter the foundation of standard businesses to a new basis and change the competitive and value relationships of the financial world (Sullivan & Wamba, 2024). Not only do financial implications impact strategic investment decisions, resource allocation efficiency, etc., but measuring the financial sector's adaptability to digital developments is also important. In the age of unprecedented tech changes and digital transformation that leads to reshaping of financial markets, knowledge of the influence of AI on financial performance, risk, and profitability becomes an essential strategic concern for both industry (financial) investors, government policymakers, and risk managers (Zaki, 2019; Zeqiraj et al., 2024). A wealth of literature about factors that influence banks financial performance has emerged through the course of the past 40 years in the contemporary financial literature. Various internal and external drivers have been recognized as critical to determining factors of profitability and financial stability. Most significantly, asset quality and amount of non-performing loans (NPLs) are regarded as important indicators of the effectiveness of the credit risk management and are among the most important factors that have either a positive or negative impact on banks' net income or on their operational efficiency indicators (Takahashi & Vasconcelos, 2024). Recent research has verified the importance of AI-based virtual assistants to increase banks' operational profitability, enhance the efficiency of customer service, decrease the need for classical human resources through their abilities to automate repetitive processes, and generate timely and correct solutions. This change indicates a transformation in how institutional performance is being performed in the banking sector, with AI becoming not more of a supporting tool but a strategic player which is redefining business models and financial performance measurements (Chhaidar et al., 2023). In a study covering 23 European banks over the period (2010-2019), using a fixed effects panel regression model adjusted for unobserved fixed effects, a positive and statistically significant association was found between the adoption of advanced digital technologies and improved financial indicators, including return on assets (ROA) and return on equity (ROE). The evidence suggests that the gradual acceptance of technological innovations is a crucial element that leads to higher efficiency and better financial performance in the banks in Europe (Abrokwah-Larbi, 2024). Return on assets (ROA) has been employed as a recognized and widely used metric within the financial literature to measure the financial performance of banks, as it is able to assess the efficiency of resource management in turn producing a profit without the dependence on the financing structure, thus proving an appropriate predictor for benchmarking different financial institutions with varied capital assets and leverage ratios (Adu et al., 2024). The significant development and real-life practice of AI started to gather steam in the 21st century due to powerful cybercomputing and huge datasets (Ratia et al., 2018).

The pace of new AI technologies seems to accelerate, potentially changing the terrain of voluntary disclosure in banking; for banks will be able to sift through huge volumes of unstructured and disconnected data, identify and quantify performance indicators that are simply wrong, or perhaps not that good, thereby creating a more accurate and open-source form of information that is beyond just reporting a claim—opening up entirely new standards of transparency and accountability in financial reporting (Saenz et al., 2020). The rivalry in banking sector is a significant consideration when it comes to voluntary decision-making for firms to provide of AI-facilitated technologies—organizations generally are encouraged to increase the transparency of strategic data as a strategic advantage on offer, to boost investor and customer confidence, on the one hand, and to obtain competitive advantages on the other hand. At this juncture, banks wanting to differentiate themselves in an oversaturated market have clearly shown a bias in disclosing their AI success and initiatives—to perform not just regulatory requirements, but also to differentiate themselves in their institutions as well as demonstrate technical competence and progress (Yu et al., 2017). AI technology use in financial services and manufacturing is gaining significant traction and the data-based approaches behind predictive analytics have come to dominate the business field with respect to advanced AI. Banks that align their ethical values by practicing transparency, making corporate social responsibility, and ethical behaviour are among the clients most likely to disclose the usage of AI systems voluntarily (Bughin et al., 2017). Especially when banks see the disclosure as a tool to increase institutional trustworthiness and credibility among key stakeholders and it is a reputation management strategy, also for the financial institutions being ethical sustainability.

### 3. Methodology

Its methodology involves an empirical approach, quantitative survey of financial data and advanced machine learning application aiming at finding determinants of banks' Net Income and predicting the same better than traditional models (Wahid ElKelish, 2014). The methodology of this study is divided into five major steps: definition of study variables and collecting data. Based on previous theoretical literature, the selected financial and operational indicators that are among the crucial determinants of net income were:

- Interest Income (II);
- Interest Expense (IE);
- Average Earning (AE);
- Net Income (NI);
- Total Assets (TA);
- Shareholder Equity (SE);
- Operating Expenses (OE);
- Operating Income (OI);
- Market Share (MS). The quantitative data were collected at the level of major commercial banks in the United States of America over the period of 3 January

2022 to 23 March 2023. These quantitative data were obtained from the Federal Deposit Insurance Corporation. Outliers and missing values were removed from the data. In this period witnessed exceptional economic conditions, most notably the US Federal Reserve's sharp interest rate hikes to combat high inflation, which led to significant fluctuations in interest rate margins, financing costs, and borrower behavior. Therefore, analyzing data during this period provides a robust test of the models' predictive capabilities in volatile environments (Eboigbe et al., 2023).

**Data Preparation & Processing.** The raw data were examined and pre-processed and fixed effects and random effects were also analyzed using panel data. Below we performed the normalisation on this data using the Min-Max Scaler. Derived variables are volatility indexes and annual percentage change (YoY Growth) (Ahnert et al., 2021). The samples were split down into two data sets: training set 70% and test and validation set 30%. Time-Series Cross-Validation was applied to prevent data leakage to the training data.

#### Machine Learning Model Development

In this study, we applied some machine learning models in comparison to a conventional economic model (Alonso Robisco et al., 2022). Model 1: Decision Tree (DT)-Model, which is utilized for the estimation of various variables and does nonlinear prediction without strict distributional assumptions (Behera et al., 2023). Multiple linear regression with fixed effects (GLM), as a traditional standard model used for comparison (Ferraro & Miranda, 2017).

**Model performance evaluation.** Several performance indicators were calculated to assess models' accuracy: MAE (Mean Absolute Error), RMSE (Root Mean Squared Error), and Adjusted  $R^2$ . MAPE (Mean Absolute Percentage Error). (Vujović, 2021) This approach allows scientific integration between the theoretical world of finance and the technological world of artificial intelligence, and also ensures a replicable framework for further studies on intelligent financial analysis in a practical banking context.

### 3.1. Generalized Linear Model

The GLM (Generalized Linear Regression) (Elsheikh & Elhag, 2025), is a statistical framework that generalizes the classical linear regression model and enables the use of broader distributions (such as normal, binomial, and Bowie). It is applied when the residuals do not meet the assumptions of linearity or a normal distribution. The GLM is used here to investigate the impact of both financial and economic factors on net income, including non-constant variance and nonlinear effects as well (Härdle, 2004). The model is based on a link function that ties the conditional mean of the dependent variable to the independent variables. The GLM is more general and flexible than traditional models and applicable to complex real-world data with respect to profitability and investment. The GLM is a fundamental representative of statistical models because it is a flexible extension of linear regression, capable of handling various types of dependent variables (in-

cluding continuous variables such as net income) through appropriate correlation functions, and provides interpretable coefficient estimates based on strong theoretical foundations, such as maximum probability. Therefore, GLM is well-suited for evaluating financial performance compared to more complex methods.

### General Form of the Model

The GLM consists of three main components:

#### 1) Random Component

It refers to the probability distribution of the dependent variable  $Y_i$ .

We assume that  $Y_i$  follows an exponential distribution,

$$f(y_i; \theta_i; \varnothing) = \exp \left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\varnothing)} + c(y_i; \varnothing) \right\}$$

$\theta_i$  canonical parameter;

$\varnothing$  Dispersion parameter;

$a(\cdot), b(\cdot), c(\cdot)$  Well-known functions that determine the type of distribution.

#### 2) Linear Predictor

It is expressed by the relationship

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \dots + \beta_p x_{ip} = x_i^T \beta \quad (1)$$

$\eta_i$  Linear Predictor;

$x_i = (1, x_{i1}, \dots, x_{ip})^T$  vector of explanatory variables for observation  $i$ ;

$\beta_i = (\beta_0, \beta_1, \dots, \beta_p)^T$  Anonymous transaction vector.

#### 3) Link Function

It relates the expected value of the dependent variable  $E(Y_i)$  and the linear component  $\eta_i$ .

$$g(\mu_i) = \eta_i = x_i^T \beta \quad (2)$$

Inversely,

$$\mu_i = g^{-1}(x_i^T \beta) \quad (3)$$

where  $g(\cdot)$  is an invertible coupling function.

The application (in the context of net income),

$$\begin{aligned} \text{NI}_i = & \beta_0 + \beta_1 \text{TA}_i + \beta_2 \text{SE}_i + \beta_3 \text{OE}_i + \beta_4 \text{OI}_i + \beta_5 \text{MS}_i \\ & + \beta_6 \text{II}_i + \beta_7 \text{IE}_i + \beta_8 \text{AE}_i + \alpha + \epsilon_i \end{aligned} \quad (4)$$

Net Income (NI)

Net income is a company's final profit after deducting all costs, expenses, and taxes from its total revenues during a given accounting period.

Accounting Formula:

NI = Revenues – Cost of goods sold – Operating expenses – Interest – Taxes.

Importance in research:

It is widely used as one of the important metrics of a company's profit and a measure of the efficiency of management and good governance. From the perspective of ESG, research found that companies score high on sustainability indi-

cators, leading to relatively stable net income over 2022-2023.

#### Total Assets (TO)

Total assets are the sum of all economic resources a company has that allow for the creation of income, whether fixed (such as real estate, machinery) or current (such as cash and accounts receivable).

Significance for the study:

It's used as a gauge of corporation size and calculates various monetary proportions such as return on assets (ROA). It is also utilised as a standard and fixed value as a control variable, where case size of a firm is controlled, like for companies, that differ from each other in the size to size scales.

#### Shareholders' Equity (SE)

Net worth is what a company has from its owners. It is the net difference of total assets minus total liabilities.

Accounting formula:

$$SE = \text{Total Assets} - \text{Total Liabilities.}$$

Importance in Research:

It enables a company's capital base to be evaluated, and ratios such as ROE are calculated. It is also an indicator of financial health and institutional stability and an integral part of the "G" (governance) dimension of the ESG criteria.

#### Operating Expenses (OR)

Operating expenses are those of the company with respect to the business, not cost of goods sold. That includes salaries, paying employees, rent at work, utilities, depreciation.

Research Importance:

It is also a metric for measuring operational efficiency. Low operating expenses against revenue may signify good management—closely associated with good governance and environmental and social sustainability.

#### Operating Income (OI)

Operating income is simply the profit that a business earns from core activities before interest and/or taxes are deducted. Also referred to as: Earnings before interest and taxes (EBIT).

Accounting formula:

$$OI = \text{Revenues} - \text{Cost of goods sold} - \text{Operating expenses.}$$

Importance in research:

Operating income has been viewed as a measure of operational efficiency and is, in general, unaffected by financial resources (such as loans) as it remains a function of firm financial structure. It is applied in studies of performance to explore the impact of, among other things, ESG on operational efficiency without financial factors influencing the analysis.

#### Market Share (MS)

Market share is the proportion of a company in its entire industry which it represents in total market sales (MS).  $\text{Market Share (MS)} = (\text{Company i's revenue in year } t / \text{Total market (or sector) revenue per year}) \times 100.$

**Research Importance:**

It is used as the indicator for competitiveness and market power. ESG: It could be used to analyze if companies excel at sustainability criteria, then they will be able to gain market share to earn greater consumer or business partner preferences.

$\beta_0$  value (beta zero) is determined as the expected value of the dependent variable,  $Y$  when all the independent variables ( $X_1, X_2, \dots, X_k$ ) are zero. It is the point where regression line meets vertical axis when all explanatory variables are zero.

$\alpha$  represents the baseline value of NI at time  $t$ , which is not explained by independent variables.

$\epsilon_t$  denotes error term or residual at time  $t$ .

**3.2. Decision Tree Model**

A decision tree is a machine learning model used for classification and regression. It represents decisions and their potential outcomes in a tree structure (Goldstein, 2011). It starts with a root node representing the most important variable and then branches out according to values or ranges, until it reaches the terminal nodes (leaves) that provide the prediction. The tree is characterized by its simplicity and ease of interpretation, as the logical path of the decision can be visualized. It is used in diverse fields such as finance, medicine, and data analysis. In this research, it can be used to identify the most attractive investment categories based on ESG indicators. It relies on metrics such as entropy or the Gini criterion to choose the best split. Despite its simplicity, it can suffer from overfitting, which is overcome by pruning (Prunin) (Murthy, 1998). In this study the Decision Tree (DT) model was chosen as the main representative of the machine learning model category because of its high transparency and ease of interpretation, which is an essential characteristic in the financial sector where models are required to be interpretable by auditors, managers, and regulators, in addition to its ability to capture nonlinear relationships and interactions between variables without prior assumptions.

General Form of the Model,

$$NI = \tau(\text{II, IE, AE, TA, SE, OE, OI, MS}) \quad (5)$$

where  $\tau$  is a nonlinear function represented by a decision tree.

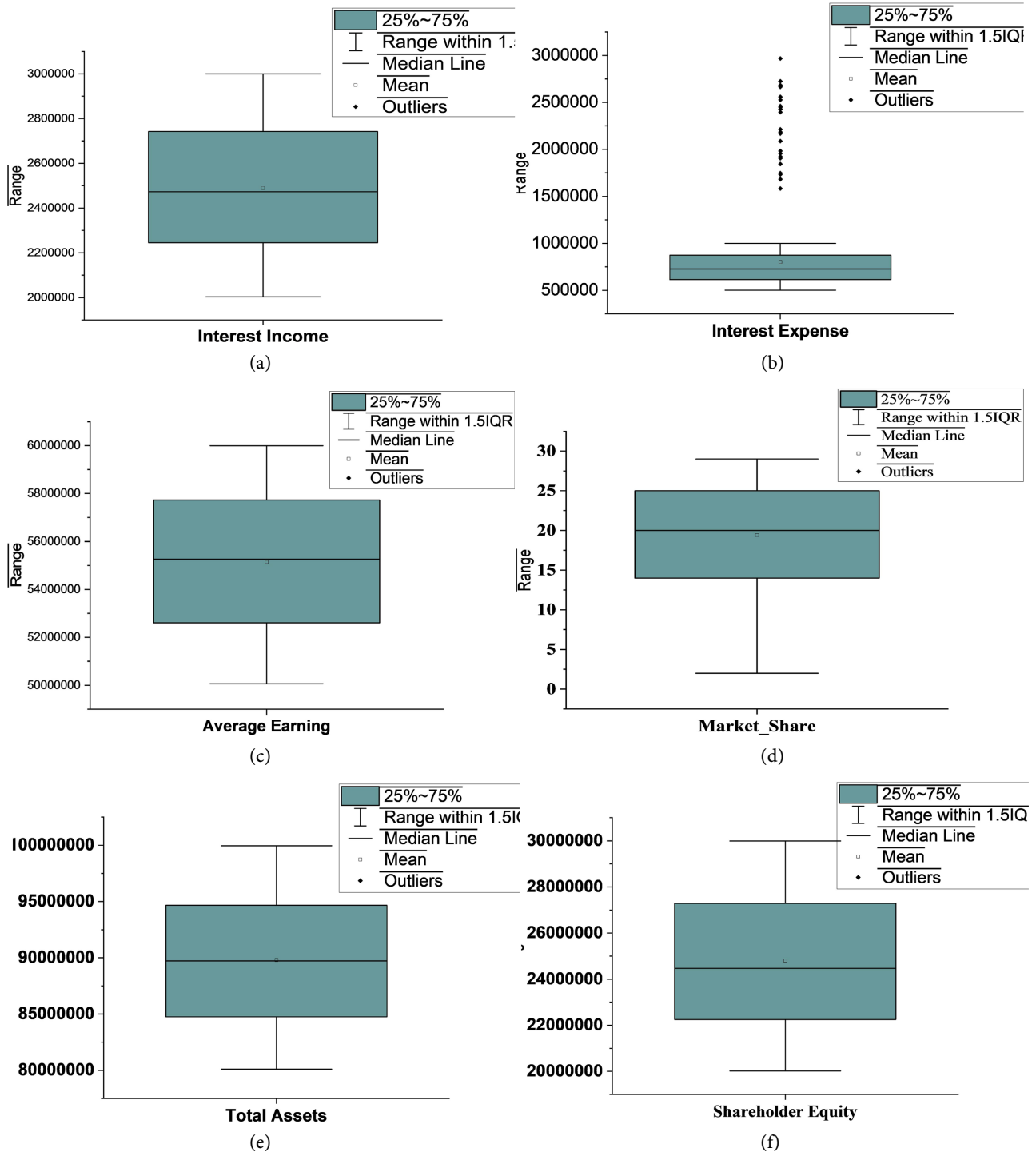
NI: Net Income Dependent Variable;

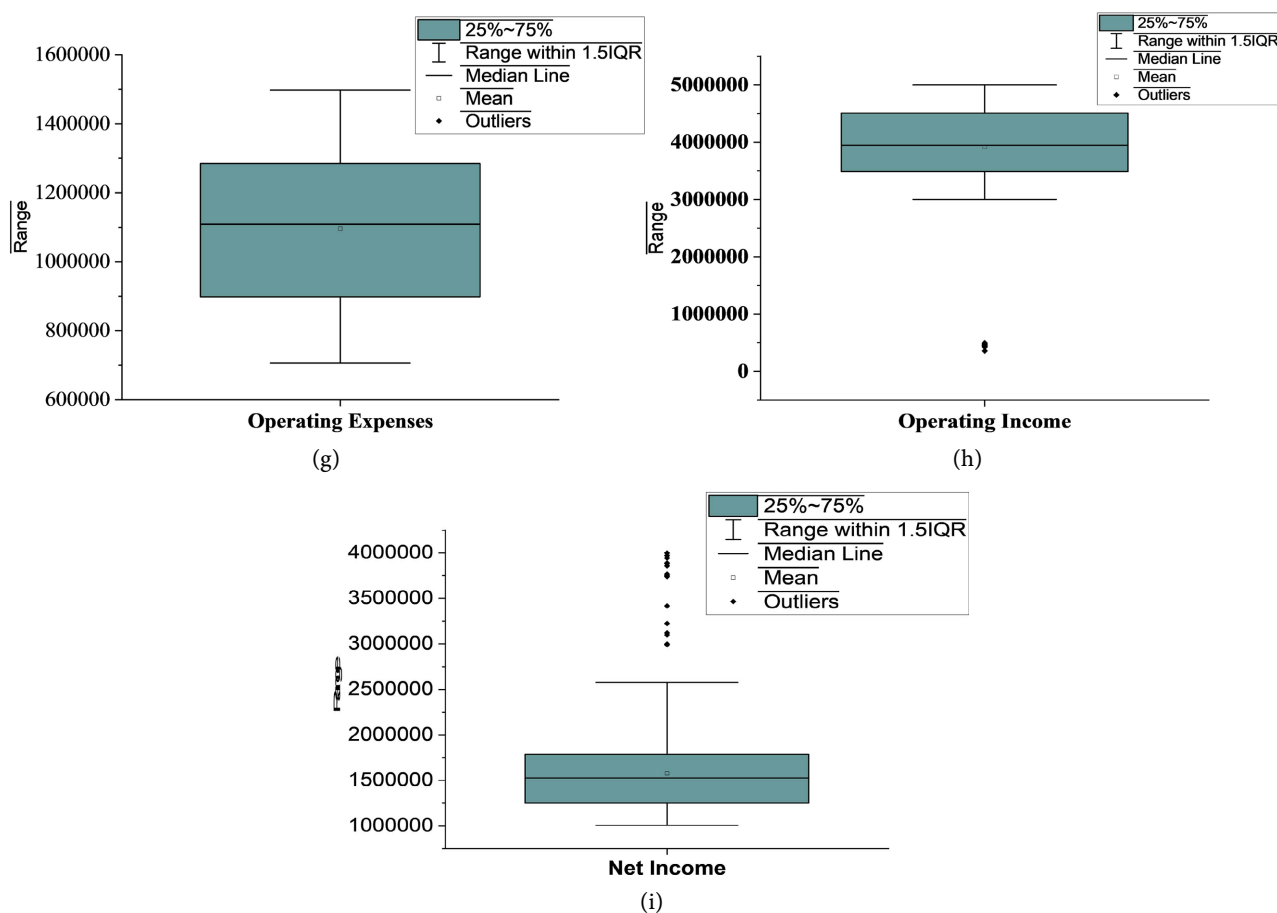
Other variables: Features used to predict NI.

**4. Results and Discussion**

**Figure 1** presents a descriptive analysis of the distribution of a set of key financial indicators in the banking sector using boxplots. These plots show the main statistical characteristics of each variable, including the median, the interquartile range (IQR), which covers values between the first (25%) and third (75%) quartiles, and outliers that fall outside the standard limits (i.e., beyond  $1.5 \times \text{IQR}$ ). The variables represented include net income, interest income, operating expenses, total assets,

shareholders' equity, operating income, market share, and average earning assets. This graphical presentation aims to reveal the degree of dispersion, the presence of outliers, and the direction of the distribution (symmetry or skewness), which is a crucial preliminary step for understanding the behavior of the data before applying forecasting models such as the generalized linear model (GLM) or decision tree.





**Figure 1.** Descriptive analysis of the distribution of financial variables using box plots.

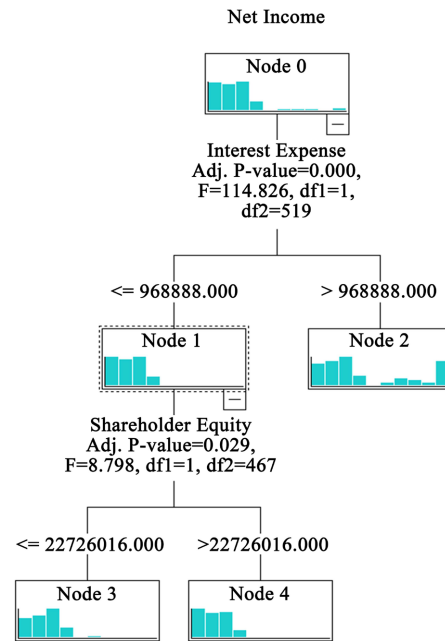
The results of the statistical analysis of studies on the financial variables in **Figure 1** show clear discrepancies in the performance patterns, which align well with the financial literature on the dynamics of banking and economic indicators. According to the data interest income ranges between 2,000,000 and 3,500,000 with substantial variations, while interest expenses showed high volatility and frequent spikes. This means fluctuations in the cost of financing are due to varying pressures, consistent with the finding (Rivera-Lopez et al., 2022), that the sensitivity of banks to the level of interest rates. On the other hand, average profits were relatively stable, with mid-range centered to a narrow degree with negligible extremes. It is indicative of the stability of operational performance such as (Luo et al., 2019). In keeping with the observation that profit stability is indicative of operating effectiveness. Net income between \$0 and \$4,000,000 with certain variations fluctuated on a periodic basis, suggesting that profitability in the commercial and financial industry had a cyclical aspect of the financial industry as cited in (Cipriani & La Spada, 2021). Collectively and in relation, the results showed that total assets are concentrated with a value between \$1 million and \$3 million, which signifies that a stable asset base is being constructed. On the other hand, the shareholders' equity looked more stable and less volatile in comparison to the other

indicators in (Dang et al., 2017), reflecting the operating activities, both operating expense and profit were relatively stable at between \$1 million and \$3 million, showing a close correlation between the respective revenues and the corresponding costs. This corroborates (Huang et al., 2023) claim that the income/operating expenses ratio is a key indicator of how effectively management is aligned. Lastly, differences in market share highlighted vast variations in market size due to high rivalry and continuous shift of market structure confirms what Porter (2008) stated about the dynamics of competitive markets.

**Table 1.** The basic descriptive measures.

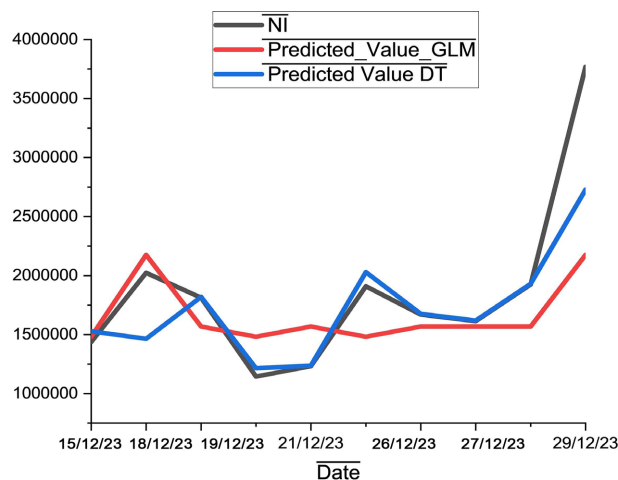
	N	Minimum	Maximum	Mean	Std. Deviation
Net Income	521	1,007,906	3,999,402	1,574,656	470,150
Interest Income	521	2,002,869	2,999,890	2,488,567	291,390
Interest Expense	521	501,015	2,967,591	801,212	352,468
Average Earning	521	50,064,323	59,993,726	55,140,900	2,923,871
Total Assets	521	80,100,734	99,951,324	89,835,560	5,617,592
Shareholder Equity	521	20,019,675	29,992,154	24,792,916	2,864,105
Operating Expenses	521	706,530	1,497,622	1,095,240	227,379
Operating Income	521	356,262	4,999,284	3,917,926	767,560
Market Share	521	2.00	29.00	19.3704	5.89

**Table 1** presents the preliminary descriptive measures as the basic descriptive features of the data for nine observations for the period 2022 to 2023 of N = 521 for a sample comprising nine individuals and a simple descriptive model was provided to assess the relations between the IVs and NI. These measures display central distribution features and the dispersion of the variables and give confidence intervals. Almost all the variables follow normal distribution around the mean to show some type of heterogeneity in performance by unit (firms or countries). Outliers are few, as the extremes are more or less consistent with the ones in the other scale. Standard deviations for both shareholders equity (SE) and interest income (IIR) as a measure of profit show reasonably low standard deviations when compared to the mean, suggesting the importance of such a stability for these indicators. Net income (NI) and operating income (OI), however, have a higher variation, showing different levels of profitability according to units. Total assets (TA): between 80 million and 100 million, the average of 89.8 million, that indicate there are fairly large entities included in the sample. Market share (MS) ranges from 2% to 29%, mean 19.37%, reflecting higher competition and large market power of the companies. Operating expenses (OE) comprise a somewhat minor fraction of revenue which could imply successful control. Operating income (OI) is comparatively high, a sizable portion of income, reflecting good operational efficiency.



**Figure 2.** Decision tree digramm.

This diagram 2 describes a decision tree structure to predict the net income from a series of financial variables. As an analytical technique, this tree is used for understanding the correlation between the financial factors on the profitability of an organization and making a choice and divide the decision through the utilization of statistical conditions e.g., adjusted P-value and F-test ratio. The decision tree indicates interest expense has a significant impact on net income with the first split based on a value of 96,888 ( $P < 0.001$ ). In a low-interest expense group, shareholder equity is the next key determinant, segregating companies with decent capital among others. This progression makes the obvious distinction that financial soundness through the substitution of less leverage and higher equity is the foundation to the sustainable profitability.



**Figure 3.** The predictions of the decision tree (DT) model and the generalized linear regression model (GLM).

The DT vs GLM for net income (NI) comparison graph (**Figure 3**) shows a chart comparing DT vs GLM for forecast performance. This shows the GLM model generates variable predictions that look like random movement, suggesting overfitting and being unable to track the trend of the entire model. The DT model can be expected to show more regular behaviour as compared to the GLM model. Thus the above results emphasize the need to adopt pruning techniques or obtain a simple but robust model on data, given that the data is very volatile.

**Table 2.** The performance of the Decision Tree (DT) and Generalized Linear Regression (GLM).

	SMAPE	MAPE	MASE
DT	0.119287262	0.067908525	0.06790853
GLM	0.311547944	0.204240856	0.20424086

The comparison **Table 2** of performance of the Decision Tree (DT) and Generalized Linear Regression (GLM) models reveals an obvious and considerable superiority of the Decision Tree model with respect to predictive accuracy for all their performance metrics. The DT model achieves 3x better predictive accuracy with the mean absolute percentage error (MAPE) of 6.79% compared to the GLM 20.42%, indicating it is better able to capture the true pattern of net income. This is further supported by the SMAPE measure, where it demonstrates a similar observation with DT achieving 11.93% as compared to 31.15% for the GLM, indicating DT excels with regard to relative change and volatility without bias. The MASE metric also indicates that both models are better than the naive expectation (where the value is less than 1), but the DT model performs super well with 0.0679, outperforming the reference model by more than 93% compared to 79.6% for the GLM. In particular, this outstanding performance of the decision tree results from its ability to model more complex nonlinear relationships and interactions between variables (which linear models such as GLM are unable to effectively represent or fit), particularly during economic shocks. Therefore, the performance of the metrics is in line with the hypothesis that the decision tree model is the best model for accurate prediction of the NET INCOME dynamics. The poor performance of the GLM model reflects the limitations of linear assumptions in capturing dynamic patterns and interactions between financial variables, resulting in a decrease in its predictive power.

## 5. Conclusion

This study represents a qualitative contribution to the field of corporate finance and economic analysis. It seeks to understand the relationship between financial and investment indicators, with a focus on developing accurate predictive models for net income. This interest stems from the critical importance of net income as a key measure of a company's profitability, financial stability, and competitiveness, especially in volatile economic environments. The study relied on a wide and

diverse dataset, including key financial indicators, to build and compare advanced predictive models. The results showed that the Decision Tree (DT) model significantly outperforms the Generalized Linear Regression Model (GLM) in predictive accuracy across all performance metrics. This superiority indicates that machine learning models are capable of capturing nonlinear patterns and complex interactions between variables such as the impact of operating income at certain levels of market share or operational efficiency that traditional statistical models cannot efficiently detect. The practical importance of the results lies in the fact that the decision tree model identified interest expenses and shareholder equity as the most prominent factors affecting net income, which indicates that improving the efficiency of the financing structure and strengthening the capital base can be among the priorities of financial management to achieve higher profitability stability. Through the graph comparing the predictions to the actual values, it is clear that: The GLM model provides relatively stable predictions, but fails to track large fluctuations. While the DT model shows greater accuracy in capturing peaks and troughs, despite some overfitting, which can be subsequently addressed through pruning or the use of hybrid models such as Random Forest. This study presents a new methodological framework that combines traditional quantitative analysis with modern machine learning tools. Accordingly, the study recommends the use of hybrid models that combine GLM and DT to maximize forecast accuracy while maintaining interpretability.

### Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

### References

- Abrokwah-Larbi, K., & Awuku-Larbi, Y. (2024). The Impact of Artificial Intelligence in Marketing on the Performance of Business Organizations: Evidence from SMES in an Emerging Economy. *Journal of Entrepreneurship in Emerging Economies*, *16*, 1090-1117. <https://doi.org/10.1108/jeee-07-2022-0207>
- Adu, D. A., Abedin, M. Z., Saa, V. Y., & Boateng, F. (2024). Bank Sustainability, Climate Change Initiatives and Financial Performance: The Role of Corporate Governance. *International Review of Financial Analysis*, *95*, Article 103438. <https://doi.org/10.1016/j.irfa.2024.103438>
- Ahnert, T., Doerr, S., Pierri, M. N., & Timmer, M. Y. (2021). *Does IT Help? Information Technology in Banking and Entrepreneurship*. International Monetary Fund.
- Alonso Robisco, A., & Carbó Martínez, J. M. (2022). Measuring the Model Risk-Adjusted Performance of Machine Learning Algorithms in Credit Default Prediction. *Financial Innovation*, *8*, Article No. 70. <https://doi.org/10.1186/s40854-022-00366-1>
- Behera, S., Nayak, S. C., & Kumar, A. V. S. P. (2023). A Comprehensive Survey on Higher Order Neural Networks and Evolutionary Optimization Learning Algorithms in Financial Time Series Forecasting. *Archives of Computational Methods in Engineering*, *30*, 4401-4448. <https://doi.org/10.1007/s11831-023-09942-9>
- Bughin, J., Hazan, E., Sree Ramaswamy, P., & Chu, M. (2017). *Artificial Intelligence: The Next Digital Frontier*. McKinsey Global Institute.

- <https://www.mckinsey.com/~media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.pdf>
- Chhaidar, A., Abdelhedi, M., & Abdelkafi, I. (2023). The Effect of Financial Technology Investment Level on European Banks' Profitability. *Journal of the Knowledge Economy*, 14, 2959-2981. <https://doi.org/10.1007/s13132-022-00992-1>
- Cipriani, M., & La Spada, G. (2021). Investors' Appetite for Money-Like Assets: The MMF Industry after the 2014 Regulatory Reform. *Journal of Financial Economics*, 140, 250-269. <https://doi.org/10.1016/j.jfineco.2020.11.005>
- Dang, T. V., Wang, H., & Yao, A. (2017). Chinese Shadow Banking: Bank-Centric Misperceptions. *SSRN*.
- Eboigbe, E. O., Farayola, O. A., Olatoye, F. O., Nnabugwu, O. C., & Daraojimba, C. (2023). Business Intelligence Transformation through AI and Data Analytics. *Engineering Science & Technology Journal*, 4, 285-307. <https://doi.org/10.51594/estj.v4i5.616>
- Elsheikh, A. M., & Elhag, A. A. (2025). Machine Learning-Based Analysis of Multi-Region Bone Fracture Detection and Classification Using Biomedical Images. *Alexandria Engineering Journal*, 128, 186-199. <https://doi.org/10.1016/j.aej.2025.05.074>
- Ferraro, P. J., & Miranda, J. J. (2017). Panel Data Designs and Estimators as Substitutes for Randomized Controlled Trials in the Evaluation of Public Programs. *Journal of the Association of Environmental and Resource Economists*, 4, 281-317.
- Goldstein, H. (2011). *Multilevel Statistical Models*. John Wiley & Sons. <https://doi.org/10.1002/9780470973394>
- Härdle, W. (2004). *Nonparametric and Semiparametric Models*. Springer Science & Business Media.
- Huang, J., Huang, Z., & Shao, X. (2023). The Risk of Implicit Guarantees: Evidence from Shadow Banks in China. *Review of Finance*, 27, 1521-1544. <https://doi.org/10.1093/rof/rfac061>
- Luo, R., Fang, H., Liu, J., & Zhao, S. (2019). Maturity Mismatch and Incentives: Evidence from Bank Issued Wealth Management Products in China. *Journal of Banking & Finance*, 107, Article 105615. <https://doi.org/10.1016/j.jbankfin.2019.105615>
- Murthy, S. K. (1998). Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey. *Data Mining and Knowledge Discovery*, 2, 345-389. <https://doi.org/10.1023/a:1009744630224>
- Porter, M. E. (2008) *Competitive Advantage: Creating and Sustaining Superior Performance*. Simon and Schuster.
- Ratia, M., Myllärniemi, J., & Helander, N. (2018). Robotic Process Automation-Creating Value by Digitalizing Work in the Private Healthcare? In *Proceedings of the 22nd International Academic Mindtrek Conference* (pp. 222-227). ACM. <https://doi.org/10.1145/3275116.3275129>
- Rivera-Lopez, R., Canul-Reich, J., Mezura-Montes, E., & Cruz-Chávez, M. A. (2022). Induction of Decision Trees as Classification Models through Metaheuristics. *Swarm and Evolutionary Computation*, 69, Article 101006. <https://doi.org/10.1016/j.swevo.2021.101006>
- Saenz, M., Revilla, E., & Simón, C. (2020). Designing AI Systems with Human Machine Teams. *MIT Sloan Management Review*.
- Subburayan, B., Sankarkumar, A. V., Singh, R., & Mushi, H. M. (2024). Transforming of

- the Financial Landscape from 4.0 to 5.0: Exploring the Integration of Blockchain, and Artificial Intelligence. In M. Irfan, K. Muhammad, N. Naifar, & M. A. Khan (Eds.), *Financial Mathematics and Fintech* (pp. 137-161). Springer International Publishing. [https://doi.org/10.1007/978-3-031-47324-1\\_9](https://doi.org/10.1007/978-3-031-47324-1_9)
- Sullivan, Y., & Fosso Wamba, S. (2024). Artificial Intelligence and Adaptive Response to Market Changes: A Strategy to Enhance Firm Performance and Innovation. *Journal of Business Research*, 174, Article 114500. <https://doi.org/10.1016/j.jbusres.2024.114500>
- Takahashi, F. L., & Vasconcelos, M. R. (2024). Bank Efficiency and Undesirable Output: An Analysis of Non-Performing Loans in the Brazilian Banking Sector. *Finance Research Letters*, 59, Article 104651. <https://doi.org/10.1016/j.frl.2023.104651>
- Vujović, Ž. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*, 12, 599-606. <https://doi.org/10.14569/ijacsa.2021.0120670>
- Wahid ElKelish, W., & Kamal Hassan, M. (2014). Organizational Culture and Corporate Risk Disclosure: An Empirical Investigation for United Arab Emirates Listed Companies. *International Journal of Commerce and Management*, 24, 279-299. <https://doi.org/10.1108/ijcoma-06-2012-0035>
- Yu, H., Kuo, L., & Kao, M. (2017). The Relationship between CSR Disclosure and Competitive Advantage. *Sustainability Accounting, Management and Policy Journal*, 8, 547-570. <https://doi.org/10.1108/sampj-11-2016-0086>
- Zaki, M. (2019). Digital Transformation: Harnessing Digital Technologies for the Next Generation of Services. *Journal of Services Marketing*, 33, 429-435. <https://doi.org/10.1108/jsm-01-2019-0034>
- Zequiraj, V., Gurdgiev, C., Sohag, K., & Hammoudeh, S. (2024). Economic Uncertainty, Public Debt and Non-Performing Loans in the Eurozone: Three Systemic Crises. *International Review of Financial Analysis*, 93, Article 103208. <https://doi.org/10.1016/j.irfa.2024.103208>