Content list available at JournalPark

Turkish Journal of Forecasting

Comparison of Machine Learning Algorithms for Predicting Financial Risk in Cash Flow Statements

Ecem ENGIN[1], Damla ILTER FAKHOURI[1,*]

[1] *Mimar Sinan Fine Arts University, Faculty of Science and Letters, Department of Statistics, Bomonti Campus, Istanbul, Turkey*

ARTICLE INFO

Article history:

Received	12	December	2023
Revision	05	February	2024
Accepted	26	February	2024
Available online	01	April	2024

Keywords:
Credit Classification
Machine Learning Algorithms
Cash Flow Statements
Risk
İnformation Criteria

RESEARCH ARTICLE

ABSTRACT

Nowadays, making financial decisions and evaluating loan applications is a complex and sensitive process. Cash flow data, which shows the financial risk status of businesses, plays a key role in evaluating loan applications. Cash flow data, which shows the financial risk status of businesses, plays a key role in evaluating loan applications. Guiding business managers in making strategic decisions and managing financial risks, quarterly data provides a detailed timeline of business performance and helps identify seasonal changes. A detailed analysis using machine learning algorithms evaluates the performance of different models built to compare businesses quarters in the loan classification process and highlights the role of cash flow data in the process. It was aimed to create effective algorithms by taking into account the suitability of the quarterly data between 2018 and 2022 of the 282 companies used in the study, and to provide a unique approach in the field of evaluating these algorithms with information criteria. The model performances of the quarters are very close to each other and a high success rate is obtained. Therefore, it was observed that quarterly periods did not make a significant difference in model performance. The model created for the 2nd quarter of 2019 was selected as the best model with 99% accuracy and 99% F1 value. It was also determined that the selection of variables with high accuracy rates in the models established for each quarter is important in terms of predicting financial risk.

1. Introduction

Increasingly complex credit classification processes in financial sector play a key role in the process of predicting the risks of companies. This process provides significant advantages for both financial institutions and customers by aiming to minimize the predictable risk by evaluating loan applications. Loan classification is carried out taking into account various criteria such as the applicants' financial history, income level, payment habits, etc. These evaluations lead to critical decisions on whether or not to accept loan requests or under what conditions loans can be granted. For all these reasons, cash flow data is important for financial institutions in the loan classification process. Cash flow data contains valuable information about critical factors such as payment habits, income level and debt repayment ability of businesses. These data, which reflect the cash movements of companies in real life, show their financial situation in detail. The statement of cash flows provides information about the cash position of the company in a certain period and helps the company to make decisions about its future situation [1]. The inclusion of this data in the credit classification process helps financial institutions to determine the risk profiles of companies more effectively and accurately.

Quarterly data provides a detailed timeline of business performance and plays an important role in identifying seasonal variations. It also guides business managers in making strategic decisions and managing financial risks. In this context, the study on credit classification models emphasises that the use of quarterly data is critical in understanding and evaluating business performance.

In this study, a more comprehensive analysis was performed by comparing the results obtained by using machine learning algorithms that gradient-based and suitable for classification. This method allowed for a more detailed analysis of the data and more precise results in the loan classification process. For the analysis, the cash flow statement data of the companies on the website of "Is Yatirim" were used. Among the companies, financial sector organizations (banks, private finance institutions, insurance companies, etc.) and conglomerates were excluded from the sample due to their very different operational and financial characteristics [2]. To determine the dependent variable, financial failure prediction model studies are analysed. Failure criteria are especially important for research results to be useful [3]. Karataş and Can showed that, the dependent variable was determined by choosing among the specified criteria in their study. [4]. A basic framework for machine learning algorithms applied to econometric data is inspired by Sadeghzadeh and Elmas [5].

This comprehensive analysis evaluates the performance of different models used in loan classification processes and is expected to provide an important perspective on the role of cash flow statement data in this process. Also, the comparison of quarterly periods over the models provides a valuable perspective to understand the changes in business performance over time and to determine the effects of these models.

2. Motivation and Overview

The use of machine learning algorithms in the financial sector plays a key role in data analysis and loan classification processes. This study, which examines machine learning algorithms that are gradient-based and suitable for classification, discusses traditional and innovative approaches to credit risk assessment. The impact of machine learning algorithms that have been shown to be of great importance in achieving high reliability and accurate prediction in terms of financial decisions such as Gradient Boosting, Extreme Gradient Boosting, Light Gradient Boosting and traditional methods such as Decision Tree, Extra Trees and Random Forest Classifier on the cash flow statement data is also analysed in detail.

Machine learning algorithms perform the learning process by detecting patterns and differences in the dataset under the independent of assumptions [6]. These algorithms, provide a foundation for more well-founded, reliable and highly accurate predict in financial decision-making processes. Therefore, this key step perform before model building increases the efficiency and reliability of loan classification processes in the financial sector. Studies show that these techniques are effective in achieving successful results in credit risk assessments and make significant contributions to financial decision-making processes [6,7,8,9,10,12,13,14,16].

In the modelling process, variable selection is critical in predict of the dependent variable. Lasso regression, linear and logistic regression analyses are variable selection techniques [11]. In this way, only the important variables remain in the model and the predicted risk can be explained depending on the variable.

Imbalanced data, which is a communal problem in financial data, negatively affects the accurate of the results after the model is built. Imbalanced data problems are one of the most critical issues to be considered in the literature when modelling. Classification algorithms assume that the training sets obtained with partition ratio are well balanced, with results generated with high accuracy ratio. This assumed balanced distribution is not usually found in clusters of real-life data, especially financial data. Representational capabilities can be defined by a large number of examples in one class and few examples in another class. In such cases, classification problems are highly likely to manifest themselves. There are many methods developed to overcome these problems. To improve the accuracy of the developed approaches, this study uses under sampling, oversampling, SMOTE approaches as well as sampling techniques such as clustering-based sampling and boosting. Under sampling aims to rebalance the dataset by removing instances of the majority class until the class distributions are equal. If the reduction in the majority class is done randomly, it is called Random Under sampling, and if it is done using statistical information, it is called

Informed Under sampling [12]. Oversampling multiplies the samples of the minority class until equal class distributions are obtained [13]. SMOTE is an oversampling process that enables the generation of synthetic data [14].

Some evaluation metrics are used to evaluate the models created with classification algorithms and to determine which classification model produces more accurate results [15]. The metrics used to evaluate model performance play a critical role in determining the reliability and applicability of the results obtained. In the study, model performance was evaluated in detail using various metrics such as Accuracy, AUC (Area Under the Curve), Recall, Precision, F1 score, MAE (Mean Absolute Error), MSE (Mean Square Error), AIC (Akaike Information Criterion) and BIC (Bayes Information Criterion). These metrics were chosen to evaluate the effectiveness of machine learning algorithms used in loan classification processes in the financial market, as well as to assess their suitability for specific financial objectives and to make the right decisions.

The confusion matrix is a metric used to evaluate the performance of a classification model. It is often used to visualize the model's true and false predictions. This matrix, used to understand the relationship between two classes (e.g. approved and rejected loan applications), contains four key terms:

- True Positive: Refers to situations where the model successfully detects the positive class. For example, if the loan application was approved and the model correctly predicted this, this is considered a TP.

- True Negative: Refers to cases where the model successfully detected the negative class. For example, if the loan application was negative and the model correctly predicted this, this is considered TN.

- False Positive: Refers to situations that the model predicts as positive class but are actually negative. For example, if the model predicts a loan application to be positive but it is actually negative, this is called FP.

- False Negative: Refers to situations where the model predicts a negative class but is actually positive. For example, if the model predicts a loan application to be negative but it is actually positive, this is called FN.

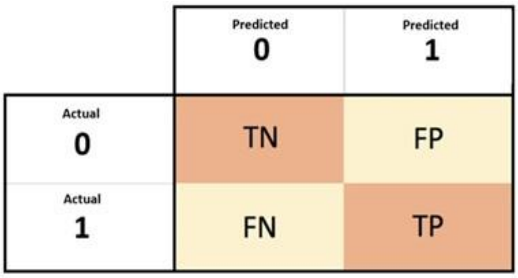


Figure 1. Confusion Matrix

Figure 1 shows the confusion matrix used to evaluate the performance of the classification model. There is a limited number of studies in the literature on loan classification with cash flow statement. In this respect, Altan and Demirci's study, which stands out as one of the few studies in the literature, examines the effectiveness of the XGBoost method for credit scoring based on cash flow statements in detail and reveals that this method exhibits superior performance compared to other algorithms. According to the results of the study, the XGBoost method was determined as the most successful model with an accuracy score of 80% and an AUC of 82%. These results show that XGBoost performs more effectively than other machine learning algorithms in the field of credit scoring [16].

A literature review on credit classification was performed. In Tütüncü and Gursakal's study evaluates the effectiveness of five different machine learning algorithms for predicting credit default risk: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting and Artificial Neural Network [17]. According to the results of the study, the Gradient Boosting algorithm has the highest accuracy rate of 91.5%. The accuracy ratio of the other algorithms are different between 85.5% and 90.5%. In Can and Gurhanli's study, while the success ratio of the random forest algorithm was 73.60%, the XGBoost algorithm was found to be the most successful application in predicting the credit eligibility of customers with a success rate of 75.60% [18]. In Aithal and Jathanna's study, the performance of different machine learning algorithms was evaluated [19]. The Random Forest algorithm achieved the highest accuracy rate (78%), the highest recall rate (93%), the highest precision rate (91%) and the highest F1-Score (85%). In Ilter and Kocadagli' s study, machine learning algorithms used in credit scoring were compared and it was found that SVM (RBF) algorithm outperformed other algorithms in terms of AUC, KS and Gini Index [20].

There are no studies found in the literature on the methods applied in the case of cash flow imbalanced data. In Erdem and Bakir's study, XGBoost model was applied using under sampling, over sampling and SMOTE methods to predict financial failure. According to the evaluation, the sampling method with the highest prediction success was determined as the SMOTE method with a success rate of 78% [21]. In the study of Kardes and Kandemir, artificial neural network and logistic regression models were applied by applying smote to predict the independent audit opinions of the companies in Borsa Istanbul. While Artificial Neural Network showed 96.5% and Logistic Regression 94.3% accurate for the classification performance [22].

A literature review of machine learning algorithms has been conducted. These algorithms have a wide range of applications in various sectors. In Hamal's study, Support Vector Machines, Naive Bayes, Artificial Neural Networks, K-Nearest Neighborhood, Random Forest, Logistic Regression and Bagging classification algorithms are compared in the prediction of business that do and do not engage in financial accounting fraud [23]. In Okochan's study, he compared Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting, Extreme Gradient Boosting and Light Gradient Boosting algorithms to predict customer churn in the banking sector [24]. In Uyanık's study, it is aimed to predict the customers who are likely to terminate their subscription in the telecommunication sector. In this study, Logistic Regression, Decision Tree, Artificial Neural Networks, Bagging and Boosting classification models were compared [25]. In Zengin's study, the attitudes of individuals over the age of eighteen residing in Istanbul towards digital banking applications, user groups and usage habits, socioeconomic and demographic aspects are examined with machine learning methods. Categorical Boosting, Lightweight Gradient Boosting Machines, Extreme Gradient Boosting, Gradient Boosting, Adaptive Boosting, K Nearest Neighbor, Decision Trees, Random Forest, Logistic Regression and Radial Basis Function Kernel Support Vector Machines classification algorithms were compared [26]. In Çalışkan's study, Logistic Regression, Linear Discriminant Analysis, Decision Tree, Nearest Neighbor Algorithm, Navie Bayes, Support Vector Machines, Adaboost Classifier, Gradient Boosting Classifier, Random Forest Classifier and Extra Tree Classifier algorithms were compared in order to diagnose epilepsy [27]. In Kılıckap's study, the problem of determining the appropriate package type for orders in the packaging process, which has an important place in terms of cost and efficiency in logistics sectors, with the help of machine learning models is emphasized. Logistic Regression, Support Vector Machines, Decision Trees, K-Nearest Neighbor Classifier, Random Forest Classifier, Gaussian Naive Bayes Classifier and XGBoost algorithms were used in the study [28].

3. Algorithms and Structures Used in Modelling

Random Forest is an ensemble learning algorithm based on decision trees. Its main goal is to create a more powerful and generalizable classification model by combining decision trees that are different from each other and are weak learners. Random Forest takes random samples from a dataset and trains many decision trees on these samples, using the bagging technique.

$$RF(x) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T} f_t(x) \\ mode(f_1(x), f_2(x), \dots, f_T(x)) \end{cases} \tag{1}$$

Equation (1) shows the mathematical equation of the random forest algorithm. Where T is the number of decision trees, and $f_t(x)$ is the prediction of the t-th decision tree for the input x [29].

Gradient Boosting Classifier, is a machine learning algorithm that builds a strong ensemble model by combining weak classifiers. In the first step, an initial model is determined on the dataset and the errors generated by the predictions of this model are calculated. Then, weak classifiers (usually decision trees) that focus on these errors are iteratively added and each one is trained to correct the errors of the previous models. The generated weak classifiers are weighted, and this process ends after a certain number of iterations or when the errors are minimized. Gradient Boosting Classifier aims to achieve high accuracy classifications by modeming complex relationships in the dataset using this incremental approach.

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{2}$$

Equation (2) shows the mathematical equation of the gradient boosting algorithm. Where $F_m(x)$ is the model's prediction at iteration *m*, $F_{m-1}(x)$ is the prediction at iteration $m - 1$, y_m is the learning rate for iteration *m*, and $h_m(x)$ is the weak model at iteration m [29].

Extreme Gradient Boosting is a powerful learning algorithm used to achieve high performance, especially on structured data sets. It is essentially an optimized version of Gradient Boosting, a tree-based modelling technique. It can make fast and efficient predictions using parallel computations and a specialized tree growth algorithm. It can successfully perform both regression and classification tasks and model complex relationships by combining tree-based models.

$$\mathcal{L}\,(\theta) = \sum_{i=1}^{n} l\,(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega\,(f_k) \tag{3}$$

Equation (3) shows the mathematical equation of the extreme gradient boosting algorithm. Where θ represents the model parameters, l is the loss function, y_i is the actual value, $\hat{y}_i$ is the predicted value, K is the number of trees, and $O(f_k)$ is a regularization term that penalizes complex models. The main goal is to accurately predict the target variable based on a large number of features [29].

Light Gradient Boosting is a fast and efficient gradient boosting algorithm for large data sets. It is an extended and optimized version of the Gradient Boosting algorithm. It utilizes tree-based learning methods to provide performance advantages, especially on large datasets. Improves overall model performance by building deep trees with fewer nodes with a "leaf-wise" growth strategy. It optimizes memory usage and offers a wide range of applications for classification/regression problems, especially in big data analysis.

Decision Tree Classifier represents the data set in a tree structure with a set of decision rules and finalization nodes. Each internal node represents a test on a feature, and each leaf node contains data points that belong to a particular class. The Decision Tree Classifier identifies and classifies the interactions and important relationships between features in the dataset. By creating a comprehensible decision tree, this method allows the user to understand and interpret the model's prediction process. Furthermore, Decision Tree Classifier has a flexible structure that can be controlled by parameters to avoid problems such as overfitting, which allows it to be successfully used in various application areas.

$$G = 1 - \sum_{i=1}^{K} p_i^2 \tag{4}$$

Equation (4) shows the mathematical equation of the decision tree classifier algorithm. K is the number of classes, and p_i is the proportion of samples that belong to the i-th class in the subset of data at that node [29].

Extra Trees Classifier is a tree-based learning algorithm used in classification tasks. This algorithm has a structure similar to Random Forest, but adds more randomness when building the trees. Extra Trees provides diversity by making more random choices when determining the decision points when building each tree. This feature can increase the generalization ability of the model and reduce the risk of overfitting. Extra Trees Classifier can be effectively used in high-dimensional data sets, noisy data and complex classification problems.

$$F(x) = \sum_{i=1}^{n} f_i(x) \tag{5}$$

Equation (5) shows the mathematical equation of the extra trees classifier algorithm. Where F(x) is the model's prediction for a new input, n is the number of trees in the ensemble, and $f_i(x)$ is the prediction of the *i*th decision tree in the ensemble [29].

Naive Bayes works based on Bayes' Theorem and is called "naive" for its assumptions in the classification process. Basically, Naive Bayes uses the assumption of independence between features when estimating the probability that a document or sample data point belongs to a class. The algorithm stands out for its fast and low computational cost in the classification process, but it can ignore real-world dependencies between features. Therefore, it can be an effective choice for simple and fast classification problems.

$$P(X|Y) = (P(Y|X)xP(X))/P(Y) \tag{6}$$

Equation (6) shows the mathematical equation of the decision tree classifier algorithm. If X, Y = probabilistic events, P (X) = probability of X being true, P(X|Y) = conditional probability of X being true in case Y is true.

4. Application

4.1. Data Set and Procedure

The dataset consists of 282 firms and 5,640 observations, covering quarterly periods between 2018 and 2022. In the study, the dependent variable is defined as the firms that are not eligible for loan applications (0) and those that are eligible (1) among the firms in the "Detention Market" among the BIST companies in the "Public Disclosure Platform". The independent variables consist of 26 items of the cash flow statement. Sampling techniques such as oversampling, undersampling and SMOTE were used for unbalanced distribution of the dependent variable. In the model with a large number of independent variables, lasso (L1 arrangement) method was applied to select features using logistic regression. In the analysis process, to control overfitting and complexity, the credit classification data were separated into training, test, and validation, and then a detailed analysis was performed by applying the cross-validation method to the train data. This cross-validation using k fold was split on a company-by-company basis and ensured that companies in each quarter were fairly represented in the training and validation sets. To understand the change in business performance over time and to determine the effects of these models, 20 different models were created for quarterly periods and model selection was made according to information criteria. In the obtained models, the effects of independent variables on the dependent variable are determined.

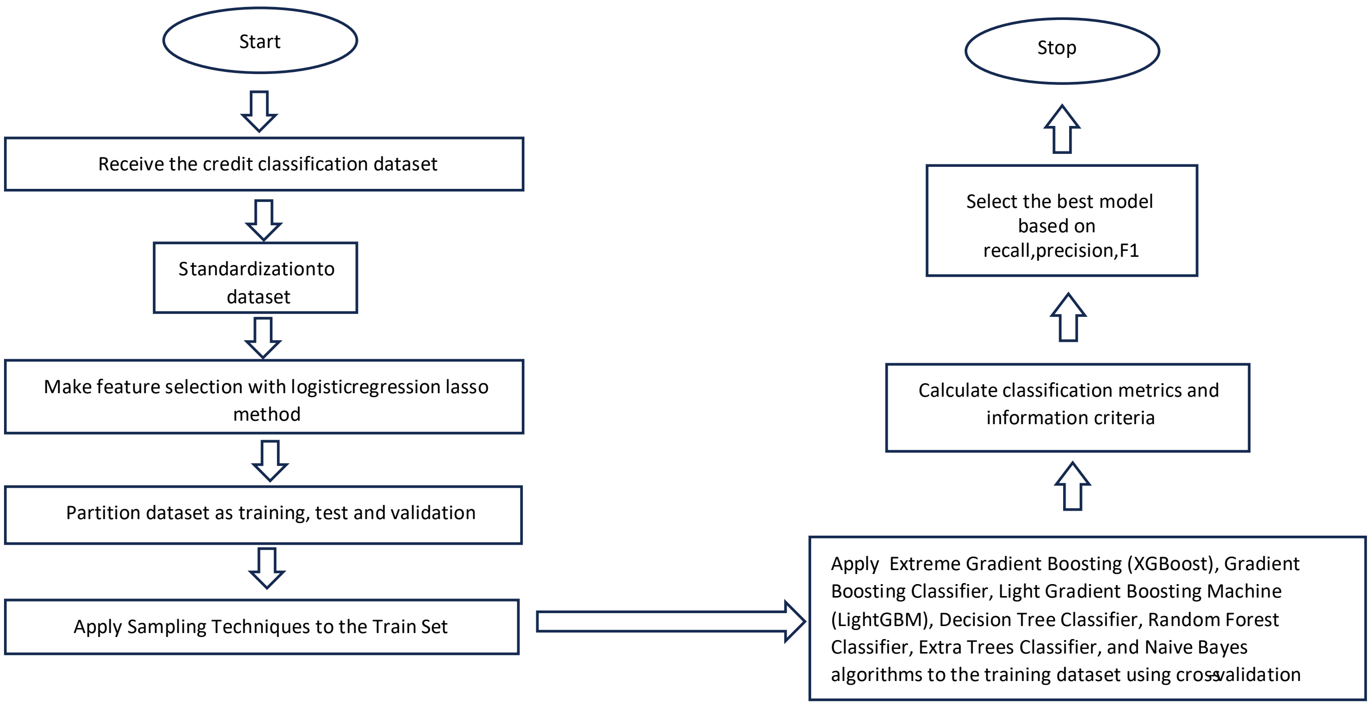


Figure 2. The flowchart of the procedure used

Figure 2 shows the procedure applied in the analysis.

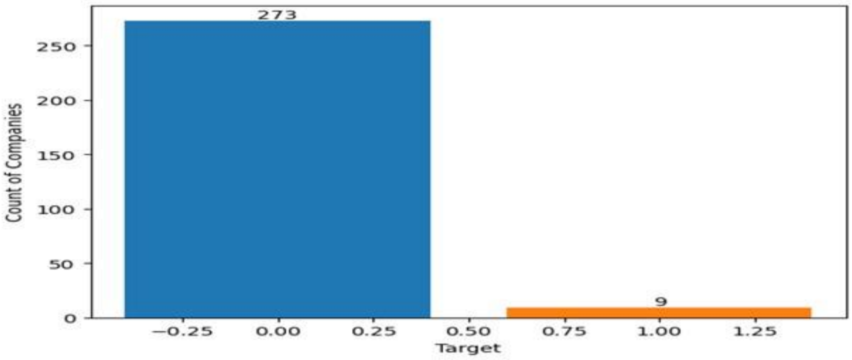


Figure 3. The unique distribution of the dependent variable

Since there are 20 observations for each company in the dataset, Figure 3 shows the unique distribution of the dependent variable across companies. Out of 282 enterprises, 9 were determined as "not suitable for credit application" and 273 were determined as "suitable for credit application".

4.2. Analysis

In the first stage of the analysis, standardisation was applied based on companies in order to balance the differences arising from wide ranges in continuous variables and to make the data more consistent. This pre-processing step ensured that the data set was more properly prepared for the analysis process. In the next stage, feature selection was performed using logistic regression lasso arrangement and 5 of the 26 continuous variables in the data set were eliminated.

Table 1. Independent Variables Used in the Models

Net Cash Generated from Operating Activities
Profit Before Adjustments
Adjustments:
Depreciation & Amortization
Change in Provisions
Other Income/Expense
Operating Profit Before Changes in Operating Working Capital (+)
Changes in Operating Working Capital
Cash from Other Operating Activities
Fixed Capital Investments
Cash from Other Investing Activities
Cash Generated from Investing Activities
Change in Financial Debts
Dividend Payments
Capital Increases
Cash from Other Financing Activities
Effect of Foreign Exchange Rate Changes on Cash and Cash Equivalents Before Adjustment
Foreign Exchange Rate Effect on Cash and Cash Equivalents
Change in Cash and Cash Equivalents
Cash at the Beginning of the Period
Cash at the End of the Period

Table 1 shows the variables selected by logistic regression lasso method. These variables were used as independent variables in the applied models. The data set is divided into 70% training, 15% testing and 15% validation. Due to the imbalance in the class distribution of the dependent variable, undersampling, oversampling and Synthetic Minority Over-sampling Technique (SMOTE) were applied to the training data set. These methods were used to correct the class imbalance and to train the models in a more balanced way.

Table 2. Class Imbalance Sampling Methods and Sample Distribution

Data	Number of Samples	Number of Negatives	Number of Positives
Original	3.948	126	3.822
Under-Sampled	252	126	126

Over-Sampled	7.644	3.822	3.822
SMOTE	7.644	3.822	3.822

Table 2 shows the different sampling methods applied on the train data and the total number of samples in the original data.

In the modelling phase, a total of 20 different models were created by considering each quarter separately in order to make a detailed analysis on quarterly data. In addition to the independent variables, a dummy variable including the quarter of that year is included in each model. This method, was applied in the same way for all sampling techniques applied to the dataset. This approach allowed us to assess the performance of companies on a quarterly basis and to understand the independent effects of each quarter. Firstly, the models were evaluated for each sampling method and the best performing model was determined. Then, the models selected in each sampling method were evaluated in detail. The best performing model was selected for each quarter. In unbalanced data sets, the accuracy metric may be insufficient due to significant proportional differences between classes. Therefore, the models were selected taking into account the criteria of recall, precision, F1 and low MSE.

Table 3. Models applied in the original data for the 4th quarter of 2022

Model	Accuracy	Recall	Precision	F1	MSE	Time(sec.)
Random Forest	0,97036	0,07143	1,00000	0,12914	0,02964	2,10372
Gradient Boosting Classifier	0,99063	0,73016	0,97059	0,82986	0,00937	7,95613
Extreme Gradient Boosting	0,98506	0,53175	1,00000	0,69218	0,01494	0,45944
Light Gradient Boosting Machine	**0,99088**	**0,72222**	**0,99048**	**0,83302**	**0,00912**	**0,21442**
Decision Tree Classifier	0,96910	0,20635	0,56349	0,29834	0,03090	0,06217
Extra Trees Classifier	0,97366	0,17460	1,00000	0,28999	0,02634	0,77227
Naive Bayes	0,93972	0,07937	0,07934	0,07883	0,06028	0,03760

Table 4. Models applied to undersampling data for the 4th quarter of 2022

Model	Accuracy	Recall	Precision	F1	MSE	Time(sec.)
Random Forest	0,88492	0,90476	0,87028	0,88647	0,11508	0,43585
Gradient Boosting Classifier	**0,89286**	**0,92857**	**0,86695**	**0,89663**	**0,10714**	**0,49897**
Extreme Gradient Boosting	0,86508	0,86508	0,86560	0,86526	0,13492	0,26227
Light Gradient Boosting Machine	0,87302	0,85714	0,88532	0,87089	0,12698	0,13596
Decision Tree Classifier	0,69048	0,65079	0,70621	0,67659	0,30952	0,04723
Extra Trees Classifier	0,83730	0,84921	0,82845	0,83715	0,16270	0,34939
Naive Bayes	0,46825	0,39683	0,44555	0,40036	0,53175	0,04158

Table 5. Models applied to oversampling data for the 4th quarter of 2022

Model	Accuracy	Recall	Precision	F1	MSE	Time(sec.)
Random Forest	0,99634	1,00000	0,99273	0,99635	0,00366	2,25055
Gradient Boosting Classifier	**0,99922**	**1,00000**	**0,99843**	**0,99922**	**0,00078**	**9,92559**
Extreme Gradient Boosting	0,99791	1,00000	0,99583	0,99791	0,00209	0,71808
Light Gradient Boosting Machine	0,99856	1,00000	0,99713	0,99856	0,00144	0,27424
Decision Tree Classifier	0,97684	1,00000	0,95588	0,97741	0,02316	0,06917
Extra Trees Classifier	0,97501	0,98299	0,96758	0,97520	0,02499	0,82480
Naive Bayes	0,59079	0,51832	0,61196	0,55145	0,40921	0,05452

Table 6. Models applied to smote data for the 4th quarter of 2022

Model	Accuracy	Recall	Precision	F1	MSE	Time(sec.)
Random Forest	0,98731	0,99686	0,97818	0,98743	0,01269	4,25266
Gradient Boosting Classifier	**0,99320**	**0,99712**	**0,98936**	**0,99322**	**0,00680**	**16,35618**
Extreme Gradient Boosting	0,99137	0,99634	0,98653	0,99141	0,00863	1,09574
Light Gradient Boosting Machine	0,99215	0,99765	0,98680	0,99219	0,00785	0,55150
Decision Tree Classifier	0,91222	0,95107	0,88339	0,91560	0,08778	0,28093
Extra Trees Classifier	0,94846	0,98064	0,92140	0,95007	0,05154	1,02194
Naive Bayes	0,58137	0,86054	0,55230	0,67275	0,41863	0,19680

The results of the models applied for the 4th quarter of 2022 are given between Table 3 and Table 6. Gradient based algorithms have attracted attention by exhibiting high performance both on the original data and on the data to which sampling techniques have been applied. It stands out for its overall accuracy, recall, precision and F1 scores. While the Naive Bayes algorithm exhibits low performance in general, it is the fastest algorithm. Furthermore, tree-based algorithms (Random Forest, Decision Tree, Extra Trees) show a significant performance improvement when sampling techniques such as oversampling and SMOTE are applied. It is observed that MSE is high for each model in undersampling data. This suggests that undersampling leads to information loss due to imbalance in the dataset and affects the accuracy of the models.

Table 7. Best model selection for the 4th quarter of 2022

Data	Model	Accuracy	Recall	Precision	F1	MSE
Orjinal Data	**Light Gradient Boosting Machine**	0,99088	0,72222	0,99048	0,83302	0,00912
Undersampling	**Gradient Boosting Classifier**	0,89286	0,92857	0,86695	0,89663	0,10714
Oversampling	**Gradient Boosting Classifier**	0,99922	1,00000	0,99843	0,99922	0,00078
Smote	**Gradient Boosting Classifier**	0,99320	0,99712	0,98936	0,99322	0,00680

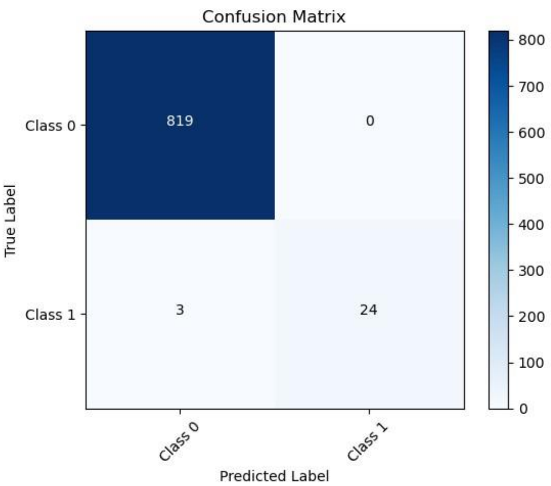
The models applied for each sampling technique and original data were evaluated and the best models were determined according to the criteria of high recall and low mean squared error. Table 7 shows the selected algorithms and the results. Gradient-based algorithms were selected as the most successful algorithms. In particular, the Light Gradient Boosting Machine algorithm stood out with high recall values and low MSE on the original data. In addition, the Gradient Boosting algorithm performed better on data with sampling techniques applied.

The same procedure was applied to evaluate the other quarter models and the best models were selected.

Table 8. Best Model Selection

Quarterly	Model	Accuracy	Recall	Precision	F1	MSE	Time(sec.)
Q4 2022	Gradient Boosting Classifier	0,99922	1,00000	0,99843	0,99922	0,00078	9,92559
Q3 2022	Gradient Boosting Classifier	0,99895	1,00000	0,99791	0,99895	0,00105	7,62514
Q2 2022	Gradient Boosting Classifier	0,99935	1,00000	0,99869	0,99935	0,00065	9,73490
Q1 2022	Gradient Boosting Classifier	0,99908	1,00000	0,99817	0,99909	0,00092	8,74177
Q4 2021	Gradient Boosting Classifier	0,99895	1,00000	0,99791	0,99895	0,00105	6,53182
Q3 2021	Gradient Boosting Classifier	0,99908	1,00000	0,99817	0,99909	0,00092	7,95505
Q2 2021	Gradient Boosting Classifier	0,99948	1,00000	0,99895	0,99948	0,00052	12,49542
Q1 2021	Gradient Boosting Classifier	0,99922	1,00000	0,99843	0,99922	0,00078	8,56730
Q4 2020	Gradient Boosting Classifier	0,99895	1,00000	0,99791	0,99895	0,00105	7,22738
Q3 2020	Gradient Boosting Classifier	0,99935	1,00000	0,99869	0,99935	0,00065	11,22264
Q2 2020	Gradient Boosting Classifier	0,99908	1,00000	0,99817	0,99909	0,00092	11,84590
Q1 2020	Gradient Boosting Classifier	0,99922	1,00000	0,99843	0,99922	0,00078	9,19991
Q4 2019	Gradient Boosting Classifier	0,99948	1,00000	0,99895	0,99948	0,00052	7,83903
Q3 2019	Extreme Gradient Boosting	0,99882	1,00000	0,99765	0,99882	0,00118	**0,64925**
Q2 2019	**Gradient Boosting Classifier**	**0,99961**	**1,00000**	**0,99922**	**0,99961**	**0,00039**	**8,11861**
Q1 2019	Extreme Gradient Boosting	0,99869	1,00000	0,99739	0,99869	0,00131	**0,96973**
Q4 2018	Gradient Boosting Classifier	0,99935	1,00000	0,99869	0,99935	0,00065	11,91560
Q3 2018	Gradient Boosting Classifier	0,99935	1,00000	0,99869	0,99935	0,00065	8,61262
Q2 2018	Gradient Boosting Classifier	0,99908	1,00000	0,99817	0,99909	0,00092	10,63663
Q1 2018	Gradient Boosting Classifier	0,99882	1,00000	0,99765	0,99882	0,00118	7,75979

Table 8 shows the best performing models among all models. Sampling with the oversampling technique stood out in the models applied for each quarter. Gradient-based algorithms provide reliable results on the data set. It is



also seen that it achieves good results in terms of performance. The model that stands out with the lowest error (0.00039), highest F1 (0.99961) and highest precision (0.99922) values is the model created with the oversampling technique in the 2nd quarter of 2019. The results of the models are similar and there is no significant difference between the quarters.

Figure 4. 2019 - 2nd quarter model complexity matrix

Figure 4 shows the confusion matrix applied on the test data of the model established for the 2nd quarter of 2019. According to the results of the confusion matrix evaluating the performance of the credit classification model, the TN value is 819, meaning that the model correctly classifies the truly negative as negative. This emphasises the reliability and accuracy of the model in negative classifications. Since the FP value is 0, the model did not classify samples that were actually negative as positive. This indicates that the model correctly classifies positives, and the probability of false alarms is low. The FN has a value of 3, which means that the model classified three instances as negative when they were actually positive. However, since this number is low, we can say that the cases where the model misses positive examples are limited. TP value is 24, meaning that the model correctly classified the truly positive ones as positive. This reflects the success of the model in positive classifications. It is seen that the overall performance of the model is positive, and its accuracy is high, especially in negative classifications.

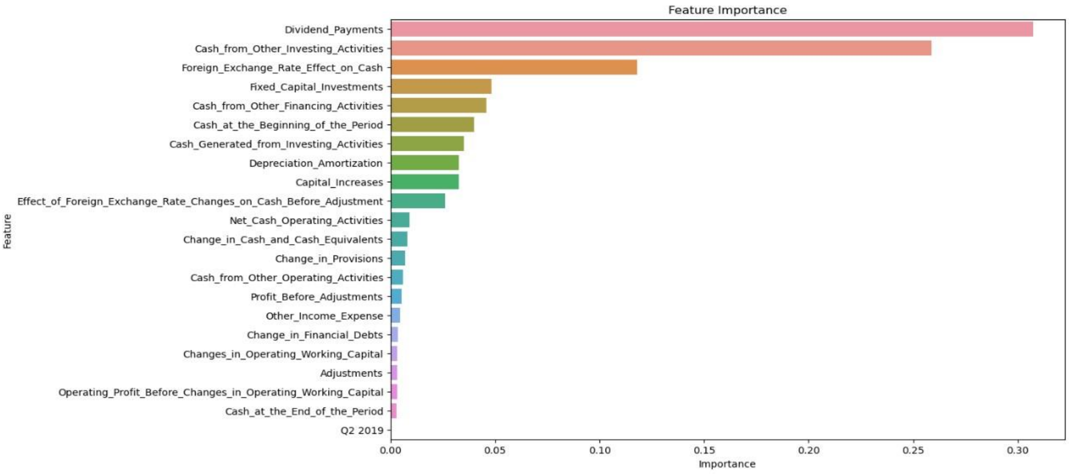


Figure 5. Impact of independent variables in the 2nd quarter 2019 model

Figure 5 shows the effect of the independent variables of the model on the dependent variables. The feature with the highest impact is "Dividend Payments" with 30.72%. This result shows that the dividend distributions of the company are determinant on the model. It can be stated that a company that distributes regular dividends may have a positive effect on loan applications. "Cash from Other Investing Activities" ranked second with 25.87%. This indicates that the company's cash flow from other investing activities has a significant impact on the model. The "Foreign Exchange Rate Effect on Cash" feature with 11.81% indicates how changes in the company's foreign exchange rates affect its cash position. It is a principal factor especially for companies doing international business. The "Fixed Capital Investments" feature with 4.81% represents the importance of the company's fixed capital investments in the model. It could be critical to the company's growth strategy and asset management. The attributes with low significance include "Net Cash Operating Activities" with 0.91%, "Change in Cash and Cash Equivalents" with 0.81%, "Change in Financial Debts" with 0.36% and "Cash at the End of the Period" with 0.28%. The low importance levels of these characteristics indicate that these factors have a limited impact on the credit assessment of the model. The data for the second quarter of 2019 is the variable with the lowest impact on credit assessment.

5. Results and Conclusions

In Q2 2019, the loan classification performance of the best model identified stands out with very impressive metrics. 99% precision indicates that the model correctly classifies loan applications, indicating that false positive predictions are low. This indicates that when loan applications are assessed positively, this assessment is highly accurate. Likewise, an F1 score of 99% indicates both the accuracy of the positive predictions and the ability of the model to successfully cover true positive cases. This indicates a balanced performance of the model in terms of precision and sensitivity. A high F1 score indicates that the credit classification model is reliable and effective. The error rate of 0.003% is a metric that reflects the overall success level of the model. A low error rate indicates that the model has an extremely low tendency to mispredict. This indicates that the loan classification model exhibits a high ability to make financially reliable and accurate predictions. The combination of these metrics shows that the best model identified in Q2 2019 provides high accuracy, reliability, and efficiency in credit assessment. This can contribute to financial institutions assessing credit applications more reliably and managing credit risk more effectively.

Among the effective variables identified, dividend payments, cash from other investing activities and foreign exchange rate effect on cash provide important financial indicators. Firstly, dividend payments may affect the liquidity of dividends paid to shareholders from the company's profits. This may be decisive in assessing the company's cash flow situation and loan repayment capacity in loan applications. Secondly, cash from other investing activities refers to the amount of cash generated by the company from investing activities. A high value may indicate that the company is executing growth strategies and seeking new opportunities. This may increase the confidence in the company's growth potential in loan applications. Finally, foreign exchange rate effect on cash reflects the impact of changes in foreign exchange rates on cash flow. This determines the company's international operations and foreign exchange risk. The assessment of foreign exchange risk is a crucial factor in loan applications. On the other hand, net cash generated from operating activities, adjustments, changes in operating working capital, operating profit before changes in operating working capital (+), change in financial debts, change in provisions, depreciation & amortisation, other income/expense, change in cash and cash equivalents, cash at the end of the period and profit before adjustments, which are among the low impact variables, have almost zero contribution to the financial model. This may suggest that these variables do not play a critical role in the evaluation of loan applications and are of less importance in financial strategies. For example, net cash generated from operating activities reflects the company's cash flow from its core business activities. if this value is low, we may think that the company's cash flow from operations is weak. similarly, change in financial debts can be an indicator of the company's financing strategy, and a high change in debt may indicate the need for the company to update the financial structure. The low impact of these elements suggests that they are not of critical importance in financial models and are less influential in the assessment of overall loan applications.

Similar model results suggest that there are no significant differences across quarters. This indicates that the models for each quarter perform similarly and that there are no significant variations in the loan classification process. The homogeneous results obtained imply a stable performance in model selection and quarterly analysis. This indicates that quarterly changes in the performance of the firm do not significantly affect the loan classification models.

As a future study, panel data analysis will be added and to improve the cross – validation methods suitable for the time series. Statistical tests and seasonality analysis will be added for more detailed interpretation of quarters. It could be applied to more comprehensive credit datasets obtained from financial institutions. The methods related to modifying procedure will be continued with novel machine learning techniques.

References

[1] S. Çil Koçyiğit, Ş. Güngör Tanç, Nakit Akışlarının Sağlandığı Faaliyetler Modeli ile İşletmelerin Nakit Akış Profillerinin İncelenmesi: BİST 30, BİST 50 Ve BİST 100 Endeksleri Karşılaştırması, Muhasebe ve Finansman Dergisi – Ağustos 2021 Özel Sayı, (2021), 137148. ISSN: 2146-3042.

[2] R. Aktaş, M. Mete Doğanay, B. Yıldız, Mali Başarısızlığın Öngörülmesi: İstatistiksel Yöntemler ve Yapay Sinir Ağı Karşılaştırılması, Ankara Üniversitesi SBF Dergisi, 58, (2003), 1-24. Doi: 10.1501/SBFder_0000001691.

[3] Y. Aker, Finansal Başarısızlık Tahmininde Makine Öğrenmesi Yöntemlerinin Kullanımı: Türkiyedeki KOBİ' ler Üzerine Bir Uygulama, Giresun Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı Doktora Tezi, (2021), 1-44.

[4] B. Karataş, A. V. Can, Bibliometric Analysis of Postgraduate Theses Published On Financial Failure Prediction In Turkey (1991-2021), Journal of Accounting and Taxation Studies, (2023), 17-55. Doi:10.29067/muvu.1139919.

[5] K. Sadeghzadeh, B. Elmas, Makroekonomik Faktörlerin Hisse Senedi Getirilerine Etkilerinin BIST'de Araştırılması, Muhasebe ve Finansman Dergisi, (2018), 211-213. Doi: 10.25095/mufad.465941.

[6] H. Altınbaş, Metaheuristic Algorithms and Modern Credit Classification Methods: A Systematic Review, Istanbul Business Research, 49(1), (2020) 146-175 Doi : 10.26650/ibr.2020.49.0033.

[7] D. Ilter, E. Deniz Howe, O. Kocadagli, Hybridized ANN Classifiers with a Novel Feature Selection Procedure based Genetic Algorithms and Information Complexity in Credit Scoring, Appl Stoch Model Bus Ind., 37(2), (2021), 203-228. Doi:10.1002/Asmb.2614.

[8] D. Ilter, E. Deniz Howe, O. Kocadagli, Hybridized ANN Classifiers with a Novel Feature Selection Procedure based Genetic Algorithms and Information Complexity in Credit Scoring, Appl Stoch Model Bus Ind., 37(2), (2021), 203-228. Doi:10.1002/Asmb.2614.

[9] A. M. Esi, Bankacılık Sektöründe Kredi Ödemelerinin Makine Öğrenimi Siniflandirma Algoritmalarina Göre Analizi, Marmara Üniversitesi sosyal Bilimler Enstitüsü Ekonometrik anabilim Dalı Yöneylem Araştırması Bilim Dalı Yüksek Lisans Tezi, (2022), 181.

[10] D. Ilter, Kredi Skorlamada Yapay Zekâ Teknikleri İle Çok Aşamali Lojiistik Modellemeyi Temel Alan Hibrit Yaklaşimlar, Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı Doktora Tezi, (2021), 1-59.

[11] L. E. Melkumova, S. Ya. Shatskikh, Comparing Ridge and Lasso Estimators for Data Analysis, Procedia Engineering, 201, (2017),746-755.

[12] S. Vimalraj, R. D. Porkodi, A Review on Imbalanced Data, Proceeding of 2018 IEEE International Conference on Current Trends toward Converging Technologies, Coimbatore, India, 3(4), (2017), 444–449. Doi:10.23883/IJRTER.2017.3168.0UWXM

[13] E. Kartal, Z. Ozen, Dengesiz Veri Setlerinde Sınıflandırma, CHAP, (2017), 109 -131. ISSN: 978-605-4735-98-3.

[14] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, Journal of artificial intelligence research, 16, (2002), 321–357.

[15] A. Alan, M. Karabatak, Veri Seti - Sınıflandırma İlişkisinde Performansa Etki Eden Faktörlerin Değerlendirilmesi, Fırat Üniversitesi Müh. Bil. Dergisi, 32(2), (2020), 531 – 540.

[16] G. Altan, S. Demirci, Makine Öğrenmesi ile Nakit Akış Tablosu Üzerinden Kredi Skorlaması: XGBoost Yaklaşımı, Journal of Economic Policy Researches, 9(2), (2022), 398-424. E-ISSN: 2148-3876.

[17] T. E. Tütüncü, S. Gürsakal, Kredi Temerrüt Riskini Tahmin Etmede Makine Öğrenme Algoritmalarının Karşılaştırılması, European Journal of Science and Technology, 50, (2023), 14-22. Doi: 10.31590/ejosat.1171611.

[18] Ö. Y. Can, Makine Öğrenmesi Teknikleri Kullanılarak Kredi Analizi, İstanbul Aydın Üniversitesi Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans Tezi, (2020), 1-57.

[19] V. Aithal, R. D. Jathanna, Credit Risk Assessment using Machine Learning Techniques, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(1), (2019), 3482-3486.

[20] D. Ilter, O. Kocadagli, N. Ravishanker, Feature Selection Approaches for Machine Learning Classifiers on Yearly Credit Scoring Data, Young Business and Industrial Statisticians Workshop on Recent Advantages in Data and Business Analytics (y-BIS2019), Istanbul, Turkey, 2019. Conference Proceeding e-Book with ISBN (978-605-5005-95-5) and Serial Number (eMSGSÜ-FEF-İST019/09-Kat1), (http://kutuphane.msgsu.edu.tr/yordambt/yordam.php?aDemirbas=EK925DE7C3), (2009), 200-204.

[21] K. Ş. Erdem, M. A. Bakır, Makine ve Ekipman İmalatı Sektöründe İzolasyon Ormanı ve Yeniden Örnekleme Yöntemleri Kullanılarak Finansal Başarısızlığın Tespit Edilmesi, Journal of Productivity, 57 (4), (2023), 719-734.

[22] Z. Kardeş, T. Kandemir, Bağımsız Denetim Görüşlerinin Tahmin Edilmesinde Lojistik Regresyon ve Yapay Sinir Ağı Yöntemlerinin Karşılaştırılması: BİST Kimya İlaç Petrol Lastik ve Plastik Ürünler Sektöründe Bir Uygulama, KMÜ Sosyal ve Ekonomik Araştırmalar Dergisi (KMUSEKAD), 25(44), (2023), 293-308.

[23] S. Hamal, Decision Making Approaches For Financial Accounting Fraud: Case Study For Small and Medium-sized Enterprises (SMEs) in Turkey, Marmara University Institute for Graduate Studies in pure and Applied Sciences Department of Industrial Engineering Master thesis, (2022), 1-236.

[24] O. A. Okocha, Machine Learning Approach to the Prediction of Bank Customer Churn Problem, Istanbul Aydın University Institute of Graduate Studies Department of Software Engineering Artificial Intelligence and Data Science Program Master thesis, (2023), 1-48.

[25] F. Uyanık, Telekomünikasyon Sektörü İçin Veri Madenciliği ve Makine Öğrenmesi Teknikleri İle Ayrılan Müşteri Analizi, İstanbul Ticaret Üniversitesi İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Yüksek Lisans Tezi, (2021), 1-57.

[26] G. Zengin, Finansal Teknoloji Alanında Kullanıcı Deneyimlerinin Makine Öğrenmesi Yöntemleri İle İncelenmesi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı Yüksek Lisans Tezi, (2022), 1-100.

[27] M. Çalışkan, Makine Öğrenme Teknikleri Kullanarak Epilepsi Teşhisi, Kahramanmaraş Sütçü İmam Üniversitesi Fen Bilimleri Enstitüsü Enformatik Ana Bilim Dalı Yüksek Lisans Tezi, (2021), 1-54.

[28] Y. C. Kılıçkap, Lojistik Sektöründe Makine Öğrenmesi Modelleri Yardımı ile Uygun Paket Türünün Seçilmesi, İstanbul Ticaret Üniversitesi Fen Bilimleri Enstitüsü Endüstri Mühendisliği Anabilim Dalı Yüksek Lisans Tezi, (2023), 1-69.

[29] G. Kim, S. Kim, B. Jang, Classification of mathematical test questions using machine learning on datasets of learning management system questions, PLoS ONE 18(10), (2023), 1-17.

[30] Is Yatirim, Stocks, https://www.isyatirim.com.tr/tr-tr/analiz/hisse/Sayfalar/default.aspx , [Date of Access: 20.03.2023].