O P E R A T I O N S  R E S E A R C H  A N D  D E C I S I O N S
No. 1 2022
DOI: 10.37190/ord220102

# A COMPARATIVE STUDY OF CORPORATE CREDIT RATING PREDICTION WITH MACHINE LEARNING

SEYYIDE DOĞAN[1], YASIN BÜYÜKKÖR[1], MURAT ATAN[2*]

[1]Department of Econometrics, Karamanoglu Mehmetbey University, Karaman, Turkey

[2]Department of Econometric, Ankara Hacı Bayram Veli University, Ankara, Turkey

Credit scores are critical for financial sector investors and government officials, so it is important to develop reliable, transparent and appropriate tools for obtaining ratings. This study aims to predict company credit scores with machine learning and modern statistical methods, both in sectoral and aggregated data. Analyses are made on 1881 companies operating in three different sectors that applied for loans from Turkey's largest public bank. The results of the experiment are compared in terms of classification accuracy, sensitivity, specificity, precision and Mathews correlation coefficient. When the credit ratings are estimated on a sectoral basis, it is observed that the classification rate considerably changes. Considering the analysis results, it is seen that logistic regression analysis, support vector machines, random forest and XGBoost have better performance than decision tree and *k*-nearest neighbour for all data sets.

**Keywords:** *credit ratings, credit risk, machine learning*

## 1. Introduction

Credit rating is widely used in financial markets as a measure of creditworthiness, investment risks, and failure probability of fund requesters such as companies, countries, and financial institutions. In general, credit ratings are a classification process based on past, current, qualitative and quantitative data of borrower institutions, and presenting prudential views [1]. A rating process is also a tool that measures the probability of non-payment of the principal and interest of the securities (bonds, commercial bills, etc.) issued by the relevant institution.

Ratings are important determinants of adjusting risk premiums and marketability of corporate bonds by reducing uncertainty for financial markets and investors [2–4]. They

*Corresponding author, email address: dogans@kmu.edu.tr
*Received 15 September 2021, accepted 3 February 2022*

also provide lower funding costs, financial flexibility, and ease of entry to capital markets for issuers of securities [5]. Credit ratings that reliable assessment of the level of credit risk and the appropriate is also important to reduce the threat of bankruptcy, not only of a particular company but of all cooperating stakeholders and all related financial institutions [6]. However, these ratings should not be interpreted as "buy" or "sell" recommendations as they do not predict the profitability of investing in the rated company. Generally, ratings are expected to reflect an opinion based on at least a three-year prospective assessment of the borrower's financial health with an annual review [7].

Credit ratings, which are frequently used by bond investors, lenders, government officials, and creditor institutions to assess investment risks, are alphanumeric symbols that reveal a company's ability to find cash and its willingness to pay to meet its financial obligations fully and on time [8, 9].

Credit ratings have become one of the primary references of financial institutions for assessing credit risk, improving cash flow, reducing potential risks, and making managerial decisions [10, 11]. In addition, following the New Basel Capital Agreement, Basel II, and Basel III as new standards are set for capital adequacy in banks, it is critical to managing the financial risk that will arise if the counterparty fails to meet its obligations. For this reason, efficient and effective assessment tools of banks regarding the borrower's credit risk have become essential for every financial institution and it is in question that companies look for a rating condition when giving credit [12, 13].

Independent credit rating agencies (CRAs) such as Standard & Poor's, Moody's and Fitch, which have gained competence in this field, have undertaken the task of obtaining the credit ratings of companies. These institutions analyse the risk situations of companies in-depth, using various statistical models, with a combination of public and private financial data from different aspects from strategic competitiveness to operational level, and the subjective judgments of field experts, publish their credit ratings and update these credit ratings periodically. Credit rating assessment is an expensive and complex process, mainly because it requires months of effort from many experts to analyse different variables that reflect the reliability of companies [14, 15].

CRAs provide international capital markets with consistent and comparable assessments of the credibility of companies and financial institutions of countries, under the principles of independence, objectivity, reliability, and transparency. They make these evaluations based on financial, economic and technical analysis, taking into account different factors such as the financial and non-financial structure of the company, the conditions of the market in which it is located, the characteristics of the industry in which it works, and the general economic and political situation [16]. Although rating agencies and many theoretical writers emphasize the importance of subjective judgment in the rating process, many researchers have achieved promising results in credit rating estimation with different statistical and artificial intelligence methods [2].

The main difference between traditional statistical methods and machine learning methods is that statistical methods usually build the model, for researchers, with a linear

structure and estimation of parameters to fit the observations, whereas, in machine learning methods, the original structure of the model is allowed to be learned from the data. Machine learning can learn the solution to the problem from the input data without human intervention [17]. This situation provides a great advantage in credit risk assessment; It will be more transparent and ethical if the answer to a question about providing credit is given by the parallel decision of machine learning algorithms with different learning processes instead of the implicit opinion of a single expert [18].

Precautionary rules, model reliability, accurate estimation of default risk, effective decision-making, and transparency have key roles in credit risk assessment [18, 19]. Many authors emphasise the importance of ML techniques for the efficient processing and tracking of loans [20, 21].

In this study, we apply modern statistics and machine learning methods in the field of credit rating. Our main motivation in this study is to expand the use of machine learning techniques in credit risk assessment and to pave the way for digitalisation in credit processing and monitoring activities. To achieve this aim, a real-world credit rating problem is considered, and a practical application of different modern statistical and machine learning methods is presented comparatively. However, most of the articles published in the field of credit rating focus on the credit ratings of the countries that the three giant credit rating agencies monitor and many of them focus on credit risk forecasts in developed countries. On the other hand, credit risk analysis for developing countries is quite limited.

The contribution of this study to the literature is the evaluation of the credit risk analysis of small and large-scale companies in the medium and high-risk loan portfolio of the Turkish public bank, both on a holistic and sectoral basis, with machine learning algorithms separately. There is no study based on real practice that makes this evaluation in Turkey's leading sectors such as manufacturing, marketing, and construction and contracting.

The following sections of this article are as follows: in Chapter 2, studies in the field of credit rating using modern statistics and machine learning methods are included. In Chapter 3, the theories of the support-vector machine (SVM), *k*-nearest neighbourhood (KNN), decision tree (DT), logistic regression analysis (LRA), random forest (RF) and extra gradient boosting (XGBoost) methods are presented. In Chapter 4, the optimal parameters of the methods and the optimal properties for the methods are determined, and the experimental results obtained are compared within the framework of different performance criteria. Finally, the conclusion and evaluation of the article are presented in Chapter 5.

## 2. Related works

The use of statistical methods for credit rating estimation dates back to 1959 when Fisher used the ordinary least square (OLS) to explain the variance of a bond's risk

Table 1. Previous studies about credit rating

| Methods | Data sets | Var | Sample size | Class | FS | Accuracy [%] | Ref. |
|---|---|---|---|---|---|---|---|
| OLS, ANN | S&P | 10 | 47 | 2 | correlation | 64.7, 88.3 | [27] |
| ANN, MDA | Bell Telephone | 7 | 56 | 2 | – | 88, 39 | [28] |
| CBR, MDA, ID3 | Korean companies | 12 | 3886 | 5 | ANOVA + MDA | 80.8, 77.8, 72.9 | [32] |
| CBR, SOM+CBR, LVQ+CBR, MDA | commercial paper issued in Korean companies | 13 | 2971 | 5 | ANOVA + stepwise | 61.1, 67.1, 69.1, 55.0 | [37] |
| LRA, SVM, ANN | Taiwan rating corporation | 16 | 74 | 5 | ANOVA | 72.97, 79.73, 75.68 | [2] |
| LRA, SVM, ANN | S&P | 14 | 265 | 5 | ANOVA | 75.09, 80.38, 80.75 | [2] |
| SVM, MDA, CBR, BPN | Korean information services | 10 | 3017 | 5 | ANOVA + MDA | 67.2, 58.7, 63.4, 59.9 | [36] |
| LRA, ANN, SVM, NFS, F-SVM, VBE, RBE | Japanese consumer credit card /UK corporations | 13/12 | 653/60 | 2 | – | 75.8, 80.7, 79.9, 77.9, 83.9, 85.2, 87.2 | [38] |
| OMSVM, MDA, MLR, CBR, ANN | manufacturing industry in Korea | 14 | 1295 | 4 | *t*-Test, Wilk's lambda, *F*-value | 67.9, 65.0, 67.1, 51.4, 69.5 | [39] |
| RST, RF+DT, DT, RF+CART, CART, RF+SVM, SVM, RF+RST | Taiwan Economic Journal (TEJ) | 18 | 2470 | 3 | RF | 90.3, 84.0, 83.5, 77.6, 71.1, 77.8, 74.4, 93.4 | [40] |
| MLP, RBF-NN, SVM, NB, RF, LDC, NMC | US/European non-financial companies | 81/43 | 852/244 | 9 | IFS-MFS-CBF -CONBF- GAF | 58.2, 56.4, 57.9, 44.0, 59.3, 56.1, 56.4 63.0, 64.3, 60.8, 68.7, 56.4, 51.3, 51.0 | [14] |
| BP, ELM, I-ELM, SVM | Moody's investor services/S&P | 7 | 418/780 | 2 | correlation | – | [9] |

| Methods | sets | Var | Sample size | Class | | Accuracy [%] | Ref. |
|---|---|---|---|---|---|---|---|
| MLR, MDA, BPNN, M- DTSVM T | Taiwan dataset | 18 | – | 3 | – | 60.0, 64.8, 70.8, 83.8 | [41] |
| LRA, RF, XGB, NN | asking loan from a bank | 181/10 | 117019 | 2 | – | AUC: 87.6, 99.3, 99.4, 97.9 | [42] |
| MLR, LDA, ANN, SVM, GPC, RF, XGB | 308 companies in S&P | 27 | 308 | 6 | – | 59.6, 61.7, 60.8, 63.6, 61.6, 64.6, 62.3 | [43] |
| OLR, RF, GB | developed and emerging economies | 27 | 880 | 2 | correlation | 26.9, 50.5, 48.8 | [44] |
| DT, RF, ANN, SVM | S&P, Moody's, Fitch | 16 financial | 6029 | 2 | – | 84.2, 82.8, 73.9 | [45] |
| DT, RF, ANN, SVM | S&P, Moody's, Fitch | 20 energy | 2009–2018 | 2 | – | 82.1, 84.45, 78.1 | [45] |
| DT, RF, ANN, SVM | S&P, Moody's, Fitch | 20 health care | 2009-2018 | 2 | – | 83.9, 82.9, 76,6 | [45] |
| LRA, SVM, RF, GB, FNN | Croatian banks | 109 | 870 710 | 2 | – | 91.6, 91.6, 94.4, 94.8, 92,8 | [46] |
| RF, XGB, LGBM, AdaB ETC, LRA, KNN, NB | mortgage credit in Romania | 10 | 3 225 | 2 | correlation | 94.8, 95.2, 95.3, 81.7,88.3, 84.8, 56.5, 57.3 | [17] |

MLR – multinomial logit regression, LDA – linear discriminant analysis, ANN – artificial neural network, SVM – support vector machines, GPC – gaussian process classifier, RF – random forest, GB – gradient boosting, XGB – extreme gradient boosting, OLR – ordered logistic regression, FNN – forward neural network, RST – rough set theory, CART – classification and regression trees, DT – decision trees, MLP – multilayer perception, LDC – linear discriminant analysis, NMC – nearest mean classifier, RBF – radial basis function, MDA – multiple discriminant analysis, CBR – case based reasoning, BPN – back propagation neural network, BP – back propagation, ELM – extreme learning machine, I-ELM – incremental extreme learning machine, ID3 – induction algorithm, NB – naive bayes, CBF – correlation based filter, GAF – genetic algorithm filter, CONBF – consistency based filter, IFS – individual future selection, MSF – mixed future selection, OMSVM – ordinal multiclass svm, M-DTSVM-RST: multiple future selection dt+ svm+ rst, NFS – neuro fuzzy system, F-SVM – fuzzy svm, VBE – volatility based ensemble, RBE – reliability based ensemble, LGBM – light gr ient boosting method, ADABOOST: adaptive boosting, ETC – extra trees classifier.

premium [22]. Pinches and Mingo [23] use multiple discriminant analysis (MDA) to derive a linear discriminant function associating a set of independent variables with a dependent variable. Other researchers also use logistic regression analysis [24] and probit analysis [25] from multivariate statistical methods.

However, recently, the most preferred methods in the field of credit rating have turned to artificial intelligence technologies. An artificial neural network [26] is one of the most frequently used methods in credit risk assessment. In the first years, the success of traditional statistical methods was mostly compared with ANN and it was observed that better results were obtained with ANN [27, 28]. Some authors study the classification accuracy of different machine learning models including ANN [26, 29, 30]. In subsequent studies, Kim, Weistroffer's [31] rule-based systems, Shin and Han's [32] case-based reasoning (CBR), Kalaivani and Shunmuganathan's [33] *k*-nearest neighbours (KNN), Zekic–Susac, Sarlija [34] and Baesens, Gestel [35] propose classification and regression trees (CART) approaches. Lee [36], Zhong, Miao [9] and Huang, Chen [2] demonstrate a good implementation of a relatively new algorithm, SVM. In all these studies, it is observed that different methods dominate over the others each time, depending on the size and complexity of the data set handled. For this reason, it will be beneficial to use several different methods in the field of credit rating rather than a single method. In Table 1, the previous studies used in the field of credit rating are expanded, the methods used for the feature selection stage are included, and the estimation accuracy rates obtained by the methods are presented.

Considering the aforementioned pioneering studies in the field of credit rating, it is clear that machine learning techniques produce better results than statistical methods. On the other hand, no machine learning method has provided absolute superiority to the others in the credit evaluation and rating studies that we have traced in the literature. Depending on the nature of the problem and the change in the sample, the prominent ML method differed. It is generally accepted that a single method should not be chosen as the estimation method. The most commonly used ML methods in the field of credit valuation are SVM, MLP, ANN, KNN and DT. Similarly, the statistical method LRA, which can explain the underlying cause and effect relationship, is also frequently used. It is observed that RF, AdaBoost. and XGBoost ensemble methods, which draw attention due to their good results and fast operation, have been used more intensively in studies published in recent years.

## 3. Methodology

### 3.1. Logistic regression analysis

Logistic regression analysis (LRA) is an old statistical method for predicting categorical dependent variables using a given set of independent variables [47, 48]. In LRA,

the dependent variable $y_i$ ($i$ = 1, 2, ..., $n$) takes only the values of 1 and 0 with the$p_i$ and $1 - p_i$ probabilities. Thus, it can be said that the dependent variable has a Bernoulli probability density function. Considering the situation where the dependent variable is equal to 1

$$P(y_i = 1) = p_i = \frac{1}{1 + e^{-\mathbf{x}\boldsymbol{\beta}}} \tag{1}$$

is obtained. Here $\boldsymbol{\beta}$ is $(k+1)\times 1$ dimensional parameter vector and $\mathbf{x}$, $k + 1$ dimensional explanatory variable matrix. For $y = 1$, the odds ratio can be shown with

$$\frac{P(y=1)}{1-P(y=1)} = \frac{p_i}{1-p_i}$$

By using the logarithm of the odds ratio value

$$\log it(y_i = 1) = \ln \frac{p_i}{1-p_i} = \mathbf{x}\boldsymbol{\beta} \tag{2}$$

is obtained. The purpose of this transformation is to make the parameter vector linear and to use the advantages of the linear regression model in logistic regression analysis. In machine learning methods, class affiliation of predicted label $x$ values is determined according to the probability values they receive. If the probability $x$ is equal or greater than 0.5, it takes the value 1, and if it is less, it takes the value 0 [49].

### 3.2. *k*-nearest neighbour

The KNN algorithm presented by Fix and Hodges [50] is a non-parametric clustering algorithm that is easy to understand and implement, with a fast and good prediction performance. Its main purpose is to find the nearest neighbour for each sample point that needs to be estimated. A distance function is usually used when searching for the nearest neighbour. The performance of this method depends on the number of neighbour $K$ determined by the researcher and the distance function [51]. The selection of the most appropriate $K$ value is made by establishing a balance between deviation and variance. Small $K$ values increase the bias and cause over fitting, while large $K$ values cause the class boundaries to become blurred and the bias to decrease too much. Simulation and cross-validation are usually done when making the selection. The distance functions frequently used for the KNN algorithm are the Euclidean, Chebyshev and Manhattan distances. These distances are:

$$d_{\text{Euclidean}}(x, y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{3}$$

$$d_{\text{Manhattan}}(x, y) = \sum_{i=1}^{n} |x_i - y_i| \tag{4}$$

$$d_{\text{Chebyshev}}(x, y) = \max_i (|x_i - y_i|) \tag{5}$$

where $x$ and $y$ represent two vectors in the feature space, $x_i$ and $y_i$ are their coordinates.

### 3.3. Support-vector machines

Support-vector machine (SVM) is a relatively newly developed method based on statistical learning theory [52] and can be applied to classification (pattern recognition) problems and perform function estimation in regression. The standard SVM classifier was introduced as a quadratic optimisation problem and was first applied to two-class classification problems.

The best hyperplane separating the $S$ training set consisting of $P$ number of attributes $(x_i, x_j)$, $i = 1, ..., m$, $x_i \in \Re^n$ input and output pairs, with **w** being the weight vector, is the plane that minimises $\eta(w) = 1/2\|\omega\|^2$.

Linear non-separating problems can be solved linearly by accepting a certain error to be assigned for misclassified samples. In this case, the problem can be written as finding the hyperplane that minimises training errors through slack variables:

$$\begin{aligned} &\min_{\omega, b, \xi} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{m}\xi_i \\ &\text{const: } y_i\left(\langle \omega, x_i \rangle + b\right) \geq 1 - \xi_i \text{ and } \xi_i > 0, \quad i = 1, ..., m \end{aligned} \tag{6}$$

Here, C is the penalty parameter on training errors and $\xi_i$ is the non-negative slack variable [53]. This problem can be solved by making use of the Lagrangian function. The dual model in which Lagrangian variables are maximised is given below [52]:

$$\begin{aligned} &\max_{\alpha} L_d(\alpha) = \sum_{i=1}^{m}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{m} y_i y_j \alpha_i \alpha_j \langle x_i, x_j \rangle \\ &\text{const}: \sum_{i=1}^{m} y_i \alpha_i = 0 \text{ and } 0 \leq \alpha_i \leq C, \quad i = 1, ..., m \end{aligned} \tag{7}$$

The mapping function $\varphi$ is used for non-linear SVM training examples. Defining a suitable kernel function based on the inner product is to do a non-linear transformation of the data from the input space to the high-dimensional (maybe infinite) feature space to make the data linearly separable. The Kernel function given in equation (8) uses instead of the inner product $\langle x_i, x_j \rangle$ that we use in the objective function of the dual model (9).

$$K\left(x_i, x_j\right) = \varphi\left(x_i\right)^T \varphi\left(x_j\right) \tag{8}$$

$$\begin{aligned} &\max_{\alpha} L_d\left(\alpha\right) = \sum_{i=1}^{m} \alpha \sum\nolimits_{i=1}^{m}{}_{i} - \frac{1}{2} \sum_{i,\, j=1}^{m} y_i y_j \alpha_i \alpha_j K\left(x_i, x_j\right) \\ &\text{const:} \sum_{i=1}^{m} y_i \alpha_i = 0 \text{ and } 0 \le \alpha_i \le C, \quad i = 1, ..., m \end{aligned} \tag{9}$$

Following the solution step in the linearly separable case, the decision function is obtained:

$$f\left(x\right) = y = \text{sign}\left( \sum_{i=1}^{m} \alpha_i^* y_i \left\langle K\left(x_i, x_j\right) \right\rangle + b^* \right) \tag{10}$$

Many kernel functions help the SVM find the optimal result. The most frequently used of these are: polynomial kernel $K\left(x_i, x_j\right) = \left(1 + x_i x_j\right)^d$ radial base kernel $K\left(x_i, x_j\right) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right)$ and sigmoid kernel $K\left(x_i, x_j\right) = \tanh\left(K\left(x_i x_j\right) - \delta\right)$ [54, 55].

### 3.4. Decision trees

The decision trees algorithm is a simple and understandable classification algorithm that models the relationship between features and outputs with the help of a tree [56]. It is also a widely used algorithm because decision trees are resistant to outliers. Decision trees are classified from top to bottom (flow chart). It starts from the root node, is divided into internal nodes according to the values of the features, and continues until finally a leaf node is reached (class label assigned) [57]. The creation of the decision tree is done by repeatedly partitioning the internal nodes. Partition criteria used for partitioning internal nodes and enabling the selection of features:

- Gini index [57],
- the information gain [58],
- chi-squared test [59, 60],
- entropy for C4.5 algorithm [58].

Even if the decision trees are accurate and efficient, sometimes due to overfitting, the tree can be very large and increase the error for the test dataset. The pruning method is generally used to eliminate the overfitting problem [61]. After pruning, the tree is reduced in size and complexity is removed, resulting in an easy-to-understand tree. There are two methods for pruning: pre-pruning and post-pruning [62]. In this study classifications and regression trees (CART) algorithm is used.

### 3.5. Random forest

The random forest (RF) algorithm, a supervised machine learning method used for both regression and classification, was first proposed by Breiman et al. [57]. RF is an ensemble method that is easy to use and often provides a good solution even without hyperparameter tuning. It was developed to control the overfitting problem that often occurs in the DT method [63]. RF consists of decision trees selected from training data by the bootstrap method. These decision trees are grown by using CART algorithm, increasing randomness, and reducing the likelihood of an overfitting problem. In the RF method, each tree in the forest performs a class prediction, and the class with the most votes becomes the model's prediction. The general process of RF is as follows:

- a bootstrap sample is chosen from data to grow each tree of RF,
- data is divided into two sets: training and testing sets. Training data is used to obtain in-bag and out of bag samples,
- to increase accuracy, a random number of features are selected for each tree and used to construct leaves and nodes,
- at the beginning of RF, a feature is appointed as the root node, and then a random tree structure is formed from top to bottom by splitting and branching the training set into subsets [64].

### 3.6. Extreme gradient boosting XGBoost)

Extreme gradient boosting (XGBoost), originally based on gradient boosted decision trees (GBDT), is a supervised machine learning method used for regression and classification problems. The most important feature of the XGBoost is that it is scalable for all scenarios. As compared to the existing machine learning methods, it has a high calculation ability and uses limited memory. According to Chen, Guestrin [65], XGBoost has the following features:

- a novel distributed weighted quantile sketch algorithm for weighted data,
- sparsity-aware for missing values,
- cache-aware algorithm for large datasets.

The objective function of XGBoost is defined as

$$L(\varphi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \tag{11}$$

where $l$ is a differentiable convex loss function measures the difference between $y_i$ and $\hat{y}_i$ which are actual and predicted values. The regularisation term $\Omega$ in the second component of equation (11) can be defined as

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \|w\|^2 \tag{12}$$

where $\gamma$ is the complexity parameter and controls the number of leaf nodes, $T$ is the number of leaf nodes, $w$ is the weight of leaf nodes, and $\lambda$ is the parameter which is used for avoiding overfitting.

# 4. Application

The proposed research design for the application part of the study is presented in Fig. 1. Here, firstly, the data set used in the study is introduced. Then, the data pre-process, which is an important step for every data mining process, is processed. How the hyper-parameters are selected for the methods preferred as the estimation method and the selected optimal parameters are explained. Finally, how the performance of the developed models will be evaluated is explained.

MATLAB 9.4 (R2018a) software was used in this study for the machine learning methods and solvers which are LR (lassoglm), DT (fitctree) and KNN (fitcknn). The SVM model was developed using the LIBSVM [66] software system. Python programming language was used for calculation of RF (random forest classsifier) and XGBoost (XGBoost classifier).

## 4.1. Research design

**Data.** The companies considered for credit rating operate in the manufacturing, marketing, as well as construction and contracting sectors. Among these companies which applied for a loan to the public bank with the largest capital in the banking sector in 2016

to find resources for their investments and whose credit rating was evaluated by the bank, a total of 1881 companies constitute the sample set, of which 916 are in the manufacturing sector, 746 in the marketing, and 219 in the construction and contracting one. Thirty-nine different financial ratios of companies and one categorical variable are used in the study. These financial ratios are given in Table 2 and their descriptive statistics in Appendix 1.

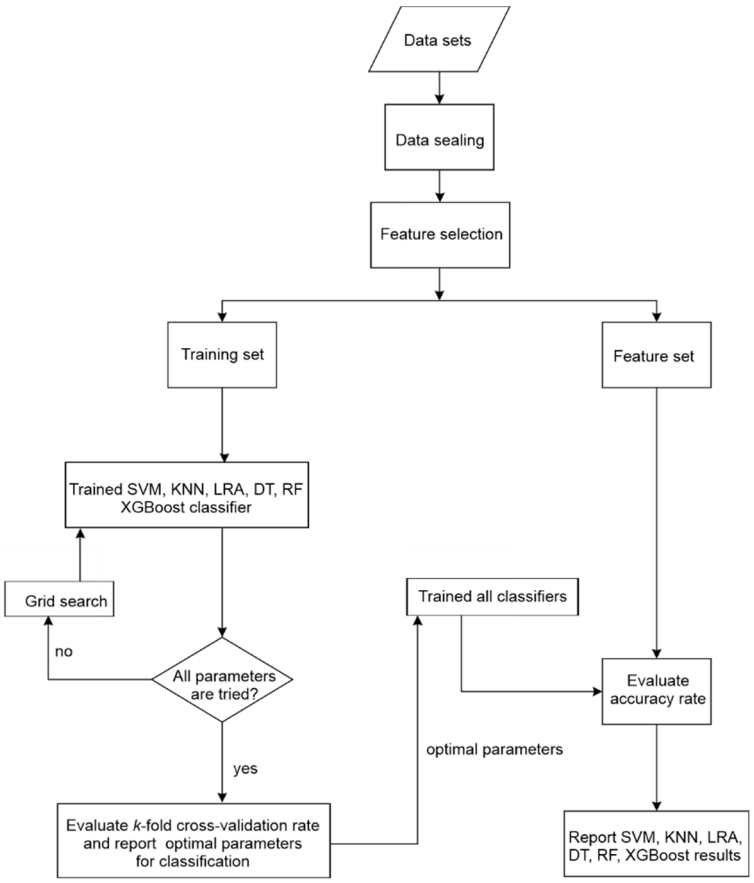


Fig. 1. The flow chart of the machine learning processes

The financial ratios in Table 2 are obtained from the annual balance sheets of the companies updated with footnotes and form the features of the forecasting model. In the study, the status of firms being eligible for and unsuitable for lending is used as a classifier variable. The credit rating of the companies is determined as B+, B–, C+, C–, D+, D–, E+, E–, and F by the credit risk monitoring department of the relevant bank. In this study, since credit rating is considered a two-class problem, companies with a credit

rating of B+, B–, C+, and C– are examined with the "creditable" class label. Since the companies in the loan portfolio of the bank's risk unit are in the medium and high-risk group due to the default rate, the number of companies with A+ and A– credit ratings in the data set is only 2; therefore, they are excluded from the study. On the other hand, companies with credit ratings of D+, D–, E+, E–, and F are considered unsuitable for lending and were evaluated with the "risky" class label.

Table 2. Financial ratios

| | | | |
|---|---|---|---|
| X1[a] | categorical variable for sector type | X21[b] | working capital turnover |
| X2[b] | current ratio | X22[b] | net working capital turnover |
| X3[b] | quick ratio | X23[b] | tangible assets turnover |
| X4[b] | cash ratio | X24[b] | fixed asset turnover |
| X5[b] | inventory/current asset | X25[b] | shareholder's equity turnover |
| X6 | inventory/total assets | X26[b] | asset turnover |
| X7 | stock dependency ratio | X27[b] | net income after tax/shareholder's equity |
| X8[b] | net working capital | X28[b] | net income before tax/shareholder's equity |
| X9[b] | total foreign asset/total assets | X29[b] | financial expense + net income before tax/total liabilities |
| X10 | shareholder's equity/total assets | X30[b] | net income after tax/total assets |
| X11[b] | foreign asset/shareholder's equity | X31[b] | operating margin/total assets-tangible-fixed assets |
| X12[b] | short-term liabilities/total liabilities | X32[b] | operating margin/net sales |
| X13 | net tangible fixed assets/shareholder's equity | X33[b] | gross sales/net sales |
| X14[b] | fixed assets/total foreign asset | X34[b] | net operating profit after tax/net sales |
| X15[b] | fixed assets/shareholder's equity | X35 | cost of goods sold/net sales |
| X16[b] | fixed assets/constant capital | X36 | operating expenses/net sales |
| X17[b] | short-term liabilities/foreign asset | X37[b] | period income/net sales |
| X18[b] | current asset/total assets | X38 | net sales growth (%) (annual) |
| X19 | net tangible fixed assets/total assets | X39[b] | asset growth (%) |
| X20 | stock turnover (annual) | X40 | inventory keeping time |

[a]Used only data-all.
[b]Selected after feature selection.

The data set used in the study was analysed in four groups to diversify the sample and to analyse it on a sectoral basis. These groups consist of (I) data-all covering all the data and adding a categorical variable for sectors, (II) data-manufacturing for the manufacturing sector, (III) data-marketing for the marketing sector, (IV) data-C&C for the construction and contracting sectors. The data set is divided into four sub-groups, taking into account the sector the companies belong to, to observe whether evaluating on a holistic or sectoral basis will make a difference. Thus, the accuracy rates of the prediction model to be established both in all the data and in the sectors are examined. The characteristics of the data set are given in Table 3.

Table 3. Features of data set

| Data sets | Number of class | Sample size | Creditable /risky | Nominal features | Numerical features | Total features |
|---|---|---|---|---|---|---|
| Data-all | 2 | 1881 | 612/1269 | 1 | 29 | 30 |
| Data-manufacturing | 2 | 916 | 203/713 | 0 | 29 | 29 |
| Data-marketing | 2 | 746 | 316/430 | 0 | 29 | 29 |
| Data-C&C | 2 | 219 | 93/126 | 0 | 29 | 29 |

Of the companies in our data-all dataset, 612 are identified as creditable and 1269 as risky. When we look at the characteristics of the data sets in the table, we notice that the data sets are distributed unbalanced due to the disparity of classes in the target variable. Unbalanced distributed datasets have been a particularly striking topic in the data mining literature in recent years, and the performance of many developed methods on unbalanced datasets is expected to be demonstrated. All of the datasets are divided into two groups – training and test datasets to develop the model and to test the validity of the developed model. This separation is randomly allocated with a method called hold-out, with 80% training and 20% testing set. In this case, there are 1505 samples in the training set and 376 samples in the test set in the data-all; 733 samples in the training set and 183 samples in the test set in the data-manufacturing; 597 samples in the training set and 149 samples in the test set in the data-marketing; 175 samples in the training set and 44 samples in the test set in the data-C&C.

**Data pre-process and feature selection.** In the data pre-process, the raw data set is either normalised or scaled with different formulas. This preliminary process is carried out to avoid numerical difficulties during the calculation, as well as to prevent the large values of the variables from being affected by the small values. In this study, the financial ratios which we consider as features, are scaled using equation

$$v^* = \frac{v - \min a}{\max a - \min a} \tag{13}$$

where $v$ is raw value, $v^*$ is scaled value, $\min a$ and $\max a$ are upper and lower bounds of the feature value, respectively.

The feature selection process is an important process that eliminates the difficulty of working with high-dimensional data analysis. It is aimed to better understand the data set by removing irrelevant and unnecessary features from the raw data set, working with data containing higher quality information and reducing the computation time. Three different approaches can be preferred for the feature selection process: filter, wrapper and embedded. In the filter models, generally, basic statistical methods ($t$-test, ANOVA, correlation matrix, etc.) are used and unnecessary features are eliminated by considering the relationship between features and class labels. In the wrapper models, where mostly

meta-heuristic methods are used, the calculation process is longer because it is based on the combinatorial calculation of the features. In the embedded models, the classifier both establishes the model and automatically reveals important features during the training process [67]. For the regression-based models, Lasso correction [68] and elastic net [69] approaches can be given as examples. In the correlation-based future selection, if two of the features are highly correlated with each other, then they have a similar effect on the class (dependent) variable. In such cases, one of the features may be excluded from the analysis. According to Hall and Smith [70], while a good feature subset has a high correlation with the class, they have a lower correlation with each other.

In this study, we observe a very high correlation between financial ratios, which are considered features, and therefore one of the features with a correlation coefficient above 0.70, is removed from the data set by using the Pearson correlation coefficient. The lasso correction is used for the LRA chosen as the classifier method.

**Performance evaluation.** A typical way to measure the prediction accuracy of models is the accuracy rate, which is calculated from the ratio of correctly classified samples to all observations. However, it is not healthy to compare the accuracy rate alone. The number of correctly classified positive samples and correctly classified negative samples also reveals the performance of the model in more detail. For this reason, performances are also evaluated with the Matthews correlation coefficient (MCC) evaluation criteria, which include all elements of the and confusion matrix, sensitivity, specificity, precision except for accuracy.

The classified output samples are examined in four categories that make up the confusion matrix elements. These are true positive (TP), estimating the risky loans as risky, false negative (FN), estimating the risky loans risk-free, true negative (TN), risk-free estimating of the risky loans, and false positive (FP), estimating loans that are risk-free as risky.

The formulas for these performance measures are as follows:

$$\begin{aligned}
\text{Accuracy} &= \frac{TP+TN}{TP+FP+FN+TN} \\
\text{Sensivity} &= \frac{TP}{TP+FN} \\
\text{Specificity} &= \frac{TN}{TN+FP} \\
\text{Precision} &= \frac{TP}{TP+FP} \\
MCC &= \frac{(TP\times FN)-(FP\times FN)}{\sqrt{(TP+FP)\times(TP+FN)\times(TN+FP)\times(TN+FN)}}
\end{aligned} \tag{14}$$

### 4.2. Experimental results

In this section, LRA, DT, SVM, KNN, RF, and XGBoost classifiers, which are preferred as estimation methods, are applied to four different data sets. Optimal parameters are searched for each method and the optimal parameters are decided with a ten-fold cross-validation rate to eliminate data dependency. Hyper-parameters are a very important preliminary process that affects the classification accuracy of machine learning methods determined by the model constructor. Although different methods can be used to determine hyper-parameters, an effective and easy way is Grid-Search. With this method, a combination of all possible values in the appropriate range is created for the parameters that need to be adjusted, and the most appropriate one is decided for each combination with a cross-validation rate of *k*-10 fold. Table 4 presents the optimal parameter ranges and optimal parameters selected for each method and each data set.

Table 4. The ranges of hyper-parameters and selected values

| Classifier | Hyper-parameter | Selected ranges | Best parameter | | | |
|---|---|---|---|---|---|---|
| | | | Data-all | Data -manufacturing | Data -marketing | Data -C&C |
| SVM | $C$ | $2^{-5,-3,\ldots,15}$ | 0.5 | 8192 | 8 | 2 |
| | $\gamma$ | $2^{-15,-13,\ldots,3}$ | 0.0313 | 0.0004 | 0.0078 | 0 |
| KNN | number of neighbours | 1, 2, …, 50 | 21 | 22 | 19 | 32 |
| | distance metric | Euclidean, cityblock, Chebychev | cityblock | Euclidean | cityblock | Euclidean |
| DT | min. leaf size | 10, 20, …, 100 | 60 | 77 | 60 | 46 |
| | split criteria | Gini diversity index | | | | |
| LRA | lambda | $10^{-1,-2,\ldots,-5}$ | 0.0093 | 0.0093 | 0.0145 | 0.022 |
| | penalty | Lasso | | | | |
| RF | *n*_estimators | 200, 300, .., 2000 | 200 | 300 | 1700 | 500 |
| | max_features | auto, sqrt | sqrt | auto | sqrt | auto |
| | max_depth | 10, 20 , ..., 110, none | 110 | 80 | none | 70 |
| | min_samples_split | 2, 5, 10 | 2 | 10 | 2 | 2 |
| | min_samples_leaf | 1, 2, 4, 5, 10 | 10 | 5 | 5 | 1 |
| | bootstrap | true, false | true | true | true | true |
| XGBoost | max_depth | 5, 10, 50, 60, 70, 80 | 10 | 50 | 50 | 10 |
| | learning_rate | 0.01, 0.1, 0.5, 1, 2 | 0.01 | 0.01 | 0.1 | 0.1 |
| | *n*_estimators | 100, 200, 300, 400, 500, 600, 700 | 200 | 200 | 100 | 200 |
| | gamma | 0.1, 1, 10 | 10 | 10 | 0,01 | 0,1 |

After determining the optimal parameters, each prediction model is repeated 100 times using these parameters. The results of the estimation accuracy on the data sets in which the credit rating data set is created both holistically and on a sectoral basis are

presented in Table 5. The mean and standard deviation values of the accuracy, sensitivity, specificity, precision, and MCC values obtained from each model are reported. According to these results, SVM is the method that gave the highest accuracy rate with 74.17% on the data set named Data-All, where the sectors are evaluated together. It was observed that SVM has the highest value in the specificity (68.39%) criteria.

Table 5. Performances on data sets

| Data | Method | Accuracy | Sensitivity | Specificity | Precision | MCC |
|---|---|---|---|---|---|---|
| Data-all | SVM | 0.7417 (0.0125) | 0.7557 (0.0143) | 0.6839 (0.0467) | 0.9119 (0.0187) | 0.3640 (0.0333) |
| | KNN | 0.7327 (0.0198) | 0.7581 (0.0264) | 0.6451 (0.0503) | 0.8833 (0.0210) | 0.3546 (0.0427) |
| | DT | 0.7171 (0.0246) | 0.7644 (0.0295) | 0.5869 (0.0549) | 0.8405 (0.0384) | 0.3261 (0.0563) |
| | LRA | 0.7402 (0.0184) | 0.7528 (0.0201) | 0.6836 (0.0574) | 0.9158 (0.0181) | 0.3566 (0.0455) |
| | RF | 0.7387 (0.0160) | 0.7642 (0.0199) | 0.6505 (0.0387) | 0.8859 (0.0188) | 0.3652 (0.0357) |
| | XGBoost | 0.7410 (0.0184) | 0.7683 (0.0221) | 0.6470 (0.0440) | 0.8841 (0.0206) | 0.3680 (0.0390) |
| Data-manufacturing | SVM | 0.8172 (0.0217) | 0.8286 (0.0238) | 0.7199 (0.0841) | 0.9375 (0.0154) | 0.38174 (0.0567) |
| | KNN | 0.8097 (0.0271) | 0.8199 (0.0298) | 0.6994 (0.1191) | 0.9691 (0.0156) | 0.3333 (0.0778) |
| | DT | 0.8169 (0.0257) | 0.8561 (0.0262) | 0.6157 (0.1018) | 0.9207 (0.0251) | 0.4154 (0.0842) |
| | LRA | 0.8280 (0.0236) | 0.8362 (0.0240) | 0.7611 (0.0924) | 0.9671 (0.0144) | 0.43644 (0.0735) |
| | RF | 0.8147 (0.0203) | 0.8419 (0.0232) | 0.6436 (0.0780) | 0.9382 (0.0182) | 0.3949 (0.0573) |
| | XGBoost | 0.8159 (0.0211) | 0.8401 (0.0248) | 0.6621 (0.0802) | 0.9426 (0.0199) | 0.4007 (0.0582) |
| Data-marketing | SVM | 0.6890 (0.0265) | 0.6984 (0.0376) | 0.6735 (0.0476) | 0.8101 (0.0350) | 0.3542 (0.0516) |
| | KNN | 0.6845 (0.0313) | 0.7098 (0.0431) | 0.6468 (0.0625) | 0.7640 (0.0492) | 0.3498 (0.0646) |
| | DT | 0.6710 (0.0346) | 0.6905 (0.0488) | 0.6453 (0.0788) | 0.7805 (0.0762) | 0.3165 (0.0739) |
| | LRA | 0.6911 (0.0331) | 0.6955 (0.0416) | 0.6811 (0.0602) | 0.8308 (0.0335) | 0.3524 (0.0692) |
| | RF | 0.6837 (0.0269) | 0.7069 (0.0354) | 0.6460 (0.0533) | 0.7783 (0.0418) | 0.3430 (0.0554) |
| | XGBoost | 0.6729 (0.0253) | 0.7046 (0.0376) | 0.6276 (0.0472) | 0.7519 (0.0384) | 0.3235 (0.0529) |

Table 5. Performances on data sets

| Data | Method | Accuracy | Sensitivity | Specificity | Precision | MCC |
|---|---|---|---|---|---|---|
| Data-C&C | SVM | 0.6221 (0.0606) | 0.6369 (0.0807) | 0.6269 (0.1245) | 0.8539 (0.0892) | 0.2228 (0.0958) |
| | KNN | 0.6247 (0.0707) | 0.6361 (0.0851) | 0.6277 (0.1788) | 0.8507 (0.0854) | 0.2168 (0.1493) |
| | DT | 0.6102 (0.0775) | 0.6363 (0.0837) | 0.5658 (0.1381) | 0.7485 (0.1125) | 0.1884 (0.1601) |
| | LRA | 0.6616 (0.0740) | 0.6535 (0.0829) | 0.6887 (0.1301) | 0.8634 (0.0593) | 0.3019 (0.1466) |
| | RF | 0.5920 (0.0474) | 0.6290 (0.0635) | 0.5386 (0.0955) | 0.7009 (0.0871) | 0.1613 (0.0996) |
| | XGBoost | 0.5944 (0.0512) | 0.6452 (0.0650) | 0.5269 (0.0855) | 0.6781 (0.0855) | 0.1681 (0.1042) |

The most dominant method for data-manufacturing, data-marketing, and data-C&C data sets, is LRA. In these data sets, LRA obtains accuracy values of 82.80%, 69.11%, and 66.16%, respectively.

Table 6. Post Hoc results of data-all

| Classification | LRA | KNN | DT | RF | XGBoost |
|---|---|---|---|---|---|
| SVM | –0.015[a] | 0.009[a] | 0.024[a] | 0.003[a] | 0.000[a] |
| | 0.992 | 0.009[b] | 0.000[b] | 0.885 | 0.999 |
| LRA | – | 0.007[a] | 0.023[a] | 0.001[a] | –0.007[a] |
| | | 0.53 | 0.000[b] | 0.992 | 0.999 |
| KNN | – | – | 0.015[a] | –0.005[a] | –0.008[a] |
| | | | 0.000[b] | 0.214 | 0.024[b] |
| DT | – | – | – | –0.021[a] | –0.023[a] |
| | | | | 0.000[b] | 0.000[b] |
| RF | – | – | – | – | –0.002[a] |
| | | | | | 0.958 |

[a]Mean difference of Tukey HSD test.
[b]Significance at 0.05.
$F_{5.594}$ is 25.667, $p$ value 0.000.

Considering the sensitivity criterion, it is observed that XGBoost, DT, and KNN methods came to the fore this time, especially in data-all and in data-manufacturing data sets. On the other hand, considering the specificity, precision, and MCC criteria, it is concluded that the dominant methods are again LRA and SVM.

In the data-all dataset, in which all sectors are evaluated together and a nominal variable is assigned for each sector, the classification accuracy rate of each method is found to be approximately 73%, and no significant difference is observed between the classification success of SVM, LRA, RF, and XGBoost methods. When the data set is considered on a sector basis, the accuracy rate for the manufacturing sector increases to approximately 82%. However, the accuracy rate in the marketing and construction and contracting sectors is approximately 68% and 63%, respectively. In summary, dividing the data into sub-categories has increased the forecasting success for some sectors, while decreasing it for others.

According to the results of the ANOVA test performed to determine whether there is a difference in accuracy rates between the methods in the Data All data set, it is determined that there was a statistically significant difference ($F$ = 25.667 and $p$-value = 0.000) between the methods. Post-hoc (Tukey – HSD) was applied to find out which method caused this difference, and the test results are given in Table 6. According to Table 6, while there is no difference between SVM, LRA, RF, and XGBoost methods, there is a difference between the KNN method, and the SVM and XGBoost methods. However, there is a significant difference between DT and all other methods considered.

## 5. Conclusion

Credit rating systems need to be developed, since providing consumer loans has some risks for banks. This article focuses on comparing six different well-known machine learning models for the credit rating problem.

In the light of the information obtained after a detailed literature review on machine learning methods to be used to predict the credit ratings of companies, it has been determined that the most frequently used methods are LRA, SVM, DT, KNN, RF, and XGBoost. After additional applications related to data pre-processing and optimal feature sets that will increase the performance of machine learning methods, main estimation methods are applied and the results of the experiments performed on four different data sets reveal that the best performing methods were SVM, LRA, RF, and XGBoost methods. This result is supported by statistical tests.

The credit rating data set used in the study consists of three leading sectors. In this study, the importance of making sector-based analysis is emphasized and analyses are made separately for each sector. Separating data into subcategories increases the accuracy rate in some sectors while decreasing it in others. Finding out where or why this difference originates can be examined as another research topic.

Many studies in the literature on diversification used the method to search for higher accuracy rates. According to the results of this study, in addition to the power of the

method used, it is thought that a better understanding of the structure of the data during the data pre-processing stage will provide more benefits in increasing the accuracy rate.

## References


[1] *Moody's Symbols and* Definitions, available from, https,//www.moodys.com/researchdocumentcontent page.aspx?docid=PBC_79004

[2] HUANG Z., CHEN H., HSU C.J., CHEN W.H., WU S., *Credit rating analysis with support vector machines and neural networks: a market comparative study*, Dec. Sup. Syst., 2004, 37, (4), 543–558.

[3] BALIOS D., THOMADAKIS S., TSIPOURI L., *Credit rating model development. An ordered analysis based on accounting data*, Res. Int. Bus. Fin., 2016, 38, 122–136.

[4] BLÖCHLINGER A., *Credit rating and pricing. Poles apart*, J. Risk Fin. Manage., 2018, 11 (2).

[5] AKBULAK Y., *Credit rating or rating. Concept and criteria*, Mali Çözüm Derg., 2012, 111, 171–184 (in Turkish).

[6] WÓJCICKA-WÓJTOWICZ A., *Credit risk mangement in finance-a review of various approaches.* Oper. Res. Dec., 2018, 28 (4), 99–106.

[7] LI J.P., MIRZA N., RAHAT B., XIONG D., *Machine learning and credit ratings prediction in the age of fourth industrial revolution*, Techn. For. Soc. Change, 2020, 161, 120309.

[8] KARIYA, T., YAMAMURA Y., INUI K., *Empirical credit risk ratings of individual corporate bonds and derivation of term structures of default probabilities*, J. Risk Fin. Manage., 2019, 12 (3).

[9] ZHONG H., MIAO C., SHEN Z., FENG Y., *Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings*, Neurocomp., 2014, 128, 285–295.

[10] HUANG, C.-L., CHEN M.-C., WANG C.-J., *Credit scoring with a data mining approach based on support vector machines*, Exp. Syst. Appl., 2007, 33 (4), 847–856.

[11] SHI J., XU B., *Credit Scoring by fuzzy support vector machines with a novel membership function*, J. Risk Fin. Manage., 2016, 9 (4), 13.

[12] DANENAS , GARSVA G., *Selection of support vector machines based classifiers for credit risk domain*, Exp. Syst. Appl., 2015, 42 (6), 3194–3204.

[13] YANENKOVA I., NEHODA Y.,DROBYAZKO S., ZAVHORODNII A., BEREZOVSKA L., *Modeling of bank credit risk management using the cost risk model*, J. Risk Fin. Manage., 2021, 14 (5), 211.

[14] HAJEK, , MICHALAK K., *Feature selection in corporate credit rating prediction*, Knowl.-Based Syst., 2013, 51, 72–84.

[15] HAN S.H., PAGANO M.S., SHIN Y.S., *Rating agency reputation, the global financial crisis, and the cost of debt*, Fin. Manage., 2012, 41 (4), 849–884.

[16] GÜR T.H., ÖZTÜRK H., *Country risk, rating agencies, disruptions and new regulations*, Sosyoekon., 2011 (2), 69–92, 238–240 (in Turkish).

[17] NICA I., ALEXANDRU D.B., PARAMON CRĂCIUNESCU S.L., IONESCU S., *Automated valuation modelling: analysing mortgage behavioural life profile models using machine learning techniques*, MDPI, 2021, 13 (9), 5162.

[18] ADDO, M., GUEGAN D., HASSANI B., *Credit risk analysis using machine and deep learning models*, 2018, 6 (2), 38.

[19] ASSEF F. M., STEINER M., NETO S., DE BARROS FRANCO D.G., *Classification algorithms in financial application. Credit risk analysis on legal entities*, IEEE Latin Am. Trans., 2019, 17 (10), 1733–1740.

[20] ANICETO, M.C., BARBOZA F., KIMURA H., *Machine learning predictivity applied to consumer credit-worthiness*, Fut.e Bus. J., 2020, 6 (1), 37.

[21] CARBO-VALVERDE S., CUADROS-SOLAS P., RODRÍGUEZ-FERNÁNDEZ F.J.O., *A machine learning approach to the digitalization of bank customers. Evidence from random and causal forests*, PLOS ONE, 2020, 15 (10), e0240362.
[22] FISHER L., *Determinants of risk premiums on corporate bonds*, J. Pol. Econ., 1959, 67 (3), 217–237.
[23] PINCHES G.E. MINGO K.A., *A Multivariate analysis of industrial bond ratings*, J. Fin., 1973, 28 (1), 1–18.
[24] EDERINGTON, L.H., *Classification models and bond ratings*, Fin. Rev., 1985, 20 (4), 237–262.
[25] GENTRY J.A., WHITFORD D.T., NEWBOLD P., *Predicting industrial bond ratings with a probit model and funds flow components*, Fin. Rev., 1988, 23 (3), 269–286.
[26] KOZODOI N., LESSMANN S., PAPAKONSTANTINOU K., GATSOULIS Y., *A multi-objective approach for profit-driven feature selection in credit scoring*, Dec. Supp. Syst., 2019, 120, 106–117.
[27] DUTTA, S., SHEKHAR S., *Bond rating. A non-conservative application of neural networks*, Proc. IEEE Int. Conf. Neural Networks, 1988.
[28] SURKAN A.J., SINGLETON J.C., *Neural networks for bond rating improved by multiple hidden layers*, IJCNN Int. Joint Conf. Neural Networks, IEEE, 1990.
[29] DAMRONGSAKMETHEE T., NEAGOE V.-E., *C4.5 decision tree enhanced with adaboost versus multilayer perceptron for credit scoring modeling*, [In:] *Computational Statistics and Mathematical Modeling Methods in Intelligent Systems*, Springer, Cham 2019.
[30] LUO C., *A comprehensive decision support approach for credit scoring*, Ind. Manage. Data Syst., 2020, 120 (2), 280–290.
[31] KIM J.W., WEISTROFFER H.R., REDMOND R.T., *Expert systems for bond rating. A comparative analysis of statistical, rule-based and neural network systems*, Exp. Syst., 1993, 10 (3), 167–172.
[32] SHIN K.-S., HAN I., *A case-based approach using inductive indexing for corporate bond rating*, Dec. Supp. Syst., 2001, 32 (1), 41–52.
[33] KALAIVANI , SHUNMUGANATHAN K., *Feature selection based on genetic algorithm and hybrid model for sentiment polarity classification*, Int. J. Data Min., Model. Manage., 2016, 8 (4), 315–329.
[34] ZEKIC-SUSAC M., SARLIJA N., BENSIC M., *Small business credit scoring: a comparison of logistic regression, neural network, and decision tree models*, IEEE Proc. 26th Int. Conf. Information Technology Interfaces, 2004.
[35] LESSMANN S., BAESENS B., SEOW H.-V., THOMAS L.C., *State-of-the-art classification algorithms for credit scoring*, J. Oper. Res. Soc., 2003, 54 (6), 627–635.
[36] LEE Y.-C., *Application of support vector machines to corporate credit rating prediction*, Exp. Syst. Appl., 2007, 33 (1), 67–74.
[37] KIM K.-S., HAN I., *The cluster-indexing method for case-based reasoning using self-organizing maps and learning vector quantization for bond rating cases*, Exp. Syst. Appl., 2001, 21 (3), 147–156.
[38] YU L., WANG S., LAI K.K., *Credit risk assessment with a multistage neural network ensemble learning approach*, Expert Systems with Applications, 2008, 34 (2), 1434–1444.
[39] KIM K.-J., AHN H., *A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach*, Comp. Oper. Res., 2012, 39 (8), 1800–1811.
[40] YEH C.-C., LIN F., HSU C.-Y., *A hybrid KMV model, random forests and rough set theory approach for credit rating*, Knowledge-Based Systems, 2012, 33, 166–172.
[41] PAI -F., TAN Y.-S., HSU M.-F., *Credit rating analysis by the decision-tree support vector machine with ensemble strategies*, Int. J. Fuzzy Syst., 2015, 17 (4), 521–530.
[42] ADDO M., GUEGAN D., HASSANI B., *Credit risk analysis using machine and deep learning models*, Risks, 2018, 6 (2), 38.
[43] WALLIS M., KUMAR K., GEPP A., *Credit rating forecasting using machine learning techniques*, Adv. Data Min. Datab. Manage. Manage. Persp. Int. Big Anal., 2019, 180–198.
[44] GRISHUNIN S., SULOEVA S., EGOROVA A., BUROVA B., *Comparison of empirical methods for the reproduction of global manufacturing companies' credit ratings*, Int. J. Technol., 2020, 11, 1223.

[45] GOLBAYANI P., FLORESCU I., CHATTERJEE R., *A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees*, North Am. J. Econ. Fin., 2020, 54C, 101251.
[46] MERĆEP A., MRČELA L., BIROV M., KOSTANJČAR Z., *Deep neural networks for behavioral credit rating*, Entropy (Basel), 2020, 23 (1), 27.
[47] HOSMER, D.W., LEMESHOW S., STURDIVANT R.X., *Applied Logistic Regression*, Wiley, New York 2000.
[48] NETER J., *Applied Linear Statistical Models*, McGraw-Hill, Irwin, Boston 1996.
[49] MUSA A.B., *Comparative study on classification performance between support vector machine and logistic regression*, Int. J. Mach. Learn. Cyber., 2013, 4 (1), 13–24.
[50] FIX E., HODGES J.L., *Discriminatory analysis – nonparametric discrimination. Consistency properties*, International Statistical Review, 1989, 57, 238.
[51] HAN J., KAMBER M., *Data mining concepts and techniques*, [In:] J. Gray (Ed.), *The Morgan Kaufmann Series in Data Management Systems*, Elsevier, 2006.
[52] CORTES C., VAPNIK V., *Support-vector networks*, Mach. Learn., 1995, 20 (3), 273–297.
[53] CRISTIANINI N., SHAWE-TAYLOR J., *An introduction to support vector machines and other kernel-based learning methods*, Cambridge University Press, Camridge 2000.
[54] BURGES C.J.C., *A tutorial on support vector machines for pattern recognition*, Data Min. Knowl. Disc., 1998, 2 (2), 121–167.
[55] LIAO Y., FANG S.-C., NUTTLE H.L.W., *A neural network model with bounded-weights for pattern classification*, Comp. Oper. Res., 2004, 31 (9), 1411–1426.
[56] SONG Y.-Y., LU Y., *Decision tree methods: applications for classification and prediction*, Shang. Arch. Psych., 2015, 27 (2), 130–135.
[57] BREIMAN L., FRIEDMAN J.H., OLSHEN R.A., STONE C.J., *Classification and regression trees*, Brooks /Cole Publishing, Monterey 1984.
[58] QUINLAN J.R., *C4.5: programs for machine learning*, Morgan Kaufmann Publishers, Inc., 1993.
[59] BIGGS D., DE VILLE B., SUEN E., *A method of choosing multiway partitions for classification and decision trees*, J. Appl. Stat., 1991, 18 (1), 49–62.
[60] KASS G.V., *An exploratory technique for investigating large quantities of categorical data*, J. Royal Stat. Soc., Ser. C (Appl. Stat.), 1980, 29 (2), 119–127.
[61] SAMUEL O.S., OLAJUBU E., *Supervised machine learning algorithms. Classification and comparison*, Int. J. Comp. Trends Technol., 2017, 48, 128–138.
[62] PATEL N.D. UPADHYAY S., *Study of various decision tree pruning methods with their empirical comparison in WEKA*, Int. J. Comp. Appl., 2012, 60, 20–25.
[63] ARAS S., *Stacking hybrid GARCH models for forecasting bitcoin volatility*, Exp. Syst. Appl., 2021, 174, 114747.
[64] FRAIWAN L., LWEESY K., KHASAWNEH N., WENZ H., DICKHAUS H., *Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier*, Comp. Meth. Progr. Biomed., 2012, 108 (1), 10–19.
[65] CHEN T., GUESTRIN K.D., *XGBoost. A scalable tree boosting system*, https://www.kdd.org /kdd2016 /papers/files/rfp0697-chenAemb.pdf 2016
[66] CHANG C.-C., LIN C.-J., *LIBSVM: A library for support vector machines*, ACM Trans. Intell. Syst. Technol., 2011, 2 (27), 1–27.
[67] CAI J., JAIWEI L., SHULIN W., SHENG Y., *Feature selection in machine learning. A new perspective*, Neurocomp., 2018, 300, 70–79.
[68] TIBSHIRANI R., *Regression shrinkage and selection via the lasso*, J. Royal Stat. Soc., Ser. B (Meth.), 1996, 58 (1), 267–288.

[69] ZOU H., *The adaptive lasso and its oracle properties*, J. Am. Stat. Assoc., 2006, 101 (476), 1418–1429.
[70] HALL M.A., SMITH L.A., *Feature selection for machine learning: Comparing a correlation-based filter approach to the wrapper*, FLAIRS Conference, 1999.

# Appendix

Table A1. Descriptive statistics of financial ratios used in the study

| Financial ratio | Mean | Standard deviation | Minimum | 25% | 50% | 75% | Maximum |
|---|---|---|---|---|---|---|---|
| X2 | 3.880 | 28.807 | –197.615 | 1.010 | 1.220 | 1.735 | 746.508 |
| X3 | 1.245 | 7.199 | –39.142 | 0.220 | 0.563 | 0.905 | 225.185 |
| X4 | 0.516 | 7.188 | –102.896 | 0.010 | 0.038 | 0.127 | 225.185 |
| X5 | 0.486 | 0.316 | –5.541 | 0.265 | 0.460 | 0.716 | 2.153 |
| X8 | 0.164 | 0.268 | –2.296 | 0.007 | 0.138 | 0.315 | 0.978 |
| X9 | 0.664 | 0.241 | 0.001 | 0.519 | 0.709 | 0.843 | 3.442 |
| X11 | 1.257 | 103.658 | –4258.558 | 1.025 | 2.328 | 5.053 | 178.889 |
| X12 | 0.576 | 0.258 | –0.005 | 0.397 | 0.609 | 0.780 | 2.905 |
| X14 | 0.884 | 5.317 | –0.004 | 0.136 | 0.321 | 0.669 | 182.391 |
| X15 | 0.385 | 30.411 | –1299.977 | 0.357 | 0.707 | 1.277 | 85.014 |
| X16 | 0.598 | 6.740 | –219.465 | 0.298 | 0.581 | 0.956 | 46.346 |
| X17 | 0.873 | 0.218 | –0.007 | 0.825 | 0.995 | 1.000 | 1.012 |
| X18 | 0.740 | 0.203 | –0.133 | 0.611 | 0.789 | 0.901 | 1.004 |
| X21 | 3.757 | 10.873 | –280.257 | 1.620 | 2.511 | 4.175 | 273.168 |
| X22 | 27.965 | 695.527 | –7748.836 | 0.857 | 6.224 | 17.405 | 23607.645 |
| X23 | –61.598 | 4958.618 | –214504.0 | 4.956 | 11.782 | 28.494 | 9636.940 |
| X24 | 45.173 | 332.202 | –130.095 | 4.257 | 9.553 | 23.210 | 9636.940 |
| X25 | 16.096 | 134.079 | –436.772 | 3.205 | 6.668 | 13.478 | 5317.810 |
| X26 | 2.560 | 3.393 | 0.004 | 1.154 | 1.820 | 2.884 | 85.690 |
| X27 | –0.054 | 6.524 | –219.511 | 0.047 | 0.121 | 0.242 | 23.023 |
| X28 | –0.058 | 7.727 | –233.744 | 0.061 | 0.153 | 0.296 | 23.468 |
| X29 | 0.072 | 0.230 | –0.586 | 0.020 | 0.046 | 0.092 | 8.683 |
| X30 | 0.053 | 0.226 | –0.586 | 0.013 | 0.031 | 0.065 | 8.683 |
| X31 | 0.079 | 0.231 | –0.812 | 0.024 | 0.056 | 0.101 | 8.592 |
| X32 | 0.039 | 0.102 | –1.700 | 0.014 | 0.034 | 0.063 | 0.936 |
| X33 | 0.149 | 0.125 | –1.368 | 0.077 | 0.123 | 0.201 | 0.991 |
| X34 | 0.024 | 0.105 | –2.171 | 0.008 | 0.019 | 0.039 | 0.946 |
| X37 | 0.030 | 0.109 | –2.171 | 0.010 | 0.023 | 0.047 | 0.946 |
| X39 | 0.849 | 7.420 | –0.999 | 0.019 | 0.258 | 0.635 | 288.982 |