Journal of Soft Computing and Decision Analytics, Volume 3, Issue 1 (2025) 50-71

Journal of Soft Computing and Decision Analytics

Journal homepage: www.jscda-journal.org
ISSN: 3009-3481

Analysis Impact of Financial Ratios on Bank Success Using Machine Learning Classification Algorithms: The Case of Turkey

İsmail Dere[1], Münevver Turanlı[2], Selçuk Alp[3*], Mefule Fındıkçı Erdoğan[4]

[1] Istanbul Ticaret University Istanbul, Türkiye
[2] Department of Economics, Faculty of Business, Istanbul Ticaret University Istanbul, Türkiye
[3] Department of Statistics, Faculty of Arts and Sciences, Yıldız Technical University, Istanbul, Turkiye
[4] Department of Economics, Faculty of Business, Istanbul Ticaret University Istanbul, Türkiye

ARTICLE INFO

Article history:
Received 26 December 2024
Received in revised form 21 January 2025
Accepted 26 January 2025
Available online 28 January 2025

Keywords:
Machine Learning Alorithms; Classification; Bank Success; Financial Ratios.

ABSTRACT

In the competitive environment of the 21st century, making accurate and swift decisions is of great importance for the success of businesses. In the banking sector, which provides financial services and is critical to the economy, decision-making processes are considered a crucial step in enhancing bank performance, operational efficiency, and customer satisfaction. However, traditional methods that generally rely on past experiences and intuition are found to be inadequate for analysing large data sets. In this context, it is considered that the use of artificial intelligence in analysing large data sets will provide significant advantages by increasing the speed and accuracy of decision-making processes. Therefore, the effective use of large data sets, optimization of decision-making processes, and the use of artificial intelligence to increase bank success can be highlighted as important tools. In this study, the factors determining the success of banks operating in the Turkish banking sector and the factors that should be considered in decision-making processes for bank success were examined using machine learning methods. The study, which covers the period from 2012 to 2022 for 24 banks, classified 43 financial ratios into six groups: capital adequacy, profitability, liquidity, asset quality, balance sheet structure, and income-expenditure structure. Thus, the effects of factors determining bank success in the context of these main groups were analysed. Additionally, a comprehensive analysis using all 43 financial ratios was conducted to provide a general examination of the factors determining bank success. The study concluded that machine learning methods, with their high accuracy rates, can be effectively used in decision-making, monitoring, and auditing processes.

1. Introduction

In today's rapidly digitalized world, technological innovations fundamentally change the working methods of businesses. Artificial intelligence (AI), which plays an important role in this change, makes significant contributions to businesses to make their processes more accurate and efficient. Machine

* *Corresponding author.*
E-mail address: alp@yildiz.edu.tr

https://doi.org/10.31181/jscda31202553

learning (ML), as a branch of AI, is a powerful tool used to solve various problems of businesses. ML techniques such as classification, clustering, regression and optimization allow businesses to make sense of data sets and make future predictions. In the banking sector, ML algorithms are effectively used in areas such as customer segmentation, credit risk analysis, fraud detection and development of marketing strategies. In the literature, machine learning has been successfully applied to analyses bank profitability, asset quality and lending behavior [1-3]. However, the application of AI to the evaluation of overall bank performance and success using financial ratios is limited in the literature. In this study, the success of the banks in the Turkish banking sector in terms of their activities is analyses by using ML algorithms. This study analyzes a comprehensive set of financial ratios of 24 banks operating in Turkey for the 2012-2022 operating period. In this scope, the ratio of net operating profit/loss, which expresses the net profits/losses of banks only within the scope of their activities, to total assets has been determined as the dependent variable of the study. This variable is averaged for each year in the 2012-2022 period, and banks above the average are coded as successful, while banks below the average are coded as unsuccessful. By averaging the variable for each year of the relevant period, it is aimed to take into account the opportunities and vulnerabilities that occur conjecturally. On the other hand, the independent variables of the study consisted of 6 main groups including 43 financial ratios: capital adequacy, profitability, liquidity, asset quality, balance sheet and income-expense.

The remainder of the study is organized as follows. In the second section, the conceptual background of AI and a related literature review are given. In the third chapter, firstly, the concepts and applications of ML are explained. Then, the classification methods in machine learning are discussed and Logistic Regression (LR), K-Nearest Neighbor (kNN), Support Vector Machines (SVM), Artificial Neural Networks (ANN), Classification and regression trees (CART), Random Forest (RF), Gradient Boosting (GBoost), Extreme Gradient Boosting (XGBoost), Light Gradient-Boosting Machine (LightGBM) and Categorical Boosting (CatBoost) algorithms used in this study are explained. The last part of the chapter, the evaluation criteria of classification model performances are presented in terms of accuracy, precision, sensitivity/recall and F1-measure. The fourth section includes the applications and the results of the applications.

2. Literature Review

In the banking and finance sector, studies in the literature focus on how it can be used effectively in various areas such as efficiency, risk management, customer relations and financial performance.

Halkos & Salamouris [4] used data envelopment analysis (DEA), a nonparametric analytical method, to evaluate the performance of the Greek banking sector. The study analyzes the efficiency of Greek banks based on financial efficiency ratios for the period 1997-1999.

Wu *et al.,* [5] integrated DEA and ANN to examine the efficiency of 142 branches of a Canadian bank in the Toronto area. The results show that the ANN combination provides more efficient results than the standard DEA.

Nur Ozkan-Gunay & Ozkan [6] explained bank failures in emerging financial markets using ANN with the case of Turkey. In the period 1999-2001, 23 out of 59 banks operating in Turkey went bankrupt, were closed down or merged with another bank, and it was argued that the symptoms of these failing banks emerged well in advance and therefore could be predicted by analyzing financial ratios.

Leo *et al.,* [7] analyzed and evaluated ML methods in the scope of risk management in the banking sector. The study focused on how risks are identified, measured, reported and managed, and determined that ML methods are used in the management of banking risks such as market risk, credit risk, operational risk and liquidity risk.

Malali & Gopalakrishnan [8] examined the application of AI and powered technologies in the banking and finance sector in India. The study also examines the AI ecosystem in the banking and finance sector, focusing on the critical unsolved problems of this field. As a result, the effects and significance of AI on the business environment of the banking and financial service industry are emphasized.

Königstorfer & Thalmann [9] discussed the use of AI in the banking sector. The study examines how AI can be applied in the banking sector. It has been stated that AI in the banking sector provides great benefits in many areas such as improving customer experience, optimizing risk management, fraud detection and creating more efficient business processes. It was also pointed out that ML algorithms are able to detect fraud attempts by analyzing large data sets, thereby increasing the security of banks.

Appiahene *et al.,* [10] performed a comparative analysis of ML algorithms for predicting operational efficiency in the banking sector. The study was carried out at a time when the financial crisis in Ghana between 2015 and 2018 raised several issues related to the efficiency of banks and the safety of depositors. The study evaluates bank efficiency and performance in the sample of 444 Ghanaian bank branches (Decision Unit) using a hybrid of DEA and three ML approaches.

Fares *et al.,* [11] presented a systematic and integrated literature review of the studies on the use of AI in the banking sector between 2005 and 2023. In the study, 44 academic articles were analyzed by content analysis method to present AI practices, opportunities, challenges and impacts in the banking sector. As a result of the study, it was determined that AI applications in the banking literature are concentrated in three main research areas (Strategic Planning, Operational Processes and Customer Relations).

Umamaheswari & Valarmathi [12] examined the impact of AI applications in the financial sector. The study emphasized that AI provides advantages such as reducing costs in the financial sector, speeding up transactions, offering security applications and being always accessible. With a survey of 500 people, the use of AI by users according to their age, income and education levels was examined, and it was determined that as age, income and education levels increase, trust in AI also increases.

Doumpos *et al.,* [13] examined how AI is currently used in the banking sector and how it could be used in the future. The study presents examples of the use of AI in various fields such as customer service, fraud detection, risk management, credit assessment and marketing, and argues how AI can be used more effectively and efficiently in the banking sector.

Gangwani & Zhu [14] proposed a systematic analysis of modelling and predicting business success in their study. First, three business-related features (Investment-Business-Market (IBM)) were introduced. Then, each of these attributes is examined and modelled from a specific perspective such as sales, management, innovation, etc. Furthermore, the use of different machine learning and deep learning methods for business modelling and prediction is described. The research provides a comprehensive review of computational approaches to business performance modelling and prediction.

3. Methodology

3.1 Machine Learning

AI aims for computer systems to have human-like intelligence. This idea was first proposed by Alan Turing in 1950 with the question “Can machines think?”. The term AI was first officially used in

1956 by John McCarthy at the “Dartmouth Workshop” [15]. AI is the ability of a machine to learn from its own experiences and make decisions based on these learnings [16]. It aims to design systems that can perform complex tasks such as data analysis, pattern recognition, automatic decision making through computer programs and algorithms. These systems include technological developments in different fields such as machine learning, deep learning, natural language processing, expert systems, and autonomous vehicles [17]. AI is also closely related to technological trends such as big data analytics, cloud computing, the Internet of Things (IoT) and cybersecurity. These technologies increase the potential of AI and provide new areas of application [18]. AI is becoming an increasingly used technology to optimize the processes of businesses, increase efficiency and provide competitive advantage [19]. At this perspective, increased productivity [20], fast and accurate decisions [21], cost savings, customer service improvements [22], competitive advantage [23] can be listed as advantages of AI.

Machine Learning (ML), a branch of AI, is an approach that enables computers to learn patterns by analyzing data. Learning is provided by various techniques such as supervised, unsupervised and reinforcement learning. ML is used in many areas such as classification, regression, clustering [24].

3.2 Machine Learning Applications

ML is used in many different application areas with its ability to learn and predict from data. In this section, the main application areas of ML will be examined under the titles of classification, clustering, regression and optimization.

3.2.1 Classification

Classification is an application of AI in the supervised learning category and aims to assign elements from a dataset to predefined classes. Classification includes two phases: model building and model testing. In the model building phase, each instance is assumed to belong to a predefined class specified by the class variable. The set of samples used for model building is called training data. Classification algorithms are used to develop models and rules for predicting the class variable. In the model testing phase, the accuracy of the model is estimated by comparing the known class value of the samples in the test data with the class value obtained as a result of the model.

Classification is of great importance, especially in areas such as medical diagnosis, financial decisions and marketing. For example, in medical diagnostic systems, patients can be classified according to their symptoms. Similarly, risk assessment of loan applications in the financial sector is also performed with classification algorithms [25, 26].

3.2.2 Clustering

Clustering algorithms are unsupervised learning algorithms that group the elements in the data set according to their features. These algorithms, which have an important place in the field of AI, aim to separate elements with similar features in the data set into groups. Clustering algorithms are used to discover hidden patterns and structures in a data set. Clustering differs from classification in that the definition of classes is not known in prior [27, 28].

Clustering algorithms are used to discover groups with similar features or to determine the differences between them. There are various clustering algorithms such as K-Means, Fuzzy C-means, Kohonen Neural Networks, K-Medoids, Canopy, Mean Shift, MinHash and Latent Dirichlet Allocation. These algorithms are widely used in determining marketing strategies, biological data analysis and social network analysis [29].

3.2.3 Regression Analysis

Regression analysis is an analysis method that examines the relationship between a dependent variable and one or more independent variables. If there is one independent variable in the problem, it is called univariate regression analysis, and if there are more than one independent variable, it is called multivariate regression analysis. Regression analysis provides information on the relationship between variables and, if there is a relationship, its degree of significance [30].

3.2.4 Optimization

Optimization refers to a procedure that aims to find the best solution to achieve a given goal. This procedure involves finding the optimal solution by taking into various parameters and constraints. ML is an effective tool for solving optimization problems and allows solving complex problems with large data sets. Especially in high-dimensional and multivariate data sets, ML algorithms are faster and more effective than traditional optimization algorithms.

Optimization algorithms present tools for improving industrial processes, reducing costs and increasing productivity. The correct execution of these algorithms not only increases the competitiveness of businesses, but also makes significant contributions in terms of sustainability and innovation [31, 32].

3.3 Classification Algorithms in Machine Learning

ML has developed rapidly in the recent years and has found a wide variety of applications. Especially the need to analyze large datasets and to extract useful results from these data increases the importance of ML. Classification is one of the main and widely used application areas of ML. The ML classification algorithms used in the application of the study are presented below.

3.3.1 Logistic Regression

LR is a statistical method widely used in classification problems. This algorithm, which is especially preferred in binary classification problems, provides effective results when the dependent variable is categorical. The logistic regression model uses independent variables to predict the probability that the dependent variable belongs to a class. This prediction is performed using a logistic function (sigmoid function) and the model provides the probabilities to take a value between 0 and 1 [33].

LR is to model probabilities with a logit transformation. This transformation takes a form that can be expressed by a linear combination and can be easily optimized. The logit function expresses the logarithmic ratio between the probability of an event occurring (p) and the probability of it not occurring (1-p) and is defined as Eq. (1).

$$Logit\ Function = log\left(\frac{p}{1p}\right) \tag{1}$$

The predictive capacity of the model is optimized using the maximum likelihood method. This method maximizes the probability of the observed data set under the model parameters (Menard, 2022). The model is expressed as Eq. (2).

$$logit(p) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \tag{2}$$

Where, β_0 represents the constant term of the model and $\beta_1, \beta_2, \dots, \beta_n$ represent the coefficients of the independent variables. Each of the independent variables has an increasing or decreasing effect on the logit probability of the dependent variable [34].

3.3.2 K-Nearest Neighbors

The KNN algorithm is one of the supervised learning methods and is widely used in classification problems. The fundamental principle of the KNN algorithm is to use the features of *k* nearest neighbors to classify new data. It makes decisions by evaluating the similarity of data with distance measurements [35].

In the KNN algorithm, the value *k* represents the number of nearest neighboring data around the new data. For example, when k=3, the three closest neighbors of the new data are determined and the class to which these neighbors belong, and the new data is considered to be in the most frequent class. The distances between data can be calculated by methods such as Euclidean Distance, Manhattan Distance and Minkowski Distance, etc. [36].

3.3.3 Support Vector Machine

SVM is an ML algorithm founded on strong statistical theories used for regression and classification. SVM primarily aims to optimally separate data belonging to two classes. This algorithm separates the data linearly with the help of a decision function obtained using the training data. This algorithm develops a model that assigns the training data to different classes, usually using a linear classifier. The line that best splits the classes into two is called the “decision line”. Although many lines can be drawn, the aim of SVM is to determine the optimal decision line [37].

3.3.4 Artificial Neural Networks

Based on the learning capacity of the human brain, ANNs are information processing systems capable of problem solving with the ability to discover, memorize, process and generate new information. These networks provide faster, practical and lower-error predictions for complex problems [15]. An ANN contains interconnected artificial neurons and these neurons are organized in layers. These layers are organized as an input layer, one or more hidden layers and an output layer. Each neuron multiplies its inputs by a given weight and produces an output using an activation function [38].

ANNs are also used effectively in prediction, clustering and classification problems [38]. One of the most important advantages of ANNs is their ability to learn complex and non-linear relationships. This feature makes ANNs successful in various applications such as image recognition, voice processing, natural language processing and financial forecasting [39].

3.3.5 Classification and Regression Trees

The CART algorithm is a powerful and flexible ML method for both classification and regression problems. Developed in 1984 by Breiman et al, this algorithm is widely used in data mining and prediction modelling [40]. In the CART algorithm, binary tree structures are used to divide the data set into more homogenous subgroups. In this process, the tree structure contains a decision rule at each node that divides the data into two subgroups. These decision rules for classification trees are based on whether a given variable exceeds a certain threshold value [41].

The CART algorithm is based on tree growing and pruning. *Growing the Tree:* The algorithm selects the best split by evaluating each variable and possible threshold values in the data set. The split is performed using measures such as the Gini index or entropy. *Pruning the tree:* The grown tree can often be very large and complex. To avoid overfitting, the tree is pruned to obtain a simpler model by cutting unnecessary branches. This process is usually supported by cross-validation [42].

The CART algorithm has been applied in various fields such as medicine, finance and marketing, helping to extract meaningful information from complex data sets. For example, it is effectively used in problems such as customer segmentation, credit risk assessment and disease diagnosis.

3.3.6 Random Forests

RF refers to a tree-based ensemble structured by random variables [43]. The RF method is a supervised ML algorithm used to solve classification and regression problems. This algorithm uses a combination of several decision trees generated by independent random sampling of the training data. In the branching of decision trees, a random subset is used instead of the best features, which increases the diversity of the model and reduces overfitting [44].

RF is based on a combination of decision trees, bagging and boosting methods and is one of the ensemble algorithms. The model is based on the principle that many decision trees created with randomly selected subspaces on the data set perform classification through majority voting [45].

3.3.7 Gradient Boosting

GBoost is an ensemble algorithm used to make forecasting models more powerful and effective. This algorithm aims to improve the accuracy of the model by adding successive weak estimators (usually decision trees). Each new model focuses on correcting the errors of the previous model and this process is optimized by a gradient descent algorithm [46].

The GBoost algorithm is an algorithm that works by using the whole data set and does not divide the data set into subgroups. This algorithm is based on building a decision tree from the data set and taking into account the errors of this tree, it builds a new decision tree. Each new tree is generated in order to minimize the errors of the previous tree. In this process, each new tree aims to reduce the error rate of the previous tree, while aiming to reduce the difference between the actual values and the predicted values closer to zero. Thus, the accuracy of the model increases and its errors gradually decrease [47].

GBoost gives effective results especially in classification and regression problems. This algorithm uses various hyperparameter settings to control model complexity and overfitting. For example, parameters such as learning rate and max depth are of great importance in tuning the performance and generalization ability of the model [48].

3.3.8 XGBoost (eXtreme Gradient Boosting)

XGBoost is an ML algorithm developed by Chen & Guestrin [49], which is widely used for various tasks such as classification, regression and ranking, and has become very popular in recent years. This algorithm is an optimized version of the GBoost algorithm and performs especially well on large datasets and high dimensional feature problems.

XGBoost is based primarily on the principles of GBoost. In the GBoost algorithm, each new model attempts to correct the errors of the previous model. XGBoost includes several improvements to make this process more efficient and effective. XGBoost uses L1 (Lasso) and L2 (Ridge) regularization techniques to control model complexity and prevent overfitting. Thanks to its parallel computing capability, it provides fast model training with modern multi-core processors. Due to its ability to deal with incomplete data, it simplifies the data pre-processing process. In addition, it prunes unnecessary branches by performing gain calculations while building tree structures, increasing the efficiency of the model and preventing overfitting. Thus, the prediction accuracy of the model gradually increases [50].

3.3.9 LightGBM

LightGBM is an ML algorithm that aims to improve the performance of the XGBoost algorithm. It is a decision tree based algorithm and is used in regression and classification problems. LightGBM sometimes shows better performance than the XGBoost algorithm. Like XGBoost, bagging and boosting operations are performed. Unlike other decision tree algorithms, LightGBM performs leaf-based growth instead of tree-based growth. There are more than 100 parameters in the algorithm, which makes the algorithm flexible [51].

One of the most important advantages of LightGBM is that it uses a histogram-based learning algorithm. This algorithm splits the data set into a series of discrete thousands and thus uses less memory to generate decision trees faster. Moreover, LightGBM creates deeper and more complex trees using a leaf-wise growth strategy. This provides performance gains, especially when working with large data sets and high-dimensional data [52].

3.3.10 Categorical Boosting

CatBoost is a GBoost algorithm that shows high performance, especially when working with categorical features. One of the main advantages of this algorithm is the ability to process categorical data directly. CatBoost reduces bias in the processing of categorical features, resulting in more accurate and reliable estimates. This provides a significant advantage, especially when working on complex datasets [53].

CatBoost is based on the GBoost algorithm, but includes several important innovations that distinguish it from other boosting algorithms. Firstly, CatBoost has the ability to process categorical data directly. This eliminates the need to pre-process categorical data and allows the model to work more efficiently with this data [53]. Secondly, CatBoost increases the overall performance of the model by reducing bias through randomized resampling in the training of each weak learner [54]. In addition, symmetric tree structures are used in the training process of the model, which enables the model to be trained faster and more efficiently. These innovations contribute to speeding up the training process and improving the overall performance of CatBoost, while enabling the algorithm to be used effectively in various application areas.

3.4 Performance Metrics for Classification Algorithms

There are various metrics for evaluating and comparing the effectiveness of ML classification algorithms. These metrics are used to determine which method gives more successful results and to compare the performance of the methods. Since it may be misleading to accept a single metric as a success indicator, it is important to evaluate more than one metric together.

The confusion matrix, as shown in Figure 1, is a table which is used to describe the performance of a classification algorithm.

The confusion matrix consists of four characteristics used to describe the performance metrics of classification algorithms. These are:

TP (True Positive): The number of data whose actual value is 1 and predicted as 1 by the model.

TN (True Negative): The number of data whose actual value is 0 and predicted as 0 by the model.

FP (False Positive): The number of data whose actual value is 0 and predicted as 1 by the model.

FP (False Negative): The number of data whose actual value is 1 and predicted as 0 by the model.

The performance measures of a classification algorithm are accuracy, precision, recall and F1 score calculated using the above values of TP, TN, FP and FN.

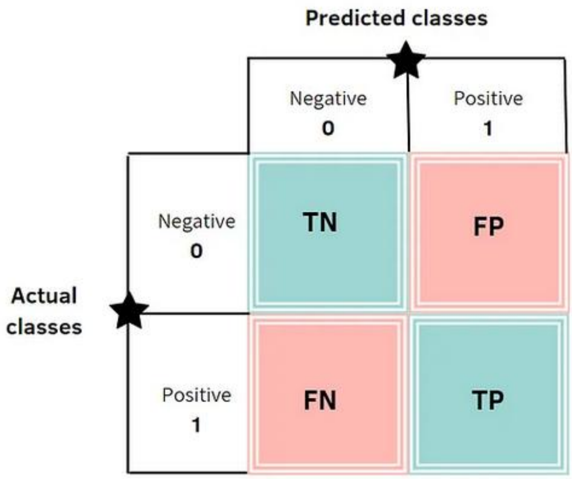

Fig. 1. Confusion Matrix

3.4.1 Accuracy

Accuracy is defined as the ratio of the number of samples correctly classified by the model to the total number of samples. In other words, it is calculated as the sum of true positives (TP) and true negatives (TN) divided by the sum of all positives (TP and FP) and negatives (TN and FN). The accuracy is expressed as in Eq. (3).

$$Accuracy = \frac{TP+TN}{TP+\mathrm{FP}+\mathrm{TN}+FN} \tag{3}$$

Although accuracy is widely used to assess the overall performance of the model, it is not a sufficient metric, especially when the dataset is unbalanced between classes. In such cases, other metrics need to be considered in order to further evaluate the performance of the model.

3.4.2 Precision

Precision shows how many of the samples that the model classifies as positive are actually positive. Precision is an important evaluation metric, especially in unbalanced data sets and when the cost of false positives is high. Precision is calculated as the ratio of true positives (TP) to the sum of true positives (TP) and false positives (FP) as in Eq. (4).

$$Precision = \frac{TP}{TP+\mathrm{FP}} \tag{4}$$

This metric is also known as the positive prediction rate. Precision is especially important in areas where the cost of false positives is high. For example, in medical diagnostics it is of great importance to correctly identify the presence of a disease.

3.4.3 Recall

Recall is a performance metric that shows how accurately a classification algorithm identifies the positive class. This metric (Eq. 5) refers to the ability of the model to accurately identify true positives.

$$Recall = \frac{TP}{TP+\mathrm{FN}} \tag{5}$$

A high recall value means that the model produces few False negatives (FN) and therefore identifies the positive class well. It is of great importance to be high especially in areas such as medical diagnosis and security. Because missing positive cases in such applications can have serious consequences.

3.4.4 F1-Score

Precision and recall alone are not sufficient to evaluate the overall performance of the model. The F1 score gives a more comprehensive evaluation of the overall performance of the model, equilibrium precision and recall. The F1 score is especially important when there is imbalance between classes in data sets or when the costs of false positive and false negative errors are high.

The F1 score is calculated as the harmonic mean of precision and recall and is given by Eq. (6):

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recal} \tag{6}$$

The high F1 score means that the model shows balanced and good performance in terms of both precision and recall.

4. Application

4.1 Dataset

The study analyses whether the banks operating in Turkey are successful in terms of their activities with respect to financial ratios. In this scope, ML algorithms are used to evaluate the financial performance of 24 selected banks operating in Turkey for the period 2012-2022. In the study, in order to evaluate the success of the banks, the ratio of *Net Operating Profit (Loss)*, which expresses the profits/losses from operations, to *Total Assets* was determined as the dependent variable. The average of the dependent variable (*Net Operating Profit (Loss)/Total Assets*) was taken for each year in the period analyzed and the banks above the average were considered successful, while the banks below the average were considered unsuccessful. In this way, it has enabled the assessment of fragilities and opportunities that arise by taking into account conjuncture differences.

The independent variables of the study consist of a wide range of financial ratios. In this respect, 43 financial ratios (Appendix 1) are categorized under six main groups: Capital Ratios, Profitability, Liquidity, Asset Quality, Balance Sheet Ratios, and Income-Expenditure Structure. By using different ML algorithms, it is aimed to determine the criteria that are effective in bank success with a wide range of financial ratios in an emerging market and to develop a model proposal that will be useful in monitoring, auditing and decision making.

The *Capital Ratios* used in the study constitutes a critical indicator of financial stability as it measures a bank's ability to withstand potential losses. While *Profitability Ratios* give an idea about the earning capacity and efficiency of a bank, Liquidity Ratios express the bank's capacity to pay its short-term liabilities [55]. Studies in the literature show that *Profitability* and *Liquidity Ratios* are strong determinants of financial distress [56]. On the other hand, while *Asset Quality* reflected in metrics such as non-performing loans and loan loss provisions shows the general structure of a bank's loan portfolio, the *Balance Sheet Ratios*, which includes the composition of assets and liabilities, provides a comprehensive view of a bank's financial position. *Income-Expense Structure Ratios* show the bank's income generation capabilities and cost management practices [57].

In order to determine the important criteria (financial ratios) that are effective in the success of the bank in different fields and scales, the independent variables are categorized into 6 main groups. In addition, an integrated analysis where all financial ratios are considered together is also performed.

4.2 Findings

Classification algorithms are used to determine which of the previously known classes a data is to be included in. In this study, 264 data of 24 banks for the period 2012-2022 were used to train ML algorithms (LR, kNN, SVM, ANN, CART, RF, GBoost, XGBoost, LightGBM and CatBoost). Using the grid search method, algorithms were run for different values of the parameters and analyses were performed according to the parameters that gave the best results. Accuracy, Precision, Recall and F1 Score metrics were used to evaluate the prediction success of ML algorithms.

4.2.1 Capital Ratios

In order to determine whether the banks are successful in terms of their operations, 7 ratios in the capital ratios group were analyzed.

The success/unsuccessful of banks using capital ratios were predicted by AI algorithms and the performance metrics are given in Table 1. LR showed the highest success level and KNN showed the lowest success level.

Table 1
Performance metrics for capital ratios

Algorithms	Accuracy	Precision	Recall	F1 Score
LR	0.77	0.77	0.77	0.77
kNN	0.62	0.62	0.61	0.61
SVM	0.66	0.68	0.66	0.66
ANN	0.66	0.66	0.66	0.66
CART	0.66	0.71	0.66	0.65
RF	0.71	0.72	0.71	0.71
GBoost	0.68	0.68	0.68	0.67
XGBoost	0.68	0.68	0.68	0.67
LightGBM	0.65	0.66	0.65	0.65
CatBoost	0.69	0.70	0.69	0.69

For capital ratios, the most effective financial ratios in determining whether the bank is successful or not in terms of its activity with the feature selection method are (D4) Shareholders' Equity / (Deposits + Non-Deposit Funds), (D2) Shareholders' Equity / Total Assets and (D3) (Shareholders' Equity - Permanent Assets) / Total Assets, respectively. The effect levels of capital ratios on bank success are presented in Table 2.

Table 2
The effect levels of capital ratios

ID	Ratios	Effect Levels
D4	Shareholders' Equity / (Deposits + Non-Deposit Funds	0.2134
D2	Shareholders' Equity / Total Assets	0.2052
D3	(Shareholders' Equity-Permanent Assets) / Total Assets	0.1663
D5	On-Balance Sheet Foreign Exchange Position/Shareholders' Equity	0.1117
D7	N(on+off) Balance-sheet Position / Total Shareholders' Equity	0.1116
D6	Net on Balance-sheet Position / Total Shareholders' Equity	0.1083
D1	Capital Adequacy Ratio	0.0853

4.2.2 Profitability

In order to determine the success of banks in terms of banking activities, 4 financial ratios in the *Profitability Ratios* group were used. The success/unsuccessful of banks using profitability ratios were predicted by AI algorithms and the performance metrics are given in Table 3. XGBoost and CatBoost showed the highest success level and kNN showed the lowest success level.

Table 3
Performance metrics for profitability ratios

Algorithms	Accuracy	Precision	Recall	F1 Score
LR	0.83	0.83	0.83	0.83
kNN	0.76	0.76	0.76	0.76
SVM	0.83	0.83	0.82	0.82
ANN	0.84	0.84	0.84	0.84
CART	0.81	0.86	0.82	0.81
RF	0.85	0.85	0.85	0.85
GBoost	0.85	0.85	0.85	0.85
XGBoost	0.88	0.88	0.88	0.88
LightGBM	0.86	0.86	0.86	0.86
CatBoost	0.88	0.87	0.87	0.87

For profitability ratios, the most effective financial ratios in determining whether the bank is successful or not in terms of its activity with the feature selection method are (D10) Income Before Taxes / Total Assets, (D8) Average Return on Assets and (D9) Average Return on Shareholders' Equity, respectively. The effect levels of profitability ratios on bank success are presented in Table 4.

Table 4
The effect levels of profitability ratios

ID	Ratios	Effect Levels
D10	Income Before Taxes / Total Assets	0.495
D8	Average Return on Assets	0.294
D9	Average Return on Shareholders' Equity	0.133
D11	Net Profit (Losses) / Paid-in Capital	0.077

4.2.3 Liquidity

In order to determine the success of banks in terms of banking activities, 5 financial ratios in the *Liquidity Ratios* group were used. The success/unsuccessful of banks using liquidity ratios were predicted by AI algorithms and the performance metrics are given in Table 5. XGBoost and CatBoost showed the highest success level and kNN showed the lowest success level.

Table 5
Performance metrics for liquidity ratios

Algorithms	Accuracy	Precision	Recall	F1 Score
LR	0.64	0.64	0.64	0.64
kNN	0.62	0.63	0.63	0.62
SVM	0.60	0.60	0.60	0.60
ANN	0.65	0.66	0.65	0.65
CART	0.56	0.56	0.56	0.56
RF	0.61	0.61	0.61	0.61
GBoost	0.65	0.65	0.65	0.65
XGBoost	0.64	0.64	0.64	0.64
LightGBM	0.68	0.67	0.67	0.67
CatBoost	0.65	0.65	0.65	0.65

For liquidity ratios, the most effective financial ratios in determining whether the bank is successful or not in terms of its activity with the feature selection method are (D16) FC Liquid Assets / FC Liabilities, (D15) Liquid Assets / (Deposits + Non-Deposit Funds) and (D14) TC Liquid Assets / Total Assets, respectively. The effect levels of liquidity ratios on bank success are presented in Table 6.

Table 6
The effect levels of liquidity ratios

ID	Ratios	Effect Levels
D16	FC Liquid Assets / FC Liabilities	0.3155
D15	Liquid Assets / (Deposits + Non-Deposit Funds)	0.1901
D14	TC Liquid Assets / Total Assets	0.1843
D13	Liquid Assets / Short-term Liabilities	0.1728
D12	Liquid Assets / Total Assets	0.1373

4.2.4 Assets Quality

In order to determine the success of banks in terms of banking activities, 6 financial ratios in the *Assets Quality Ratios* group were used. The success/unsuccessful of banks using assets quality ratios were predicted by AI algorithms and the performance metrics are given in Table 7. CART showed the highest success level and SVM and LR showed the lowest success level.

Table 7
Performance metrics for assets quality ratios

Algorithms	Accuracy	Precision	Recall	F1 Score
LR	0.65	0.65	0.65	0.65
kNN	0.69	0.69	0.69	0.69
SVM	0.65	0.66	0.66	0.65
ANN	0.73	0.72	0.72	0.72
CART	0.79	0.79	0.79	0.79
RF	0.75	0.75	0.75	0.75
GBoost	0.71	0.71	0.71	0.71
XGBoost	0.76	0.76	0.76	0.76
LightGBM	0.74	0.74	0.73	0.74
CatBoost	0.73	0.72	0.72	0.72

For assets quality ratios, the most effective financial ratios in determining whether the bank is successful or not in terms of its activity with the feature selection method are (D22) Consumer Loans / Total Loans, (D17) Financial Assets (Net) / Total Assets and (D21) Permanent Assets / Total Assets, respectively. The effect levels of assets quality ratios on bank success are presented in Table 8.

Table 8
The effect levels of assets quality ratios

ID	Ratios	Effect Levels
D22	Consumer Loans / Total Loans	0.3519
D17	Financial Assets (Net) / Total Assets	0.1534
D21	Permanent Assets / Total Assets	0.1435
D19	Total Loans / Total Deposits	0.1395
D20	Loans under follow-up (gross) / Total Loans	0.1118
D18	Total Loans / Total Assets	0.1000

4.2.5 Balance-Sheet Quality

To determine the success of banks in terms of banking activities, 9 financial ratios in the *Balance-Sheet Ratios* group were used.

The success/unsuccessful of banks using Balance-Sheet ratios were predicted by AI algorithms and the performance metrics are given in Table 9. CatBoost showed the highest success level and ANN showed the lowest success level.

Table 9
Performance metrics for Balance-Sheet ratios

Algorithms	Accuracy	Precision	Recall	F1 Score
LR	0.64	0.64	0.64	0.64
kNN	0.57	0.58	0.57	0.56
SVM	0.60	0.60	0.60	0.60
ANN	0.54	0.54	0.54	0.54
CART	0.56	0.56	0.56	0.56
RF	0.63	0.62	0.62	0.62
GBoost	0.62	0.62	0.62	0.62
XGBoost	0.64	0.64	0.63	0.63
LightGBM	0.57	0.57	0.57	0.57
CatBoost	0.65	0.65	0.65	0.65

For Balance-Sheet ratios, the most effective financial ratios in determining whether the bank is successful or not in terms of its activity with the feature selection method are (D28) TC Deposits / Total Deposits, (D30) Total Deposits / Total Assets and (D27) FC Assets / FC Liabilities, respectively. The effect levels of Balance-Sheet ratios on bank success are presented in Table 10.

Table 10
The effect levels of Balance-Sheet ratios

ID	Ratios	Effect Levels
D28	TC Deposits / Total Deposits	0.1621
D30	Total Deposits / Total Assets	0.1558
D27	FC Assets / FC Liabilities	0.1236
D31	Funds Borrowed / Total Assets	0.1216
D29	TC Loans and Receivables / Total Loans and Receivables	0.1136
D23	TC Assets / Total Assets	0.1007
D24	FC Assets / Total Assets	0.0961
D26	FC Liabilities / Total Liabilities	0.0648
D25	TC Liabilities / Total Liabilities	0.0616

4.2.6 Income-Expenditure Structure Quality

To determine the success of banks in terms of banking activities, 12 financial ratios in the *Income-Expenditure Structure Ratios* group were used.

The success/unsuccessful of banks using Income-Expenditure Structure ratios were predicted by AI algorithms and the performance metrics are given in Table 11. XGBoost showed the highest success level and kNN showed the lowest success level.

Table 11
Performance metrics for income-expenditure structure ratios

Algorithms	Accuracy	Precision	Recall	F1 Score
LR	0.81	0.81	0.81	0.81
kNN	0.75	0.77	0.76	0.75
SVM	0.80	0.81	0.80	0.80
ANN	0.78	0.80	0.78	0.77
CART	0.78	0.77	0.77	0.77
RF	0.86	0.87	0.87	0.86
GBoost	0.84	0.84	0.84	0.84
XGBoost	0.86	0.87	0.87	0.86
LightGBM	0.84	0.84	0.84	0.84
CatBoost	0.80	0.81	0.80	0.80

For Income-Expenditure Structure ratios, the most effective financial ratios in determining whether the bank is successful or not in terms of its activity with the feature selection method are (D39) Total Income / Total Expense, (D32) Net Interest Income After Specific Provisions / Total Operating Income and (D36) Other Operating Expenses / Total Operating Income, respectively. The effect levels of Income-Expenditure Structure ratios on bank success are presented in Table 12.

Table 12
The effect levels of income-expenditure structure ratios

ID	Ratios	Effect Levels
D39	Total Income / Total Expense	0.2443
D32	Net Interest Income After Specific Provisions / Total Assets	0.1247
D36	Other Operating Expenses / Total Operating Income	0.0978
D42	Interest Income / Total Expenses	0.0962
D43	Interest Expense / Total Expenses	0.0698
D37	Provision For Loan or Other Receivables Losses / Total Assets	0.0678
D33	Net Interest Income After Specific Provisions / Total Operating Income	0.0676
D38	Interest Income / Interest Expense	0.0636
D35	Non-Interest Income (Net) / Other Operating Expenses	0.0574
D34	Non-Interest Income (Net) / Total Assets	0.0468
D40	Interest Income / Total Assets	0.0337
D39	Total Income / Total Expense	0.2443

4.2.7 Analysis Using All Financial Ratios

Following the analysis of bank success according to financial ratio groups, in this stage of the study, the success of banks in the banking sector in the scope of their activities is analyses by considering all financial ratios at the same time. The success/unsuccessful of banks using All Financial Ratios were predicted by AI algorithms and the performance metrics are given in Table 13. LR showed the highest success level and kNN showed the lowest success level.

Table 13
Performance metrics for All Financial ratios

Algorithms	Accuracy	Precision	Recall	F1 Score
LR	0.91	0.91	0.91	0.91
kNN	0.71	0.74	0.72	0.71
SVM	0.79	0.79	0.79	0.79
ANN	0.79	0.79	0.79	0.79
CART	0.81	0.82	0.82	0.81
RF	0.88	0.87	0.87	0.87
GBoost	0.82	0.83	0.83	0.83
XGBoost	0.90	0.90	0.90	0.90
LightGBM	0.84	0.84	0.84	0.84
CatBoost	0.90	0.90	0.90	0.90

For All Financial ratios, the most effective financial ratios in determining whether the bank is successful or not in terms of its activity with the feature selection method are (D10) Income Before Taxes / Total Assets, (D8) Average Return on Assets and (D9) Average Return on Shareholders' Equity, respectively. The effect levels of All Financial ratios on bank success are presented in Table 14.

Table 14
Performance metrics for All Financial ratios

Ratios ID	Effect Levels	Ratios ID	Effect Levels	Ratios ID	Effect Levels
D10	0.098	D5	0.017	D29	0.013
D8	0.075	D34	0.017	D33	0.012
D9	0.055	D35	0.016	D6	0.012
D2	0.054	D38	0.016	D40	0.011
D39	0.054	D21	0.016	D30	0.011
D11	0.051	D7	0.016	D20	0.011
D4	0.048	D23	0.016	D41	0.011
D3	0.045	D24	0.015	D14	0.010
D36	0.035	D17	0.015	D25	0.010
D42	0.027	D12	0.015	D26	0.010
D22	0.022	D13	0.015	D19	0.009
D32	0.021	D37	0.014	D31	0.008
D15	0.018	D18	0.013	D1	0.008
D16	0.018	D27	0.013		
D43	0.017	D28	0.013		

Although the use of all data in analyzing the success of banks in terms of their activities gives the best results, it is evaluated that the analyses made within the scope of financial ratio grouping can be an important approach in cases where all data sets cannot be available or when the needs or interests of the user differ. The algorithms with the highest accuracy values in financial ratio groups are presented in Table 15.

Table 15
Accuracy of Financial Ratios

Financial Ratios	Algorithms	Accuracy
Capital	LR	0,77
Profitability	XGBoost	0,88
Liquidity	LightGBM	0,68
Assets Quality	CART	0.79
Balance-Sheet	CatBoost	0,65
Income-Expenditure Structure	XGBoost	0.86
All	LR	0.91

5. Conclusions and Recommendations

The banking sector provides financial services for households and enterprises and has a significant role in the economy in terms of consumption, investment and savings. The banking sector, which provides important services for the economy such as facilitating payments, providing financing and supporting advisory services, also has a critical role in infrastructure and private financing investments. The realization of these functions of banks, which have a significant role in the functioning of the modern market economy, is possible only if they have a reliable structure. A reliable structure of banks is important both for the effective continuation of their activities and for ensuring economic stability.

The reliable structure of banks is primarily related to their profitability levels. Banks with low profitability are likely to experience difficulties in securing financing and remain unprotected in conditions of risk and instability. This situation, in addition to the bank's own organizational identity, is closely related to the whole banking sector and thus to the whole economy. A solid structure of the economy, its continued development, and its resistance to risks and uncertainties are only possible through the stability of banks, in other words, through the success of banks. Banks need to be managed effectively in order to improve their profitability levels and achieve a reliable structure. At this point, analyzing the banking sector and the banks operating in the sector is critical for both financial and economic growth and stability.

In this study, whether the banks in the banking sector are successful or not in terms of their operations is analyzed in a broad perspective in terms of financial ratios. In this respect, ML algorithms are used to evaluate the financial performance of 24 selected banks operating in Turkey for the period 2012-2022. In the study, the ratio of net operating profit (loss), which refers to the profits/losses generated by the bank within the scope of its activities, to total assets is taken as the dependent variable. In the period in review, this variable was averaged for each year and banks above the average were classified as successful, while banks below the average were classified as unsuccessful. The independent variables of the study consisted of 43 financial ratios categorized under 6 groups: capital adequacy, profitability, liquidity, asset quality, balance sheet structure, income-expense structure.

For the classification algorithms used in the application, the best parameters were determined by combinations with the Grid search method and the analysis was performed according to these parameters. Accuracy, Precision, Recall and F1 Score metrics were used to measure the prediction success of ML algorithms. The following comments are based on Accuracy. According to the findings, the highest accuracy value in terms of capital ratios was obtained by LR algorithm and the lowest value was obtained by kNN algorithm. In terms of Profitability ratios, the highest accuracy values were found by XGBoost and CatBoost, and the lowest accuracy value was found by kNN algorithm; in terms of Liquidity ratios, the highest value was found by LightGBM, the lowest value was found by

CART algorithm; in terms of Asset Quality ratios, the highest value was found by CART, the lowest value was found by SVM and LR algorithms; in terms of Balance Sheet Structure ratios, the highest value was found by CatBoost, the lowest value was found by ANN algorithm; in terms of Income-Expense Structure, the highest value was found by XGBoost, the lowest value was found by kNN algorithm. In the case that all financial ratios are considered together, the highest accuracy value is obtained with LR, while the lowest value is obtained with kNN algorithm.

While the accuracy values varied between 0.54-0.88 in the analyses conducted for groups, it increased up to 0.91 in the combined analysis where all financial ratios were used together. Although the best results are obtained when all financial ratios are used, in cases where all ratios are not available, it has been observed that analyses by dividing the analysis into groups according to the user's area of interest also provide important findings.

In the study, the effect of financial ratios is also determined using the feature selection method. Firstly, when the findings of the analysis in terms of Capital Adequacy ratios are analyses, it is seen that the most effective financial ratios that determine the success of the bank in terms of its operation among the 7 financial ratios are (D4) Shareholders' Equity / (Deposits + Non-Deposit Funds), (D2) Shareholders' Equity / Total Assets and (D3) (Shareholders' Equity - Permanent Assets) / Total Assets, respectively.

According to the findings of the analysis in terms of profitability ratios are analyses, it is determined that the most effective financial ratios in determining whether the bank is successful or not in terms of its operation among the 4 financial ratios are (D10) Income Before Taxes / Total Assets, (D8) Average Return on Assets and (D9) Average Return on Shareholders’ Equity, respectively. In addition, in the analyses made in terms of liquidity ratios, it was determined that the most effective financial ratios among the 5 financial ratios considered in determining the success of the bank were (D16) FC Liquid Assets / FC Liabilities, (D15) Liquid Assets / (Deposits + Non-Deposit Funds) and (D14) TC Liquid Assets / Total Assets, respectively, the most effective financial ratios among the 6 indicators considered in terms of asset quality ratios were (D22) Consumer Loans / Total Loans, (D17) Financial Assets (Net) / Total Assets and (D21) Permanent Assets / Total Assets, respectively, the most effective financial ratios among the 9 indicators considered in terms of balance sheet structure ratios were (D28) TC Deposits / Total Deposits, (D30) Total Deposits / Total Assets and (D27) FC Assets / FC Liabilities, respectively, and the most effective ratios among the 12 financial ratios considered in terms of income-expense structure ratios were (D39) Total Income / Total Expense, (D32) Net Interest Income After Specific Provisions / Total Operating Income and (D36) Other Operating Expenses / Total Operating Income, respectively.

Finally, the results of the analyses using all financial ratios showed that the most effective financial ratios in determining whether the bank is successful in terms of operation among the 43 financial ratios considered are (D10) Income Before Taxes / Total Assets, (D8) Average Return on Assets and (D9) Average Return on Shareholders’ Equity, respectively.

The findings obtained both from the analyses conducted within the groupings and from the analysis where all financial ratios are considered show that it is extremely important to establish/maintain a solid capital base, optimize profitability, ensure adequate liquidity and manage asset quality for the S (Net Operating Profit (Loss) / Total Assets) ratio, which is considered as a success criterion in bank operations. This also points to criteria that need to be taken into account in order to ensure the long-term sustainability of the bank's operations and to enhance and/or maintain competitiveness. The findings of the study provide a detailed assessment of the success and financial health of banks in their operations, as well as a model proposal to enable stakeholders to make informed decisions and develop strategies for sustainable growth and stability.

Appendix 1. Financial Ratios

Financial Ratios	ID	Ratio	Mean ± Stdev
Capital	D1	Capital Adequacy Ratio	18.19 ± 4.78
	D2	Shareholders' Equity / Total Assets	11.15 ± 3.95
	D3	Shareholders' Equity-Permanent Assets) / Total Assets	7.24 ± 4.47
	D4	Shareholders' Equity / (Deposits + Non-Deposit Funds	13.99 ± 5.89
	D5	On-Balance Sheet Foreign Exchange Position/Shareholders' Equity	75.22 ± 66.26
	D6	Net on Balance-sheet Position / Total Shareholders' Equity	-48.36 ± 59.16
	D7	N (on+off) Balance-sheet Position / Total Shareholders' Equity	4.23 ± 34.98
Profitability	D8	Average Return on Assets	1.39 ± 1.70
	D9	Average Return on Shareholders’ Equity	12.00 ± 16.65
	D10	Income Before Taxes / Total Assets	1.61 ± 1.92
	D11	Net Profit (Losses) / Paid-in Capital	146.83 ± 518.46
Liquidity	D12	Liquid Assets / Total Assets	26.72 ± 12.07
	D13	Liquid Assets / Short-term Liabilities	50.05 ± 26.61
	D14	TC Liquid Assets / Total Assets	10.51 ± 10.68
	D15	Liquid Assets / (Deposits + Non-Deposit Funds)	33.25 ± 15.49
	D16	FC Liquid Assets / FC Liabilities	34.06 ± 12.38
Assets Quality	D17	Financial Assets (Net) / Total Assets	22.95 ± 13.71
	D18	Total Loans / Total Assets	60.06 ± 11.53
	D19	Total Loans / Total Deposits	100.37 ± 28.93
	D20	Loans under follow-up (gross) / Total Loans	4..48 ± 5.19
	D21	Permanent Assets / Total Assets	3.90 ± 2.23
	D22	Consumer Loans / Total Loans	16.37 ± 14.18
Balance-Sheet	D23	TC Assets / Total Assets	59.39 ± 14.17
	D24	FC Assets / Total Assets	40.61 ± 14.17
	D25	TC Liabilities / Total Liabilities	51.88 ± 13.20
	D26	FC Liabilities / Total Liabilities	48.12 ± 13.20
	D27	FC Assets / FC Liabilities	83.83 ± 13.94
	D28	TC Deposits / Total Deposits	51.01 ± 16.13
	D29	TC Loans and Receivables / Total Loans and Receivables	66.07 ± 14.88
	D30	Total Deposits / Total Assets	62.31 ± 11.58
	D31	Funds Borrowed / Total Assets	11.71 ± 11.03
Income-Expenditure Structure	D32	Net Interest Income After Specific Provisions / Total Assets	2.93 ± 2.03
	D33	Net Interest Income After Specific Provisions / Total Operating Income	52.85 ± 66.84
	D34	Noninterest Income (Net) / Total Assets	1.32 ± 1.08
	D35	Noninterest Income (Net) / Other Operating Expenses	81.57 ± 74.06
	D36	Other Operating Expenses / Total Operating Income	43.31 ± 32.18
	D37	Provision For Loan or Other Receivables Losses / Total Assets	1.10 ± 0.99
	D38	Interest Income / Interest Expense	216.35 ± 122.55
	D39	Total Income / Total Expense	157.19 ± 49.84
	D40	Interest Income / Total Assets	8.46 ± 2.44
	D41	Interest Expense / Total Assets	4.49 ± 1.84
	D42	Interest Income / Total Expenses	86.54 ± 9.39
	D43	Interest Expense / Total Expenses	67.11 ± 15.14

Acknowledgement
This research was not funded by any grant.

Conflicts of Interest
The authors declare no conflicts of interest.

References
[1] Assous, H. F. (2022). Prediction of banks efficiency using feature selection method: comparison between selected machine learning models. *Complexity*, 2022(1), 3374489. https://doi.org/10.1155/2022/3374489.
[2] Ozgur, O., Karagol, E. T., & Ozbugday, F. C. (2021). Machine learning approach to drivers of bank lending: evidence from an emerging economy. *Financial Innovation*, 7, 1-29. https://doi.org/10.1186/s40854-021-00237-1.
[3] Bonello, J., BrÉdart, X., & Vella, V. (2018). Machine learning models for predicting financial distress. *Journal of Research in Economics*, 2(2), 174-185. https://doi.org/10.24954/JORE.2018.22.
[4] Halkos, G. E., & Salamouris, D. S. (2004). Efficiency measurement of the Greek commercial banks with the use of financial ratios: A data envelopment analysis approach. *Management accounting research*. 15(2), 201-224. https://doi.org/10.1016/j.mar.2004.02.001.
[5] Wu, D. D., Yang, Z., & Liang, L. (2006). Using DEA-neural network approach to evaluate branch efficiency of a large Canadian bank. *Expert systems with applications*, 31(1), 108-115. https://doi.org/10.1016/j.eswa.2005.09.034.
[6] Nur Ozkan-Gunay, E., & Ozkan, M. (2007). Prediction of bank failures in emerging financial markets: an ANN approach. The Journal of Risk Finance, 8(5), 465-480. https://doi.org/10.1108/15265940710834753.
[7] Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. *Risks*, 7(1), 29. https://doi.org/10.3390/risks7010029.
[8] Malali, A. B., & Gopalakrishnan, S. (2020). Application of artificial intelligence and its powered technologies in the Indian banking and financial industry: An overview. *IOSR Journal of Humanities and Social Science*, 25(4), 55-60.
[9] Königstorfer, F., & Thalmann, S. (2020). Applications of Artificial Intelligence in commercial banks–A research agenda for behavioral finance. *Journal of behavioral and experimental finance*, 27, 100352. https://doi.org/10.1016/j.jbef.2020.100352.
[10] Appiahene, P., Missah, Y. M., & Najim, U. (2020). Predicting bank operational efficiency using machine learning algorithm: comparative study of decision tree, random forest, and neural networks. Advances in fuzzy systems, 2020(1), 8581202. https://doi.org/10.1155/2020/8581202.
[11] Fares, O. H., Butt, I., & Lee, S. H. M. (2023). Utilization of artificial intelligence in the banking sector: A systematic literature review. *Journal of Financial Services Marketing*, 28(4), 835-852. https://doi.org/10.1057/s41264-022-00176-7.
[12] Umamaheswari, S., & Valarmathi, A. (2023). Role of artificial intelligence in the banking sector. *Journal of Survey in Fisheries Sciences*, 10(4S), 2841-2849. https://doi.org/10.17762/sfs.v10i4S.1722.
[13] Doumpos, M., Zopounidis, C., Gounopoulos, D., Platanakis, E., & Zhang, W. (2023). Operational research and artificial intelligence methods in banking. *European Journal of Operational Research*, 306(1), 1-16. https://doi.org/10.1016/j.ejor.2022.04.027.
[14] Gangwani, D., & Zhu, X. (2024). Modeling and prediction of business success: a survey. *Artificial Intelligence Review* 57(44), 1-51. https://doi.org/10.1007/s10462-023-10664-4.
[15] Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson.
[16] Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). Machine learning: An artificial intelligence approach. Springer Science & Business Media.
[17] Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard business review*, 96(1), 108-116.
[18] Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46-60. https://doi.org/10.1016/j.futures.2017.03.006.
[19] Javaid, M., Haleem, A., Singh, R. P., Rab, S., & Suman, R. (2022). Exploring impact and features of machine vision for progressive industry 4.0 culture. *Sensors International*, 3, 100132. https://doi.org/10.1016/j.sintl.2021.100132.
[20] Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. WW Norton & Company.
[21] Davenport, T., & Harris, J. (2017). Competing on analytics: Updated, with a new introduction: The new science of winning. Harvard Business Press.

[22] Li, C. (2010). Groundswell. Winning in a world transformed by social technologies. *Strategic Direction*, 26(8), 136138. https://doi.org/10.1108/sd.2010.05626hae.002.
[23] Porter, M. E., & Heppelmann, J. E. (2014). How smart, connected products are transforming competition. *Harvard business review*, 92(11), 64-88.
[24] Shukla Shubhendu, S., & Vijay, J. (2013). Applicability of artificial intelligence in different fields of life. International Journal of Scientific Engineering and Research, 1(1), 28-35. https://doi.org/10.70729/1130915.
[25] Duda, R. O., & Hart, P. E. (2006). Pattern classification. John Wiley & Sons.
[26] Robu, R., & Holban, S. (2011, May). A genetic algorithm for classification. In Recent Researches in Computers and Computing-International Conference on Computers and Computing, ICCC (Vol. 11, No. 05).
[27] Bao, J., Zheng, Y., Wilkie, D., & Mokbel, M. (2015). Recommendations in location-based social networks: a survey. *GeoInformatica*, 19, 525-565. https://doi.org/10.1007/s10707-014-0220-8.
[28] Cummins, S., W. Peltier, J., A. Schibrowsky, J., & Nill, A. (2014). Consumer behavior in the online context. *Journal of Research in Interactive Marketing*, 8(3), 169-202. https://doi.org/10.1108/JRIM-04-2013-0019.
[29] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, 31(8), 651-666. https://doi.org/10.1016/j.patrec.2009.09.011.
[30] Maulud, D., & Abdulazeez, A. M. (2020). A review on linear regression comprehensive in machine learning. *Journal of Applied Science and Technology Trends*, 1(2), 140-147.
[31] Theissler, A., Spinnato, F., Schlegel, U., & Guidotti, R. (2022). Explainable AI for time series classification: a review, taxonomy and research directions. *IEEE Access*, 10, 100700-100724. https://doi.org/10.1109/ACCESS.2022.3207765
[32] Nocedal, J., & Wright, S. J. (1999). Numerical optimization. New York: NY: Springer New York.
[33] Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons.
[34] Hellevik, O. (2009). Linear versus logistic regression when the dependent variable is a dichotomy. *Quality & Quantity*, 43, 59-74. https://doi.org/10.1007/s11135-007-9077-3.
[35] Dhanabal, S., & Chandramathi, S. (2011). A review of various k-nearest neighbor query processing techniques. *International Journal of Computer Applications*, 31(7), 14-22. https://doi.org/10.5120/3836-5332.
[36] Mucherino, A., Papajorgji, P. J., Pardalos, P. M., Mucherino, A., Papajorgji, P. J., & Pardalos, P. M. (2009). K-nearest neighbor classification. *Data mining in agriculture*, 83-106. https://doi.org/10.1007/978-0-387-88615-2_4.
[37] Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.
[38] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. http://dx.doi.org/10.1038/nature14539.
[39] Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117. https://doi.org/10.48550/arXiv.1404.7828.
[40] Shu, X., & Ye, Y. (2023). Knowledge Discovery: Methods from data mining and machine learning. *Social Science Research*, 110, 102817. https://doi.org/10.1016/j.ssresearch.2022.102817.
[41] Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: springer.
[42] Breiman, L. (2017). Classification and regression trees. Routledge.
[43] Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. *Journal of Hydrology*, 598, 126266. https://doi.org/10.1016/j.jhydrol.2021.126266.
[44] Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues (IJCSI*), 9(5), 272.
[45] Liaw, A., & Wiener, M. (2002). Classification and regression by Random Forest. *R news*, 2(3), 18-22.
[46] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
[47] Feng, Z., Xu, C., & Tao, D. (2018, September). Historical Gradient Boosting Machine. In GCAI (pp. 68-80).
[48] Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7, 1-21. https://doi.org/10.3389/fnbot.2013.00021.
[49] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
[50] Ramraj, S., Uzir, N., Sunil, R., & Banerjee, S. (2016). Experimenting XGBoost algorithm for prediction and classification of different datasets. *International Journal of Control Theory and Applications*, 9(40), 651-662.
[51] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.

[52] Wang, D. N., Li, L., & Zhao, D. (2022). Corporate finance risk prediction based on LightGBM. *Information Sciences*, 602, 259-268. https://doi.org/10.1016/j.ins.2022.04.058.

[53] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, 1-23. https://doi.org/10.48550/arXiv.1706.09516.

[54] Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.

[55] Albulescu, C. T. (2015). Banks' profitability and financial soundness indicators: A macro-level investigation in emerging countries.*Procedia economics and finance*,23,203-209.https://doi.org/10.1016/S2212-5671(15)00551-1.

[56] Sagita, B., & Nugraha, N. (2022, July). Does Liquidity or Profitability Influence Firm Financial Distress Most? Empirical Study on Manufacturing Companies Listed in Indonesia Stock Exchange (2015-2019). In 6th Global Conference on Business, Management, and Entrepreneurship (GCBME 2021) (pp. 51-56). Atlantis Press.

[57] Aupperle, K. E., & Sarhan, M. H. (1995). Assessing financial performance in the capstone, strategic management course: A proposed template. *Journal of Education for Business*, 71(2), 72-77. https://doi.org/10.1177/2379298120909513.