STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 11, Issue 3, 2024
September–December
DOI Journal: 10.37591/JoAIRA

https://journals.stmjournals.com/joaira

Review JoAIRA

Evaluation of Credit Risk of Bank Customers with a Hybrid Approach of Data Mining Techniques

Mehrnaz Bagheri[1,*], Mohammad Taghipour[2]

Abstract

Credit risk poses the most significant threat to financial and monetary institutions. Banks strive to offer loans that generate high returns while minimizing risk. Achieving this requires the ability to accurately identify and classify credit customers, both individuals and legal entities, according to their likelihood of fully meeting their obligations. This classification is done using relevant financial and non-financial criteria. The primary goal of this study is to assess the factors that influence the evaluation of bank customers' creditworthiness through a hybrid data mining approach, thereby enhancing decision-making processes. To accomplish this, financial and qualitative data from 1,000 customer samples were sourced from the UCI University site, utilizing 24 explanatory variables. Various prediction models, including decision trees, the Naïve Bayes algorithm, support vector machines (SVM), neural networks, and combinations of these techniques, were employed to predict the risk of loan default. Additionally, the factors influencing customer creditworthiness were ranked. The findings indicated that the Naïve Bayes algorithm outperformed the decision tree method in predicting loan repayment risk. Furthermore, the key variables impacting customer validation were ranked using these techniques, highlighting the superior effectiveness of the Naïve Bayes algorithm in guiding banks' decisions on credit issuance.

Keywords: Credit risk, data mining, decision tree, Naïve Bayesian algorithm, support vector machine (SVM), neural networks

INTRODUCTION

Global competition, dynamic markets, and innovation and technology are reducing, and this is a big challenge for the banking industry. We should use support systems to improve decision-making processes in these organizations. One of the major problems of banking and financial systems is credit risk management as many monetary resources in these institutes are presented as credit to the people and the return of these resources depends upon the development of institutes. Thus, the evaluation of credit of customers to pay the loan is an important process and various methods are presented [1].

*Author for Correspondence
Mehrnaz Bagheri
E-mail: mehrnaz.bagheri000@gmail.com

[1]Master, Department of Industrial Engineering, ABA Institute of Higher Education, Qazvin, Iran
[2]Assistant Professor, Department of Industrial Engineering, North Tehran Branch, Islamic Azad University, Tehran, Iran

Received Date: August 08, 2024
Accepted Date: August 23, 2024
Published Date: September 11, 2024

Citation: Mehrnaz Bagheri, Mohammad Taghipour. Evaluation of Credit Risk of Bank Customers with a Hybrid Approach of Data Mining Techniques. Journal of Artificial Intelligence Research & Advances. 2024; 11(3): 63–81p.

Decisions on granting facilities to Iranian bank clients were formerly made conventionally, with individual assessments of the risk of non-repayment serving as the basis [2]. However, the nation's banks, financial institutions, and households use more sophisticated techniques, such as statistical methods, to lower the risk of facility non-repayment because of the growing demand for bank facilities from these groups of people and businesses, as well as the rise in intense commercial competition. This is the situation. Currently, banks categorize applicants and forecast the likelihood of a short payback of facilities based on the credit ratings of their customers. Exploiting advanced technologies such as data mining will certainly be a constant challenge for bank professionals. This is because they are creative and strive to innovate. As a result, it is realistic to employ data mining techniques to verify Mob Bank customers [3]. This research uses the creation of a credit risk prediction system with the help of a neural network, decision tree, simple Bayes algorithm, and Pashtian SVM, which saves time, and money, eliminates personal judgments, and increases accuracy. All types of facilities have advantages in evaluating applicants.

PROBLEM STATEMENT

To be aware of the requirements and behavior of customers in giving credit loans, banks should identify their features. This leads to a reduction in bank risk including credit risk. Various studies have validated the identification of good and bad customer accounts in banks. Credit risk is one of the most important risks affecting monetary and financial institutions. Credit risk is the risk by which the loan receiver cannot pay out the principal and interest of the loan in accordance with the contract terms. In other words, according to this risk, payouts are performed with delay, or they are not collected. This leads to problems with cash flow turnover in banks. Despite the existing innovations in the financial service sector, this type of risk is the major reason for the failure of financial institutions. Normally, 80% of a bank's balance sheet refers to some aspects of this type of risk. Credit risk is one of the main causes of bank bankruptcy. Based on the economic structure of our country and for reasons such as the lack of development of capital markets and other non-bank networks, funding of real sectors of the economy is dedicated to the bank network of the country. Banks attempt to provide loans to companies with high returns and low-risk. This is fulfilled if banks can identify their credit customers, including real and legal entities, and classify them based on their tendency to pay out of obligations using appropriate financial and non-financial criteria. Under such a system, loans are given to those with low credit risk and high payout potential. These assets can be used as financial sources for subsequent loans; they can play an important role in the increase in investment, growth, and economic development. Data mining is used to solve certain problems, and it uses fields such as databases, artificial intelligence, machine learning, neural networks, statistics, model identification, knowledge-based systems, knowledge acquisition, data recovery, high-speed computation, and data visual representation. Valuable data are obtained via a data mining technique. The data mining process involves knowledge extraction of existing data and turning it into an understandable structure to use the data. Data mining, as a special technique to obtain useful information, has been of great importance in recent years. Data mining is the process of defining existing models in the data. These models are valid, unique, and potentially useful. In other words, data mining techniques are used to find interesting models in databases hidden in large amounts of data.

IMPORTANCE AND NECESSITY OF RESEARCH

Since 1950, when the computer was used in data analysis and storage, the volume of information stored doubled after approximately 20 years and doubled once every 2 years as the data volume in the database improved. It is still faster than ever before the volume of stored information has increased in database technology, and its extensive use in various applications has led to a large amount of data collection. These data have created the need for powerful tools for analyzing past data. Because we are currently wealthy, we lack information, so the need to extract information and knowledge for decision-making is increasing day by day [4]. Obtaining information from bulk data is essential for effective management. The intensity of competition in the scientific, social, and economic fields has also doubled the importance of speed or access to information. On the other hand, there is a need for systems that can explore information rapidly with less human intervention and turn to statistical methods. It was tailored to the size of the data.

Data mining is a process that emerged at the beginning of the 1990s and deals with a new approach to information extraction from databases. In our country, organizations, companies, and institutions are increasingly building or purchasing database software and mechanizing their information systems. It shows. Therefore, given the growing growth of databases in organizations and the need to analyze these data in their decisions, it can be said that job data approaches are the most appropriate option for managers [5].

Information, if not properly managed instead of helping the company, will waste resources and confusion and disrupt the work process. Effective information can be extracted using data mining approaches, which can offer decision-making management. Data extraction is facilitated by data mining techniques, such as clustering valuable and relevant information. On the other hand, we are speaking at a time when it comes to satisfying different and growing human needs and understanding the different and complex physical and mental aspects of the human being. Given the limitations of various sources, the use of tools and methods such as categorizing individuals into different classes can help meet these needs. To find hidden information in data, use decision-makers, and decrease data volume, it is important to apply a categorizing strategy in conjunction with a combined data mining approach [6].

RESEARCH METHODOLOGY

This research is practical from the point of view of the goal, and in terms of method, it is descriptive-analytical research in terms of investigating and identifying factors affecting the risk of bank customers. Therefore, the library method was used to review the literature on the subject. In the next step, by examining library studies, studying previous research, and using the experiences of bank experts, indicators of customer credit that affect their performance in repaying facilities were identified.

Coding of the neural network and decision tree models and other data mining techniques was performed using Rapid Miner 5.3 software and by providing the criteria data required by customers as input to the developed approaches, the output results in each repetition were presented and the credit rating of customers was determined. In addition, with the help of data mining techniques, we attempted to identify the effective factors in the credit risk of bank customers. Finally, after the application of data mining, banks, and financial institutions can provide facilities to customers with more confidence and less risk. Acting yourself [7].

RESEARCH GOAL

The main purpose of this research is to evaluate the factors affecting the creditworthiness of bank customers with a combined approach of data mining to improve decision-making. The purpose of this research is divided into three groups:

1. Evaluation of effective factors on the credit of bank customers.
2. Determine which data mining technique is suitable to evaluate the credit of bank customers.
3. To get the best credit risk prediction outcomes, compare data mining methodologies.

STUDY QUESTIONS

This study evaluates bank customers’ credit risk using a hybrid approach of data mining techniques. This study aimed to answer the following questions:

1. What are the effective factors in the assessment of credit of bank customers?
2. Which of the data mining techniques is suitable to evaluate customers’ credit?

DIFFERENT TYPES OF DATA MINING

Data mining was either directed or undirected. Directed data mining has a special goal variable following a special model, but undirected data mining finds models or similarities between groups of data without having a special target goal or a set of pre-defined models and sets [8].

Data Mining Duties

What can we do with data mining? Data mining consists of six important functions and most of these issues can be considered as follows:

1. Classification
2. Estimation
3. Prediction

4. Similarity grouping
5. Clustering
6. Description and indexing

The first three cases are directed data mining, and they aim to determine the value of the variable of a special target. Similarity grouping and clustering are undirected data mining, in which the goal is to find the latent structure of data without considering a special goal. Indexing is a descriptive action that is both directed and undirected. Each is explained briefly below [9,10].

Classification

The classification and prediction of the two operations are used for data analysis and model extraction for important datasets and the prediction of their future behavior. Classification models are used in discrete and classified data analyses, while prediction or regression models work on continuous data. For instance, a classification model can be used to divide bank loans into risky and non-risky categories. Prediction models can predict customers' expenditures and costs based on their job and income features.

Classification is the process of finding a model by identifying sets or data concepts that predict the unknown sets of other objects. Classification is a learning function that maps a data item onto a pre-defined set. The existing data were divided into training and tests. The training data to learn the rules were used by the system, and the test data were used to evaluate the model accuracy. Classification models were used for discrete and classified data analyses. The classification was supervised learning.

Decision Tree

Decision tree is one of the most famous classification models. In decision-tree-based classification algorithms, the output knowledge is presented as a tree of different feature states.

Decision trees are employed in accordance with the principles of decision-making, and if we state the result of classification in the form of risky loans against low-risk loans, purchases, or non-purchases, decision trees are an efficient method.

One of the advantages of a decision tree is better recognition of important fields as in the decision tree, as important fields are automatically transferred to the upper nodes of the tree, and the decision tree ignores less important fields.

The decision tree model is a set of it and then rules and the data are completely shown in most cases.

CART ALGORITHM

This operator, with the learning capability of nominal or numerical input data, makes a decision tree as an output model. In the CART method, the Gini index parameter is selected to apply the entropy method [11, 12].

- *First model:* CART
- *Input data:* All fields except the target variable
- *Purpose:* Good customer account and bad customer account

As shown in Figures 1 and 2, a general report on the role of fields, name of fields, type of fields, statistics index, field range, and lost values can be observed.

As shown in Figure 3 and Table 1, the model accuracy during training was 75.30.

METHOD

The dataset is divided into two equal parts. One of the two parts was considered as the training data, and the model was based on it. Then, another dataset was used to evaluate the model. We have now changed the position to two datasets. The first dataset was used for evaluation, and the second dataset

Journal of Artificial Intelligence Research & Advances
Volume 11, Issue 3
ISSN: 2395-6720

for operation evaluation was used for training and modeling. The mean accuracy in the two stages is introduced as the final accuracy. This assessment method is called 2-fold cross-validation .If instead of 2 times, we do it K times instead of two times, the K-fold cross-validation method is achieved, as shown in Figure 4.

The larger the value of K, the more reliable the accuracy of each classifier, and the more comprehensive the knowledge.

As shown in Figure 5 and Table 2, the model accuracy in the test section was 71.10%.

As shown in Figure 6, the most important fields include:

- Status of existing checking account
- Duration in month
- Personal status and sex
- Property
- Credit amount
- Number of existing credits

Table 1. The criterion comparison in training in the CART model.

Recall	Precision	Class
85.43%	80.48%	1= Good
51.67%	60. 31 %	2 = Bad
	Accuracy=75.30	
	Classification Error = 24.70	

Table 2. Criterion comparison in test section in CART model.

Recall	Precision	Class
82.14%	77.81%	1= good
45.33%	52. 11 %	2 = bad
	Accuracy=71.10	
	Classification Error = 28.90	

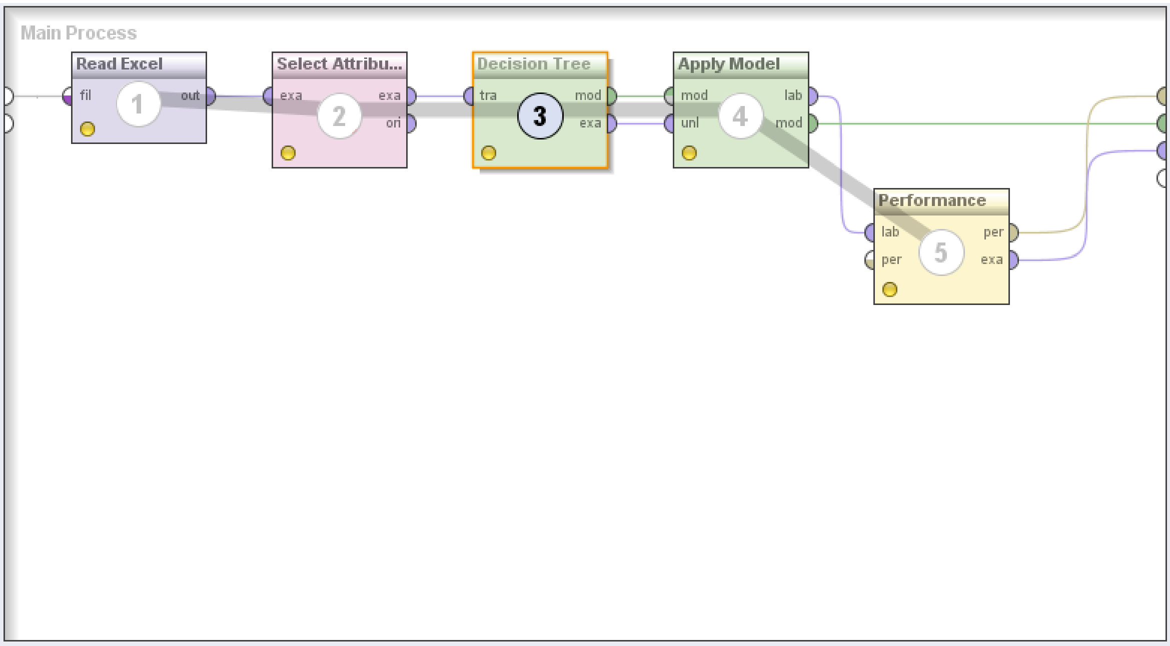

Figure 1. The structure of operators in the CART model.

Role	Name	Type	Statistics	Range	Missings
label	y	binominal	mode = 1 (700), least = 2 (300)	1 (700), 2 (300)	0
confidence_1	confidence(1)	real	avg = 0.700 +/- 0.213	[0.265 ; 0.950]	0
confidence_2	confidence(2)	real	avg = 0.300 +/- 0.213	[0.050 ; 0.735]	0
prediction	prediction(y)	binominal	mode = 1 (743), least = 2 (257)	1 (743), 2 (257)	0
regular	Status of existing checking account	polynominal	mode = 4 (394), least = 3 (63)	1 (274), 2 (269), 4 (394), 3 (63)	0
regular	Duration in month	integer	avg = 20.903 +/- 12.059	[4.000 ; 72.000]	0
regular	Credit history	polynominal	mode = 2 (530), least = 0 (40)	4 (293), 2 (530), 3 (88), 0 (40), 1 (49)	0
regular	purpose	polynominal	mode = radio/tv (280), least = retraining	radio/tv (280), education (50), furniture/ε	0
regular	Credit amount	integer	avg = 2.105 +/- 1.580	[1.000 ; 5.000]	0
regular	Savings account/bonds	polynominal	mode = 3 (339), least = 1 (62)	5 (253), 3 (339), 4 (174), 1 (62), 2 (172)	0
regular	Present employment since	polynominal	mode = 3 (548), least = 1 (50)	3 (548), 2 (310), 1 (50), 4 (92)	0
regular	Installment rate	integer	avg = 2.845 +/- 1.104	[1.000 ; 4.000]	0
regular	Personal status and sex	polynominal	mode = 3 (332), least = 4 (154)	1 (282), 2 (232), 4 (154), 3 (332)	0
regular	other parties	polynominal	mode = none (907), least = 'co applicar	none (907), guarantor (52), 'co applican	0
regular	Present residence since	integer	avg = 2.675 +/- 0.706	[1.000 ; 3.000]	0
regular	Property	polynominal	mode = 1 (633), least = 4 (6)	2 (333), 1 (633), 3 (28), 4 (6)	0
regular	Age in years	integer	avg = 1.155 +/- 0.362	[1.000 ; 2.000]	0
regular	Other installment plans	polynominal	mode = 1 (596), least = 2 (404)	2 (404), 1 (596)	0
regular	Housing	polynominal	mode = 1 (963), least = 2 (37)	1 (963), 2 (37)	0
regular	Number of existing credits	integer	avg = 0.234 +/- 0.424	[0.000 ; 1.000]	0
regular	Job	polynominal	mode = 0 (897), least = 1 (103)	0 (897), 1 (103)	0
regular	Number of people being liable	integer	avg = 0.907 +/- 0.291	[0.000 ; 1.000]	0

Figure 2. A general report of the role of fields.

accuracy: 75.30%

	true 1	true 2	class precision
pred. 1	598	145	80.48%
pred. 2	102	155	60.31%
class recall	85.43%	51.67%	

Figure 3. The model accuracy in training in CART model.

Journal of Artificial Intelligence Research & Advances
Volume 11, Issue 3
ISSN: 2395-6720

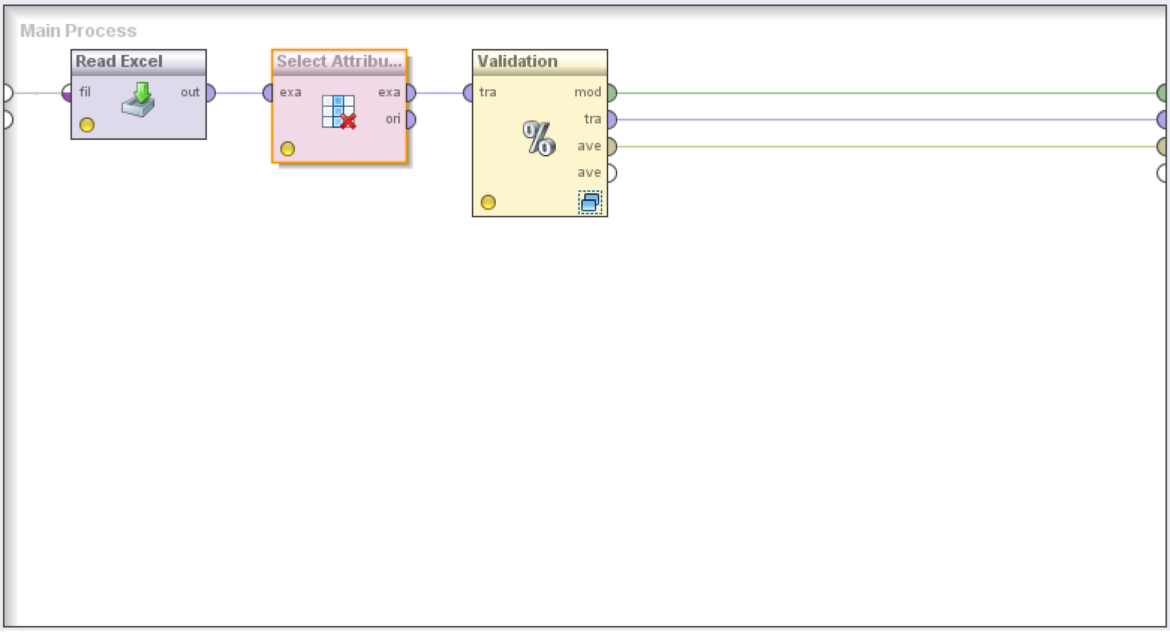

Figure 4. The structure of operators in test in CART model.

accuracy: 71.10% +/- 3.11% (mikro: 71.10%)

	true 1	true 2	class precision
pred. 1	575	164	77.81%
pred. 2	125	136	52.11%
class recall	82.14%	45.33%	

Figure 5. Model accuracy in test in CART model.

Analysis of Decision Tree

- If the Status of the existing checking account is equal to 3 or 4, then that person is in class 1.
- If the Status of the existing checking account is equal to 2, personal status, and sex 4, then that person is in class 2.
- If the Status of the existing checking account is equal to 2, personal status, and sex 2, then that person is in class 1.
- If the Status of the existing checking account is equal to 1 and the duration in month 11, then that person is in class 1.
- If the Status of the existing checking account is equal to 1, the duration in a month is bigger than 11, Duration in months is larger than 31, then that person is in Class 2.
- If the Status of the existing checking account is equal to 1, the duration in months is bigger than 11, the Duration in months is smaller than or equal to 31, and the job is equal to 1, then that person is in class 1.

NAÏVE BAYES

A probable framework is used to solve classification problems.

Naïve Bayes can create a probable model by combining the observation evidence of the real world, and occurrence probability is achieved via the set of irrelevant features, as shown in Figure 7.

- *Second model:* Naïve Bayes
- *Input data:* All fields except the target variable
- *Purpose:* Good customer account and bad customer account

As shown in Figure 8 and Table 3, the model accuracy for the training was 75.80%.

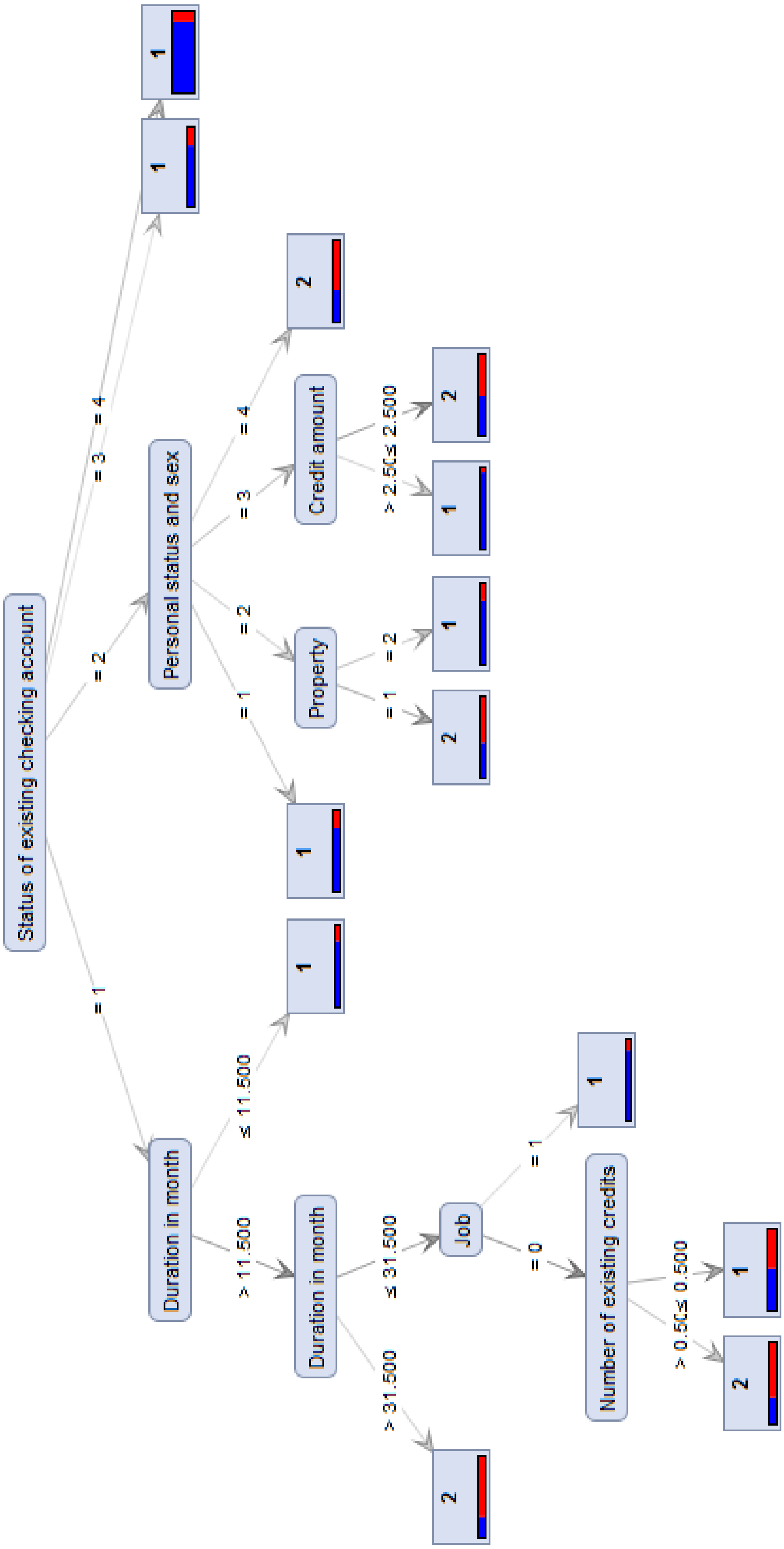

Figure 6. Output of decision tree in training.

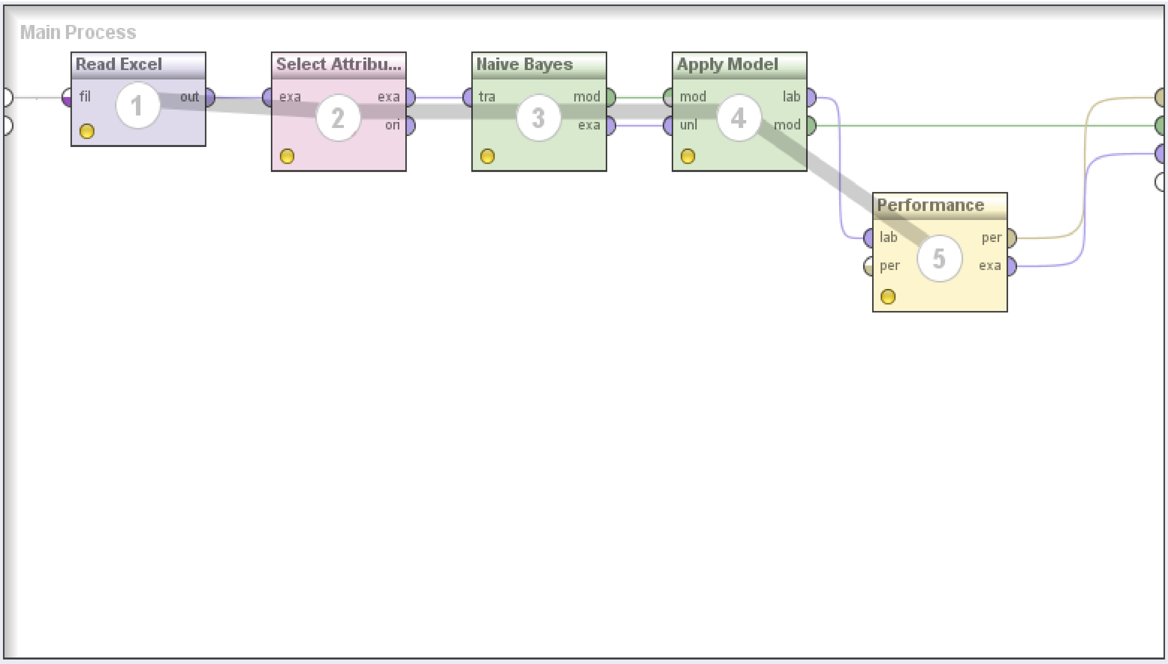

Figure 7. The structure of operators in training in Naïve Bayes.

accuracy: 75.80%

	true 1	true 2	class precision
pred. 1	587	129	81.98%
pred. 2	113	171	60.21%
class recall	83.86%	57.00%	

Figure 8. Model accuracy in training in the Naïve Bayes model.

Table 3. The criterion comparison in training in Naïve Bayes.

Recall	Precision	Class
83.86%	81.98%	1= good
57.00%	60. 21 %	2 = bad
	Accuracy=75.80	
	Classification Error = 24.20	

The evaluation of the classification algorithms is shown in Figure 9, where k represents K-fold cross-validation with a value of K= 10.

As shown in Figure 10 and Table 4, the accuracy of the model in the test section is 73.60%.

Probability analysis as shown in Figure 11 is described as:

- If the status of the existing checking account is equal to 1 with a probability of 45%, it is in class 2.
- If the status of the existing checking account is equal to 2 with a probability of 35%, it is in class 2.
- If the status of the existing checking account is equal to 3 with a probability of 49%, it is in class 1.
- If the credit history is equal to four, with a probability of 34%, it is in class 1.
- If Present employment is equal to 3, with a probability of 57%, it is in class 1.
- If Personal status and sex were equal to 1, with a probability of 31%, it was classified as class 1.
- If the Other installment plans are equal to 1, with a probability of 62%, it is in class 2.
- If the job is equal to zero with a probability of 94%, it is in class 2.
- If the foreign worker is equal to zero, with a probability of 84%, it is in class 1.

ALGORITHM OF SUPPORT VECTOR MACHINE

Support vector machines (SVM) is a classification method applied to a big dataset, as shown in Figure 12. A large dataset contains more node predictions [13].

- *Third model:* SVM algorithm
- *Input data:* All fields except the target variable
- *Purpose:* Good and bad customer account

As shown in Figure 13 and Table 5, the model accuracy for the training was 72.00%.

Table 4. Criterion comparison in test section in Naïve Bayes.

Recall	Precision	Class
83.00%	80.03%	1= good
51.67%	56. 57 %	2 = bad
	Accuracy=73.60	
	Classification Error = 26.40	

Table 5. The criterion comparison in training in SVM.

Recall	Precision	Class
97.71%	72.15%	1= good
12.00%	69. 23 %	2 = bad
	Accuracy=72.00	
	Classification Error = 28.00	

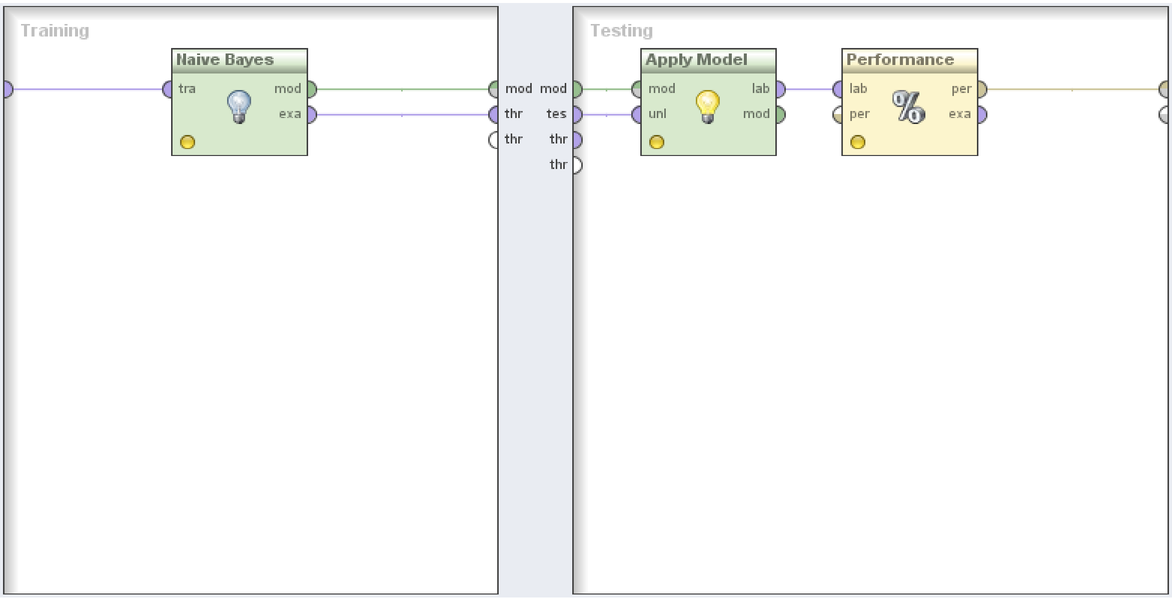

Figure 9. Operators structure in test in Naïve Bayes model.

accuracy: 73.60% +/- 3.32% (mikro: 73.60%)

	true 1	true 2	class precision
pred. 1	581	145	80.03%
pred. 2	119	155	56.57%
class recall	83.00%	51.67%	

Figure 10. The model accuracy in the test section in the Naïve Bayes model.

Evaluation of classification algorithms, as shown below in Figure 14, where k is K-fold cross-validation with a value of K= 10.

As shown in Figure 15 and Table 6, the model accuracy was 71.20%.

Table 6. The criterion comparison in testing in SVM

Recall	Precision	Class
97.00%	71.78%	1= good
11.00%	61. 11 %	2 = bad
	Accuracy=71.20	
	Classification Error = 28.80	

Attribute	Parameter	1	2
Status of existing checking account	value=1	0.199	0.450
Status of existing checking account	value=2	0.234	0.350
Status of existing checking account	value=4	0.497	0.153
Status of existing checking account	value=3	0.070	0.047
Status of existing checking account	value=unknown	0.000	0.000
Duration in month	mean	19.207	24.860
Duration in month	standard deviation	11.080	13.283
Credit history	value=4	0.347	0.167
Credit history	value=2	0.516	0.563
Credit history	value=3	0.086	0.093
Credit history	value=0	0.021	0.083
Credit history	value=1	0.030	0.093
Credit history	value=unknown	0.000	0.000
purpose	value=radio/tv	0.311	0.207
purpose	value=education	0.040	0.073
purpose	value=furniture/equipment	0.176	0.193
purpose	value='new car'	0.207	0.297
purpose	value='used car'	0.123	0.057
purpose	value=business	0.090	0.113
purpose	value='domestic appliance'	0.011	0.013
Attribute	Parameter	1	2
purpose	value=repairs	0.020	0.027
purpose	value=other	0.010	0.017
purpose	value=retraining	0.011	0.003
purpose	value=unknown	0.000	0.000
Credit amount	mean	2.290	1.673
Credit amount	standard deviation	1.651	1.303
Savings account/bonds	value=5	0.270	0.213
Savings account/bonds	value=3	0.336	0.347
Savings account/bonds	value=4	0.193	0.130
Savings account/bonds	value=1	0.056	0.077
Savings account/bonds	value=2	0.146	0.233
Savings account/bonds	value=unknown	0.000	0.000
Present employment since	value=3	0.574	0.487
Present employment since	value=2	0.287	0.363
Present employment since	value=1	0.043	0.067
Present employment since	value=4	0.096	0.083
Present employment since	value=unknown	0.000	0.000
Installment rate	mean	2.843	2.850
Installment rate	standard deviation	1.108	1.095
Personal status and sex	value=1	0.317	0.200

Attribute	Parameter	1	2
Personal status and sex	value=2	0.230	0.237
Personal status and sex	value=4	0.124	0.223
Personal status and sex	value=3	0.329	0.340
Personal status and sex	value=unknown	0.000	0.000
other parties	value=none	0.907	0.907
other parties	value=guarantor	0.060	0.033
other parties	value='co applicant'	0.033	0.060
other parties	value=unknown	0.000	0.000
Present residence since	mean	2.726	2.557
Present residence since	standard deviation	0.659	0.793
Property	value=2	0.344	0.307
Property	value=1	0.619	0.667
Property	value=3	0.031	0.020
Property	value=4	0.006	0.007
Property	value=unknown	0.000	0.000
Age in years	mean	1.156	1.153
Age in years	standard deviation	0.363	0.361
Other installment plans	value=2	0.416	0.377
Other installment plans	value=1	0.584	0.623
Other installment plans	value=unknown	0.000	0.000
Attribute	Parameter	1	2
Housing	value=1	0.953	0.987
Housing	value=2	0.047	0.013
Housing	value=unknown	0.000	0.000
Number of existing credits	mean	0.207	0.297
Number of existing credits	standard deviation	0.406	0.458
Job	value=0	0.877	0.943
Job	value=1	0.123	0.057
Job	value=unknown	0.000	0.000
Number of people being liable	mean	0.907	0.907
Number of people being liable	standard deviation	0.290	0.291
Telephone	value=0	0.967	0.940
Telephone	value=1	0.033	0.060
Telephone	value=unknown	0.000	0.000
foreign worker	value=0	0.844	0.767
foreign worker	value=1	0.156	0.233
foreign worker	value=unknown	0.000	0.000
u	value=1	0.753	0.620
u	value=0	0.247	0.380
u	value=unknown	0.000	0.000
v	value=0	0.979	0.977

Figure 11. Probability output in Naïve Bayes.

NEURAL NETWORK

Neural networks are simple modules of the human neural system. The basis is a neuron organized in some layers. A neural network called a multi-layer perceptron is a simplified model of data processing in the human brain. Neural networks simulate a series of internal connections among neurons, as shown in Figure 16.

- Fourth model: Neural Network
- Input data: All fields except the target variable
- Purpose: Good and bad customer account

As shown in Figure 17 and Table 7, the model accuracy was 96.30%.

The structure of the operators in training in neural net is shown in Figure 18.

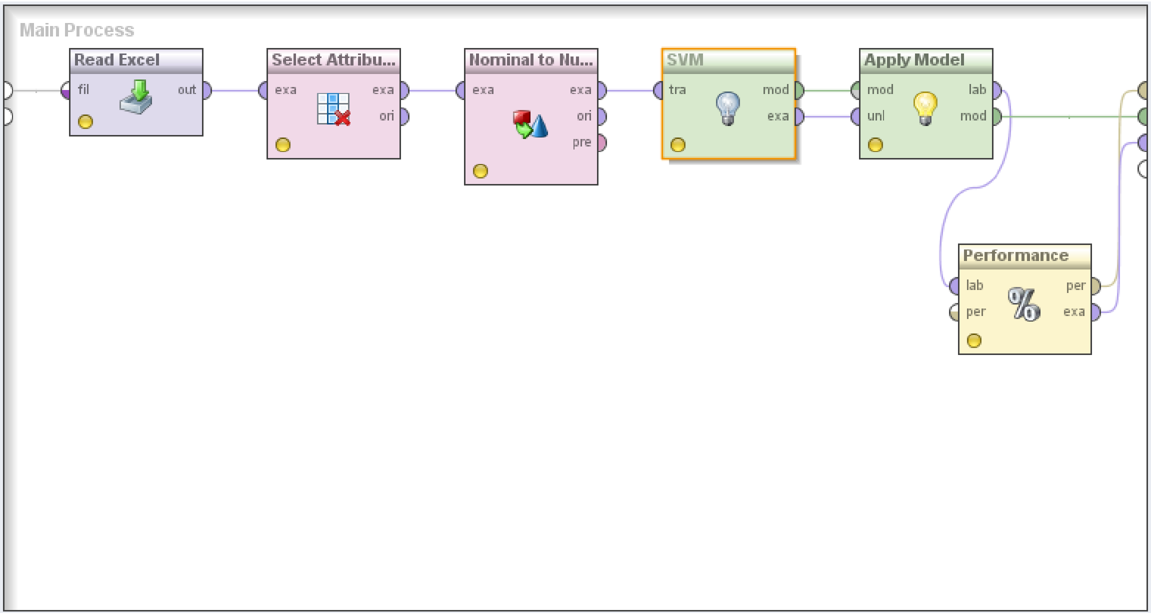

Figure 12. The operators' structure in training in the SVM model.

accuracy: 72.00%

	true 1	true 2	class precision
pred. 1	684	264	72.15%
pred. 2	16	36	69.23%
class recall	97.71%	12.00%	

Figure 13. The model accuracy in training in SVM.

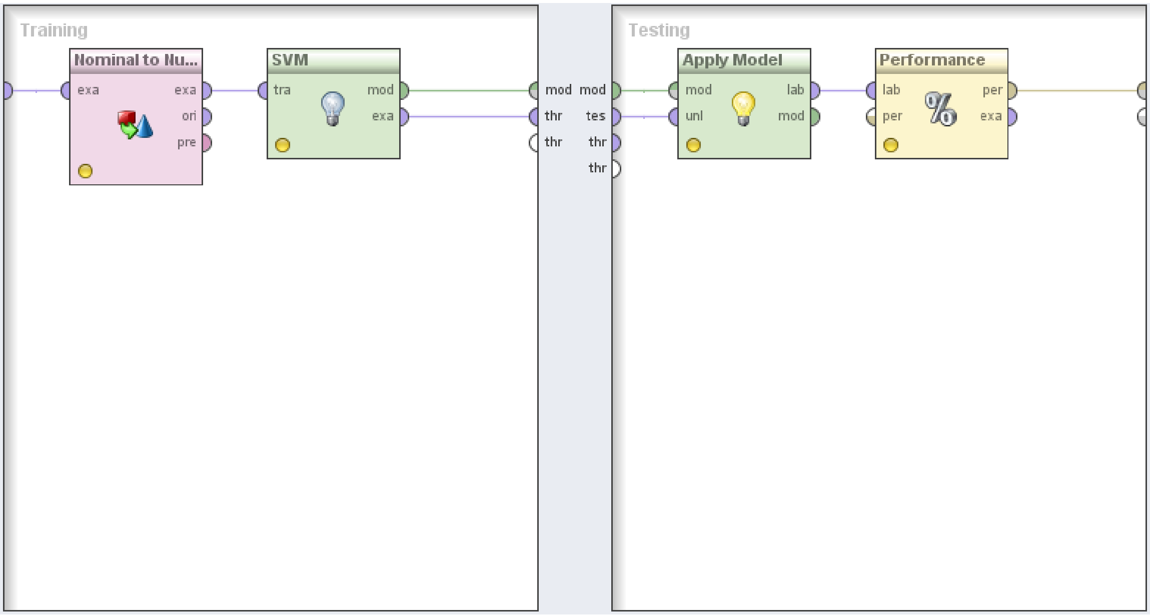

Figure 14. The operators' structure in testing in the SVM model.

Table 7. The criterion comparison in training in a neural net.

Recall	Precision	Class
98.86%	95.98%	1= good
90.33%	97. 13 %	2 = bad
	Accuracy=96.30	
	Classification Error = 3.70	

accuracy: 71.20% +/- 1.08% (mikro: 71.20%)

	true 1	true 2	class precision
pred. 1	679	267	71.78%
pred. 2	21	33	61.11%
class recall	97.00%	11.00%	

Figure 15. The model accuracy in testing in SVM algorithm.

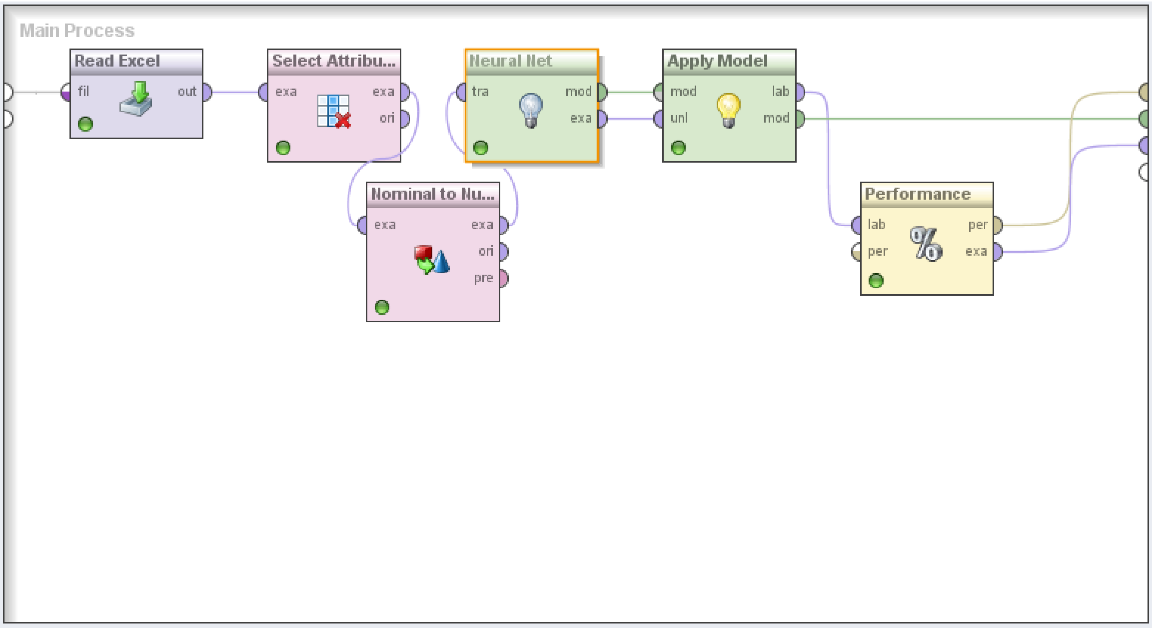

Figure 16. The structure of operators in training in neural net.

accuracy: 96.30%

	true 1	true 2	class precision
pred. 1	692	29	95.98%
pred. 2	8	271	97.13%
class recall	98.86%	90.33%	

Figure 17. The model accuracy in training in neural net,

Table 8. The criterion comparison in training in neural net.

Recall	Precision	Class
86.86%	73.08%	1= good
25.33%	45. 24%	2 = bad
	Accuracy=68.40	
	Classification Error = 31.60	

As shown in Figure 19 and Table 8, the accuracy of the model was 68.40%.

The Majority Vote Hybrid Method

The main idea of this operator is the selection of the majority vote for classification and the mean of the estimated values for the regressions. The vote operator had a microprocessor with two modeling operators. To estimate the tag of a new sample, each model in the vote operator microprocessor states their proposed class tag as a vote. Finally, by counting votes, the majority vote is used as the tag of class, as shown in Figures 20–22. This method can considerably increase the model accuracy, and the model is stable against a set of data with noise or lost values.

The model accuracy in the combined structure of four algorithms is shown in Figure 23.

RESULT

As shown in the models and above sets we have:

The total accuracy of the model in classes 1 and 2 is presented in Table 9.

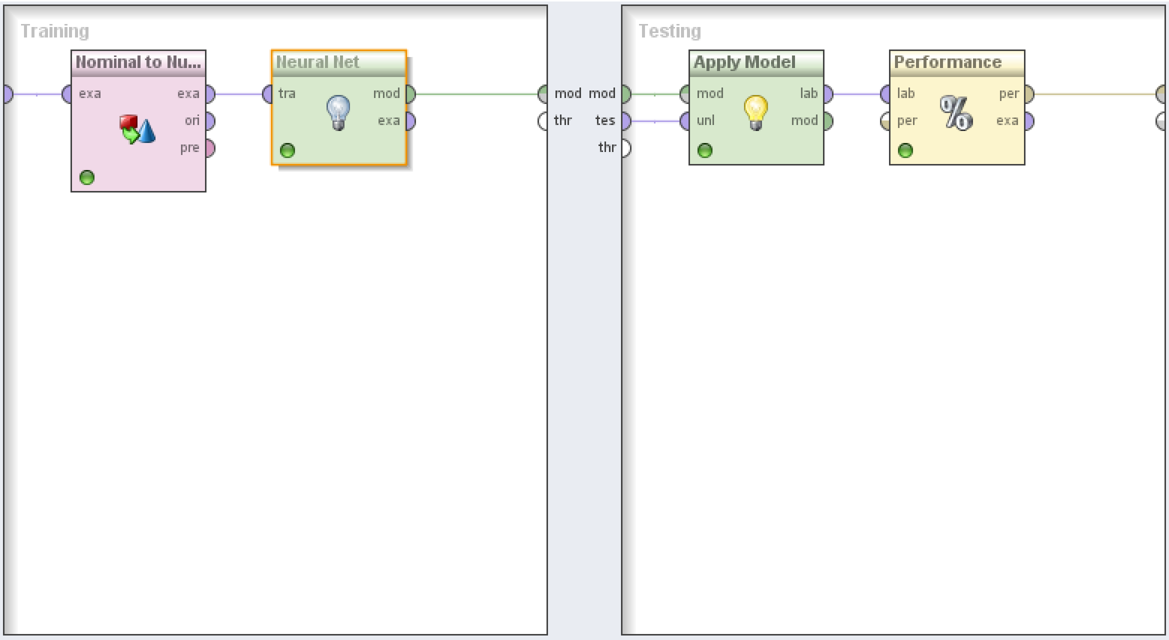

Figure 18. The operators' structure in training in neural net.

accuracy: 68.40% +/- 4.74% (mikro: 68.40%)

	true 1	true 2	class precision
pred. 1	608	224	73.08%
pred. 2	92	76	45.24%
class recall	86.86%	25.33%	

Figure 19. The model accuracy in testing in neural net.

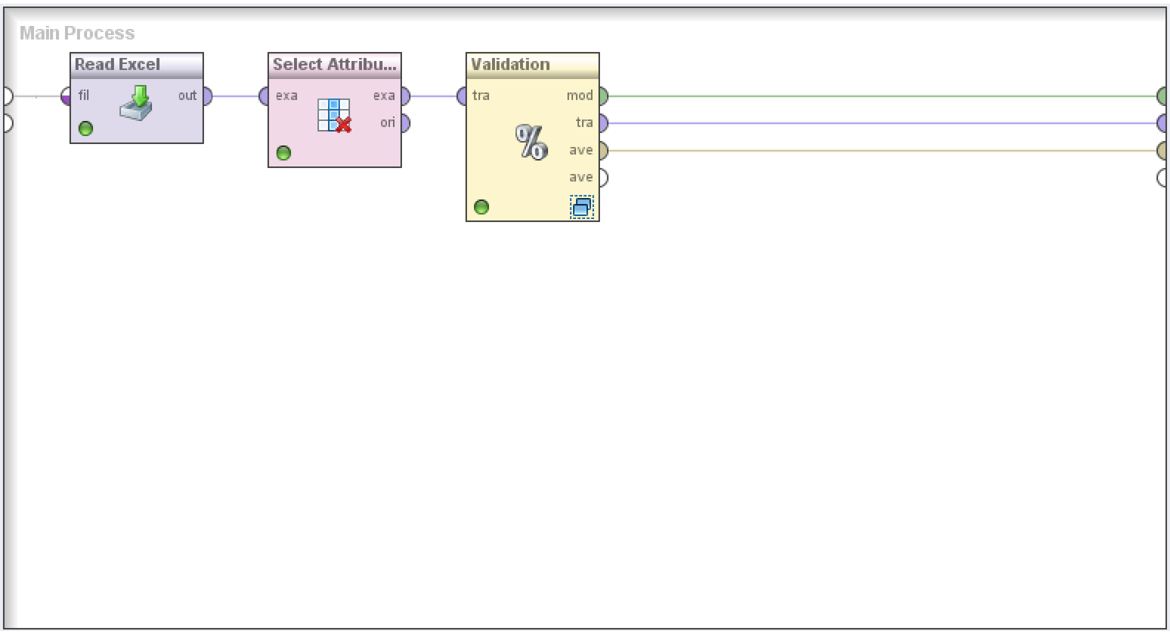

Figure 20. The structure of operators in the main process.

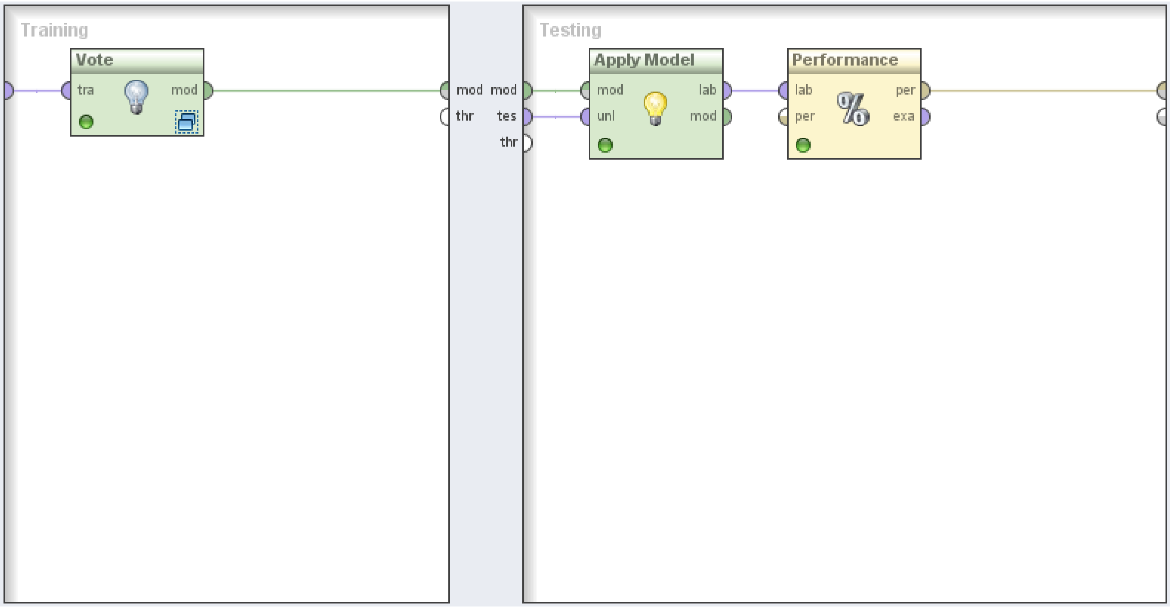

Figure 21. The operators' structure in training and testing mode.

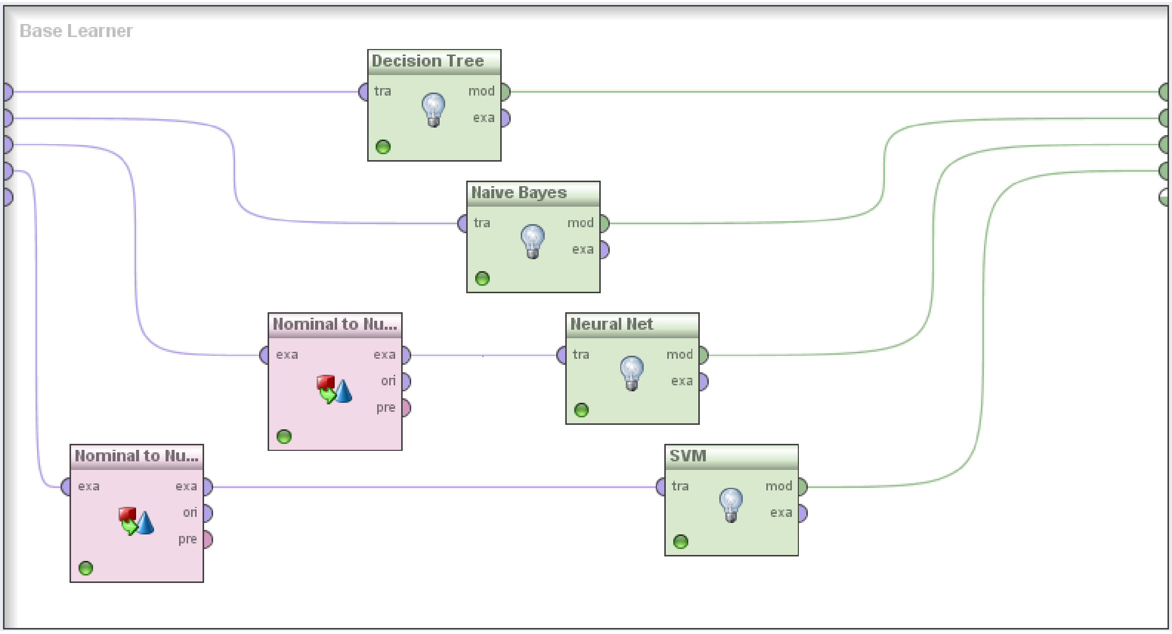

Figure 22. The operators' structure in the base learner.

accuracy: 72.50% +/- 2.94% (mikro: 72.50%)

	true 1	true 2	class precision
pred. 1	646	221	74.51%
pred. 2	54	79	59.40%
class recall	92.29%	26.33%	

Figure 23. The model accuracy in the combined structure of models.

Precision of Good Class Using Precision Criterion

Precision criterion: This is based on the classifier prediction accuracy, and it shows how much trust we can have in the output of the classifier, as shown in Table 10.

The accuracy of each model and its comparison with the combined structure of the four models using the majority vote are shown in Table 11.

Table 9. Total accuracy in two classes 1, 2.

Model	Model accuracy in training	Model accuracy in testing
C4.5	75.30%	71.10%
Naïve Bayes	75.80%	73.60%
SVM	72.00%	71.20%
Neural Net	96.30%	68.40%

Table 10. Accuracy of good class using precision criterion.

Model	Model accuracy in training	Model accuracy in testing	Model
CART	Precision	80.48%	77.81%
Naïve Bayes	Precision	81.98%	80.03%
SVM	Precision	72.15%	71.78%
Neural Net	Precision	95.98%	73.08%

Table 11. The accuracy of each model and its comparison with the four models.

Model	Model accuracy in testing
CART	71.10%
Naïve Bayes	73.60%
SVM	71.20%
Neural Net	68.40%
VOTE – Majority vote	72.50%

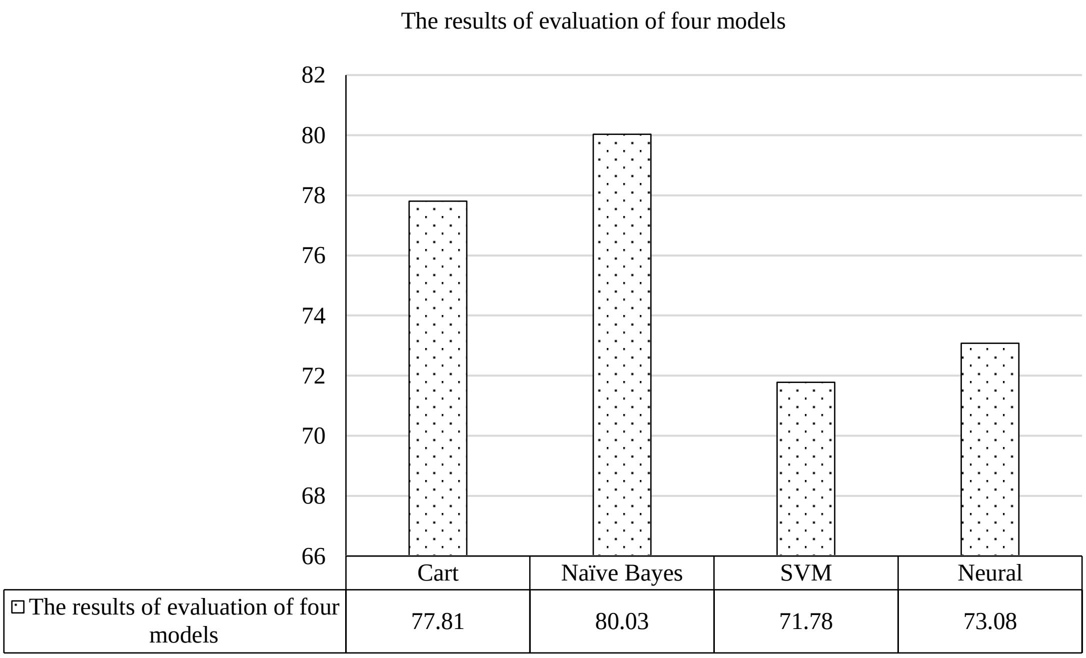

Figure 24. The assessment results of four models.

For effective competition in competitive markets, banks should have better perceptions of customers and the market. The banking industry has imposed significant changes to its activities. Pioneer banks use data mining tools for dividing customers, validating, and confirming them, predicting non-payouts, marketing, and identifying fraud models. This study attempts to use data mining in the banking industry to predict customer lending. Decision-making trees, Naïve Bayes, support vector machines, and neural networks have acceptable accuracy in data mining of collected data regarding bank customers. The results show that Naïve Bayes has better performance compared to other algorithms in the prediction of giving loans to customers, as shown in Figure 24. By selecting good and bad customer accounts as the target variable and 24 predictive variables, it was predicted that the status of the existing checking account, duration in months, personal status and sex, Property, Credit amount ,and number of existing credits were recognized as important variables.

CONCLUSION

The main purpose of this research was to investigate the factors that affect bank customers' credit using a combined approach of data mining to improve decision-making. We know that banks are trying to grant their facilities to customers who, while having low-risk, can have a return commensurate with the interest of the granted facility. This project attempted to use data mining in the banking industry to predict the payment of credit facilities to customers. Decision trees, simple Bayes, support vector machine, and neural network, despite their simplicity, provided acceptable accuracy results in the data mining of the collected data related to bank customers, and the results showed that simple Bayes performed better than other algorithms in predicting credit facilities for customers.

Author Contributions

Conceptualization, MB, and MT; methodology, MB, and MT; software, MT, and MB; validation, MB, and MT; formal analysis, MB, and MT; investigation, MT, and MB; resources, MB, and MT; data curation, MB, and MT; writing—original draft preparation, MB, and MT; writing—review and editing, MT, and MB; visualization, MT, and MB; supervision, MB, and MT; project administration, MB, and MT. All authors have read and agreed to the published version of the manuscript.

Conflict of Interest

The authors declare no conflict of interest.

REFERENCES

1. Ghadamzan Jalali A, Habibi Machiani H, Taghipour M, Moshtaghi S. Explain the relationship between intellectual capital, organizational learning and employee performance of Parsian Bank Branches in Gilan Province. Educ Adm Res Q. 2020;10:127-142.
2. Najafi S, Saati S, Tavana M. Data envelopment analysis in service quality evaluation: An empirical study. J Ind Eng Int. 2015;11:319-330. doi:10.1007/s40092-014-0090-6.
3. Blose JE, Tankersley WB, Flynn LR. Managing service quality using data envelopment analysis. Qual Manag J. 2005;12:7-22. doi:10.1080/10686967.2005.11919246.
4. Shirouyehzad H, Lotfi FH, Shahin A, Aryanezhad MB, Dabestani R. Performance evaluation of hotels by data envelopment analysis based on customers' perception and gap analysis. Int J Serv Oper Manag. 2012;12:447-467. doi:10.1504/IJSOM.2012.047953.
5. Baghipour Sarami F, Bozorgiamiri A, Mououdi MA, Taghipour M. Modeling of nurses' shift work schedules according to ergonomics: A case study in Imam Sajjad (As) Hospital of Ramsar. J Ergonomics. 2016;4:1-12. doi:10.21859/joe-04011.
6. Taghipour M, Habibi MH, Amin M. The impact of working capital management on the performance of firms listed in Tehran Stock Exchange (TSE). J Multidiscip Eng Sci Technol. 2018;7:24-32.
7. Taghipour M, Soofi Mowloodi E, Mahboobi M, Abdi J. Application of cloud computing in system management in order to control the process. Manag Int Technol Sci Publ (ITS). 2020;3:34-55.
8. Dastyar B, Kazemnejad H, Sereshgi AA, Jabalameli MA. Using data mining techniques to develop knowledge management in organizations: A review. J Eng Proj Prod Manag. 2017;7:80-89. doi:10.32738/JEPPM.201707.0004.
9. Akbarnezhadbaei K, Mohamadi M, Kouloubandi A, Taghipour M. Determining a model for evaluating the knowledge management system in order to improve industries with the focus on educational technology and applying data mining concepts. J Mod Thoughts Educ. 2023.

10. Akbarnezhadbaei K, Mohammadi M, Kouloubandi A, Taghipour M. Modeling the application of knowledge management system in order to improve the technology governance in the automotive industry of Iran using the data mining environment. Educ Adm Res. 2023;14.
11. Bel L, Allard D, Laurent JM, Cheddadi R, Bar-Hen A. CART algorithm for spatial data: Application to environmental and ecological data. Comput Stat Data Anal. 2009;53:3082-3093. doi:10.1016/j.csda.2008.09.012.
12. Rutkowski L, Jaworski M, Pietruczuk L, Duda P. The CART decision tree for mining data streams. Inf Sci. 2014;266:1-15. doi:10.1016/j.ins.2013.12.060.
13. Salcedo-Sanz S, Rojo-Álvarez JL, Martínez-Ramón M, Camps-Valls G. Support vector machines in engineering: An overview. WIREs Data Min Knowl Discov. 2014;4:234-267. doi:10.1002/widm.1125.

