Jurnal Teknik Informatika (JUTIF)
Vol. 3, No. 6, Desember 2022, hlm. 1739-1746
DOI: https://doi.org/10.20884/1.jutif.2022.3.6.615
p-ISSN: 2723-3863
e-ISSN: 2723-3871

IMPROVED SUPPORT VECTOR MACHINE PERFORMANCE USING PARTICLE SWARM OPTIMIZATION IN CREDIT RISK CLASSIFICATION

Aditiarno Manik[1], Erna Budhiarti Nababan[*2], Tulus[3]

[1,2]Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Indonesia
[3]Faculty of Mathematics and Natural Sciences, Universitas Sumatera Utara, Indonesia
Email: [1]tiarnoaditia@gmail.com, [2]ernabrn@usu.ac.id, [3]tulus@usu.ac.id

(Naskah masuk: 19 September 2022, Revisi : 5 Oktober 2022, diterbitkan: 26 Desember 2022)

Abstract

In Classification using Support Vector Machine (SVM), each kernel has parameters that affect the classification accuracy results. This study examines the improvement of SVM performance by selecting parameters using Particle Swarm Optimization (PSO) on credit risk classification, the results of which are compared with SVM with random parameter selection. The classification performance is evaluated by applying the SVM classification to the Credit German benchmark credit data set and the private credit data set which is a credit data set issued from a local bank in North Sumatra. Although it requires a longer execution time to achieve optimal accuracy values, the SVM+PSO combination is quite effective and more systematic than trial and error techniques in finding SVM parameter values, so as to produce better accuracy. In general, the test results show that the RBF kernel is able to produce higher accuracy and f1-scores than linear and polynomial kernels. SVM classification with optimization using PSO can produce better accuracy than classification using SVM without optimization, namely the determination of parameters randomly. Credit data classification accuracy increased to 92.31%.

Keywords: *Classification, Credit Data, Particle Swarm Optimization, Support Vector Machine, SVM Parameter.*

1. INTRODUCTION

Credit has become an important stimulus in encouraging the economy to grow and is also the main activity of every bank everywhere [1]. The provision of credit provides opportunities for the community to improve or increase their consumption and provide opportunities for companies to make investments that were previously difficult to do using their own funds. In allocating or providing loans to customers or potential borrowers, commercial banks and financial institutions must consider the credit risk that will be given, whether the customer will repay the credit or not (non-performing loans). If the credit default is too large, it can cause cash flow to the creditor company to be hampered and all the positive benefits expected from the provision of credit will not occur[2]. Therefore, credit giving companies need a good analysis in their credit planning so that the possibility of bad loans can be known beforehand. The credit risk level classification method has played an important role in contemporary practice of credit risk management activities and contributes to the main credit approval decisions so as to efficiently and accurately measure the level of credit risk proposed by prospective borrowers.

Classification itself has been widely used in various fields, including experimental classification of diabetic retinopathy using retinal images[3]**.** The important objective of the credit classification method is to separate the group of borrowers who are unable to repay and the group of borrowers who are able to repay loans. Today, classification performance has become increasingly important for credit scoring, because even a small percentage of the increase can already provide a large number of advantages for financial institutions[4].

SVM which is included in supervised learning is able to model a classification by utilizing training data and used in predicting test data which is new data from observations and then the model is further validated. SVM can provide optimal accuracy results and the error rate tends to be small because it has the ability to generalize and produce good performance compared to other methods[5],[6]**.** The identification of SVM as an important tool to choose among researchers for credit model development has been reviewed in various studies[7],[8]**.**

The biggest challenge in establishing the SVM classification model is in terms of determining the hyper parameter values[9]. While it is known that the determination of the right parameter values can improve the accuracy of the performance on the SVM. Classification model[10]**.** Therefore, in order to obtain parameter values that can provide the best performance results on the SVM classification model, it is necessary to search for SVM parameters which are expected to optimize classification performance. The search for SVM parameters in question is an activity to determine the hyperparameters of the SVM classification model that can provide optimal performance results.

This study implemented the Particle Swarm Optimization (PSO) method to find the optimal SVM parameter value with several advantages. PSO as an evolutionary computational technique, is able to obtain a globally optimal solution in the search space by considering the interactions between individuals in a swarm of particles. Each particle provides information on its best position to the other particles while adjusting the position and speed of change in position based on the information obtained regarding the best position. If a comparison is made with similar algorithms, such as the Genetic Algorithm (GA), then PSO is a simpler model that does not require more procedure steps, such as selection, mutation or crossover steps of the GA method. Not only that, the PSO technique is also successful in finding optimal parameters in other machine learning methods that are known from several previous studies[11], as in the classification of breast cancer datasets[12] and in predicting agricultural water consumption[13], where SVM-PSO is judged to be simpler to adjust compared to other methods such as GA with minimum adjustments to its parameters. On credit assessment alone, PSO is very suitable to be combined with SVM[14] and PSO is also suitable to be combined with the Artificial Neural Network method[15].

The application of PSO in previous research has given results that illustrate the increase in classification accuracy in determining SVM parameters that optimize accuracy. Seeing the ability of the PSO method, this research uses the PSO method to be applied in the search for SVM parameters so that an increase in classification performance will be obtained.

This study will review the performance of SVM by selecting parameters using PSO to classify credit risk which will be compared with SVM using default parameters. An issue with the credit scoring model that should be noted is the difficulty of obtaining real-world credit data, as customer credit data is confidential in most financial institutions and researchers cannot gain access to this data. To evaluate the classification performance, this study will apply the SVM classification using the German credit data set from UCI and the private data set which is a credit data set issued by a local bank in North Sumatra.

2. REVIEW OF RELATED STUDIES

Many studies related to credit risk classification have been carried out previously and various classification methods or algorithms have been studied and applied. Research [16] studied public data in the form of credit approval datasets obtained from the UCI repository and also used private data, which were credit approval datasets sourced from local banks. Each attribute in their research was weighted using an information acquisition algorithm. Each attribute has an unequal weight based on the calculations performed. The K-NN algorithm used for classification results in an increased accuracy of 3.26% for local datasets and 7.53% for UCI datasets. Their research shows that the output generated from the classification model is not affected by any attribute after the attribute selection in the two datasets.

Some research examines the advantages of computational intelligence as well as soft computing methods, namely a new hybrid approach to improve credit risk management. In their research, they apply a modeling method under conditions of uncertainty, where the parameters of the neural network, including weights and errors, are considered in the form of fuzzy numbers. In this method, the underlying system is first modeled using ANN and then, using fuzzy inference, the optimal decision will be determined with the highest degree of excellence.

The empirical results of using the proposed method show the high efficiency and accuracy of this method in analyzing credit rating issues. Research [17] studied the classification between good and bad customers with twenty variables from the German Credit Dataset. Two non-statistical techniques were used in their research namely ANN and SVM. Because there are various advantages and disadvantages in implementing the model, it can be said that the model with the highest prediction success, according to the data set used is the SVM method. Paper [18] studied a profit-driven approach to classifier construction and simultaneous variable selection based on linear SVM. Its main purpose is to incorporate business-related information such as variable acquisition costs, Type I and II error costs, and profits generated by properly classified examples, into the modeling process. The proposed framework is studied in credit scoring issues for Chilean banks, and yields superior performance with respect to business-related objectives.

Research [19] studied that class imbalance occurs when instances in a class are much higher than in other classes. This major machine learning issue can affect predicted accuracy. Their research shows that SVM is a reliable and appropriate method in dealing with class imbalance problems but weak in biased data distribution. Paper [20] studied the resampling-based learning paradigm using SVM based on Deep Believe Network (DBN) to overcome the problem of data imbalance in credit classification. The experimental results show that classification performance increases effectively when the DBN-based ensemble strategy is integrated with resampling techniques, especially on imbalanced-data problems. The applied DBN-based SVM resampling ensemble learning paradigm can be used as a promising tool for credit risk classification with unbalanced data.

Another research [21] studied that the use of SVM for credit risk classification is interesting to do because SVM is characterized by versatility, resilience, and computational simplicity. At the same time, even in the case of a limited number of samples, SVM can obtain better classification results. Other Paper [22] studied the SVM classification method, concluded that the selection of kernel functions and their parameters played an important role in the results. Radius Basis Function (RBF) is a commonly used kernel. For RBF-SVM, two parameters, *c* and, are used to control SVM performance. The penalty parameter c determines the trade-off between minimizing installation errors and maximizing margins between classes. The parameter determines the RBF bandwidth. Therefore, in this study, an effective technique is implemented to find pairs of parameters c and that will optimize the results of SVM classification.

3. OBJECTIVES OF THE STUDY

The problem in the application of the SVM method is the challenge in finding the parameter values that can provide optimal results. Determining the hyperparameter value of SVM is the biggest problem in setting up the SVM model, even though the accuracy of the SVM classification model can be improved by determining the right parameter values. So we need a method to be applied in finding the right parameters which can provide optimal SVM outputs with the best accuracy results in credit risk classification. The expected goal of this research is to improve the performance of SVM in credit risk classification.

4. METHODS

The purpose of this study is to develop the best predictive model in classifying credit risk (current or non-current/loss) using the SVM classification method with the best parameter search based on PSO. In general, the stages of the research methodology to be carried out are described as follows.

4.1. Input Data

In this first stage, the selection of data to be used for testing is carried out. This study uses two credit data sets, namely: (1) German Credit Data set which is a benchmark credit data set published by the UCI dataset, and (2) Private Credit Data set which is a credit data set issued by a local bank in North Sumatera. In the credit risk analysis model, one of the problems that needs to be considered is the unavailability of real credit data from financial institutions providing credit, because customer credit data is confidential in most financial institutions. Currently, researchers are obtaining private data which is a credit data set issued by a local bank in North Sumatra. The data consists of 390 instances with 16 attributes.

The data set obtained from this local bank is a data set of capital loans (credit) given to prospective borrowers for certain purposes, including: working/business capital, investment or consumptive. The data recorded by the bank is partly data related to prospective borrowers and data related to the amount of credit applied for. Of the 390 credit borrowers whose data has been recorded, there are 162 borrowers who fall into the bad credit group, which is denoted as class 1, and there are 228 borrowers who fall into the current credit group, which is denoted by class 0. The data characteristics of each attribute are exist in the dataset can be seen in Table.1. Henceforth, this set of credit data from local banks is called Private Credit Data. The characteristics of the Private Credit Data set are described as follows:

Tabel 1. Private Credit Data Set Attribute Information

Number	Attribute	Information
Attribute 1	Gender (data type: categorical)	1: Man 2: Woman
Attribute 2	Age in years (data type: numeric)	-
Attribute 3	Marital Status (data type: categorical)	1: Married 2: Single 3: divorced
Attribute 4	Number of Dependents (data type: integer)	
Attribute 5	Job Type (data type: categorical)	1000: Agriculture 3000: Industry 6000: Trading 8000: services 9990: others
Attribute 6	Length of Service (data type: categorical)	1: < 1 year 2: between 1 – 4 years 3: between 4 – 7 years 4: > 7 years
Attribute 7	Residential Ownership Status (data type: categorical)	1: One's own 2: Rent
Attribute 8	Credit Term, in months (data type: numeric)	-
Attribute 9	Initial Ceiling (data type: numeric)	-
Attribute 10	Outstanding (data type: numeric)	-
Attribute 11	Rate (data type: numeric)	-
Attribute 12	Interest Installments (data type: numeric)	-
Attribute 13	Principal Installment (data type: numeric)	-
Attribute 14	Purpose of credit application (data type: categorical)	10: Working capital 20: Investation 39: Consumptive
Attribute 15	Nature of Credit (data type: categorical)	0: Combined 1: Credit by Agreement – Co-Financing 2: Credit by Agreement – Other Banks/Institutions 3: Credit by Agreement – Bank Debtor 4: Credit by Agreement – Others 6: Short-term loan 9: Credit Without Agreement
Attribute 16	Nature of Credit (data type: categorical)	0 : Current Credit

	1: Bad Credit (failed)

- Dataset Name: Kredit Privat Data Set
- Attribute Characteristics: *Categorial, Numerik, Integer*
- Number of Instances: 390
- Number of Attributes: 16 (8 numerik, 7 categorial, 1 class label)

Sector: Finance

4.2. Preprocessing Data

Preprocessing data aimed at being able to provide optimal impact on the classification results, consists of 5 stages, namely: (a) redundant handling, (b) missing value handling, (c) outlier handling, (d) categorical data transformation, and (e) data normalization. The five preprocessing stages will be applied to the private credit data set, while for the German credit benchmark data set, only steps (d) and (e) are required.

a. handling redundant or duplicate data. Handling redundant begins with the detection of redundant data, to find the presence of repetitive data (dual data). After the redundant data is found, then handling the redundant data that has been detected, by deleting the redundant data and leaving 1 data from the redundant data.

b. Handling missing values.

The technique for handling missing values or missing data is using the mean substitution method. The mean substitution method is the replacement of missing values by using the average value (mean), which is an estimation technique that is very often used where the missing value is replaced by the average value of the variable.

c. Handling data outliers.

Outlier handling begins with the detection of outlier data, to find any abnormal data (outliers). The technique for detecting outliers is by comparing the data with the standard deviation of each attribute. After the outliers are found, then the outlier data that has been detected is handled by using the median substitution method. In the matlab program, outlier detection and handling uses the command:

$$y = \text{hampel (sort (DATA(:, x))}$$

Using the command above applies a Hampel filter to the input vector, DATA, to detect and remove data outliers in column x. For each DATA sample, the function calculates the median of the window consisting of the sample and its six surrounding samples, three per side. It also estimates each sample's standard deviation about the window median using the median absolute deviation. If the sample differs from the median by more than three standard deviations, it is replaced by the median. If DATA is a matrix with many columns (attributes), then it's easy to treat each x column as an independent channel.

d. Categorical Data Transformation

For data sets that have categorical attribute data, it is necessary to convert them to numeric form first. The changes can be done by replacing the data with certain numbers as long as they are consistent. In the German data set, there are 20 features in the data, of which there are 13 features which are categorical data. In the second data set, namely the Private credit data set which is also used in the test, in total there are 15 attributes and 1 class label in the data, where from 15 attributes there are 7 attributes which are categorical data. The categorical data in each attribute is transformed into numeric form with the help of the Matlab application. Replacement of categorical data into numeric, namely by transforming categorical data into a number with a value that has the same weight, where data consisting of 2 categories is transformed into values 1 and 2; while the data consisting of 3 categories, are transformed into values 1, 2 and 3. And so on for the data with the next number of categories.

e. Normalization

The overall credit data set has been in numeric form, all of its attributes have different value ranges. The difference in the range of data (the upper and lower limits of the data) will affect the weight of the attribute influence on the classification results. Where, attributes with a larger data range will be judged to have more influence on the classification results than attributes with a small data range. For this reason, the next step is to normalize the data for all the attributes used, so that the data will have the same range, which is between 0 and 1. Then the data set that has been normalized is ready to be used in the next process stage, namely the classification process using the SVM method and classification using SVM+PSO method.

4.3. SVM Classification

The data sets that have been preprocessed are then classified using the SVM classification method with three kernels, namely the linear kernel, the RBF kernel and the polynomial kernel. Determination of test data and training data is done by validation using 10-fold. In the model, it is recommended to use 10 fold cross validation as the best number of folds for validity testing [23]. Determination of SVM parameters for each kernel (linear, RBF and polynomial) is done randomly. Even though it is done randomly, the parameters used must still be limited, which is in the range between the smallest parameter value and the largest parameter value. The search interval for the SVM parameter value used is the parameter values c and that promise to exist in the interval $c = [2^{-5}, 2^{7}]$ and $\gamma = [2^{1}, 2^{6}]$ [22] while the parameter d is sought at interval d= $[2,7]$ [24]..

4.4. SVM + PSO Classification

Furthermore, credit data classification is carried out using the SVM+PSO classification method, namely the SVM classification with the kernel parameters determined using the PSO method. Initialization of PSO using the best parameters obtained from the SVM classification. The SVM+PSO classification process will stop if the stopping criteria are met, that is, if in 10 iterations there is no increase in accuracy. The process of searching for SVM parameters using PSO shown in Figure.1. The following describes the steps for implementing PSO in the SVM parameter search.

4.5. SVM + PSO Initialization

In Figure.1, it can be seen the SVM+PSO classification process starting from the initialization of PSO particles, including the determination of several PSO parameters and the test parameters that were determined at the beginning. In this study, the PSO parameters and test parameters used are as follows:

K-fold validation : 10-fold.
Number of particles : 30 particles
Inertia weight (ω_k) : 2
Cognitive weight (c_1): 2
Social weight (c_2) : 2

PSO parameters, namely inertia weight (ω_k) is used as a parameter that controls the effect of the previous particle velocity. If the inertia value is too large, the velocity will continue to increase so that the particles will diverge. The particle distance to the optimum value will continue to increase with each iteration. Another parameter in the PSO algorithm is the cognitive weight parameter (c_1) and social weight parameters (c_2), namely the acceleration constant which has an influence on the speed of convergence. The value of ω_k, c_1, and c_2 , c_2 used in this research test is based on the parameters commonly used in the application of PSO.

4.6. Update PSO Particle Position and Evaluation of Fitness Value

Next, the initial position of the PSO particle which represents the initial parameter value of SVM will be updated by applying formula.1 and formula.2. The initial position of the particle is symbolized by the notation $x_{i,j}^k$. The notation *i* represents the particle *i*, *k* represents the iteration *k*, where *j* represents parameter *j*. The latest position of the particle (the result of the position update) is symbolized by $x_{i,j}^{k+1}$ in iteration s=1, where the stopping criterion used is if in 10 consecutive iterations, there is no increase in accuracy of more than 0.001%.

$$v_{ij}^{k+1} = \omega_k * v_{ij}^k + c_1 * rand * \left(pbest_{ij}^k - x_{ij}^k\right) + c_2 * rand * \left(gbest_{ij}^k - x_{ij}^k\right) \quad (1)$$

$$x_{ij}^{k+1} = x_{ij}^k + v_{ij}^{k+1} \quad (2)$$

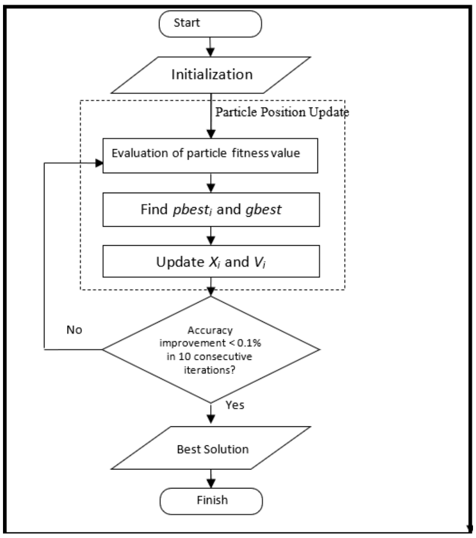

Figure 1. SVM Parameter Search Flowchart With PSO

The accuracy of the SVM and SVM+PSO classification methods was validated using a 10-fold cross validation technique. Accuracy is calculated using the values obtained from the confusion matrix table. The average accuracy of 10 pieces of accuracy to the final accuracy. The classification accuracy of SVM and SVM+PSO is then compared to see how far the impact of PSO is in improving SVM performance.

5. RESULT AND DISCUSSION

From the classification results of generating 30 random combinations of parameters on credit data in each kernel, the highest accuracy is seen in each credit data set, which is presented in Table 2.

Tabel 2. Highest Accuracy Generated by SVM Classification

Kernel	**Credit Privat Data Set (%)**	**Credit German Data Set (%)**
Linear	90.26	62.50
RBF	89.74	70.00
Polinomial	88.97	55.80

In Table 2 it can be seen that the test results on the Private Credit data set, the highest SVM classification accuracy is 90.26% from the linear kernel, while the test results on the German Credit data set, the highest SVM classification accuracy is 70.00% from the RBF kernel.

Comparison The experimental results between the highest accuracy of SVM classification and the highest accuracy of SVM+PSO classification in both data sets can be seen in Table 3.

Testing the SVM + PSO Linear classification on Private data sets, the value of c which is initialized in the PSO process is the value of c which produces the highest accuracy in the SVM test, namely the value of c = 12.1595, the iteration process stops at the 13th iteration with an accuracy of 90.77% . The results of this final accuracy, when compared with the initial accuracy, there is an increase in accuracy of 1.79%, with the parameter that produces the highest accuracy, c = 0.0336. Testing on private data sets, the value of c which is initialized is the value of c = 0.1595, the iteration process stops at the 13th iteration with an accuracy of 63%. The results of this final accuracy, when compared with the initial accuracy, there is an increase in accuracy of 0.5%, with the parameter that produces the highest accuracy, namely c = 0.0321.

Tabel 3. Comparison of Classification Accuracy on Private Credit Data Sets

Set Data	Kernel	SVM	SVM + PSO	Iteration	Time (s)
Credit Privat Dataset	Linear	90.26%	90.77%	13	159.60
	RB	89.74%	92.31%	35	236.73
	Polinomial	88.97%	89.49%	11	454.24
Credit German Data Set	Linear	62.5%	63%	13	398.2
	RB	70%	74.25%	15	262.14
	Polinomial	55.8%	63.16%	11	419.87

Testing the SVM+PSO RBF classification on a private data set, the parameter values c and which are initialized in the PSO process on the RBF kernel are the values that produce the highest accuracy in the SVM test. The process stopped at the 35th iteration and obtained an accuracy result of 92.31% with a combination of parameters that produced the highest accuracy in the RBF kernel, namely the parameters c = 8.9540 and = 3.5291. Testing the SVM+PSO kernel RBF classification on the Credit German data set, the iteration process stops at the 15th iteration with an accuracy of 74.25%. The results of this final accuracy, when compared with the initial accuracy, there is an increase in accuracy of 4.25% with the parameters that produce the highest accuracy in SVM RBF, namely the parameters c = 2.3287 and = 58.0708.

Testing the SVM+PSO Polynomial classification on the Private Credit data set, the process stops at the 11th iteration after there is no increase in accuracy of 10 iterations in a row, and an accuracy result of 89.49% is obtained. The combination of parameters that produces the highest accuracy in the Polynomial kernel is at parameters c = 5.5674 and d = 2. Testing the SVM + PSO Polynomial kernel classification on the Credit German data set, the iteration process stops at the 11th iteration with an accuracy result of 63.16%. The results of this final accuracy, when compared with the initial accuracy, there is an increase in accuracy of 7.36% with the parameters that produce the highest accuracy in SVM Polynomial, namely the parameters c = 9.5472 and d = 2.

PSO can improve accuracy which is better than the accuracy generated by SVM classification with random parameters. This proves that PSO is effective for finding parameter values that can produce high accuracy. Table 4 summarizes the results of increasing the accuracy of each data on each kernel.

Tabel 4. Improved SVM classification accuracy with optimization method

SVM Kernel	Credit Privat Data Set (%)	Credit German Data Set (%)
Linear	0.51%	0.5%
RBF	2.57%	4.25%
Polinomial	4.62%	7.36%

The increase in the accuracy of linear SVM+PSO classification in the Private data set is 0.51% and in the German credit data set there is an increase in accuracy of 0.5%. The increase in the classification accuracy of SVM+PSO RBF in the private data set is 2.57% and in the Credit German data set there is an increase in accuracy of 4.25%. The increase in the accuracy of SVM+PSO Polynomial classification in the Private Credit data set is 4.62% and in the German Credit data set there is an increase in accuracy of 7.36%.

If we compare the results of the SVM classification whose parameters are determined randomly with the results of the SVM classification whose parameters are determined by parameter search using the PSO method, the results of the SVM classification using PSO optimization are better than without PSO (random parameters). Comparison of accuracy results without optimization method using PSO optimization method is presented in table-5.

Tabel 5. Comparison of classification accuracy using optimization method with no optimization method

Set Data	Kernel	SVM	SVM + PSO
Credit Privat Dataset	Linear	90.26%	90.77%
	RB	89.74%	92.31%
	Polinomial	88.97%	89.49%
Credit German Data Set	Linear	62.5%	63%
	RB	70%	74.25%
	Polinomial	55.8%	63.16%

In testing using the Private Credit data set, the highest classification accuracy was obtained when using the RBF kernel with an accuracy of 92.31% with the parameter values c and respectively 8.9540 and 3.5291. In testing using the Credit German data set, the highest classification accuracy was obtained when using the RBF kernel with an accuracy of 74.25% with parameter values c and respectively 2.3287 and 58.0708.

If the f1-score value is calculated from each kernel using the best parameters that have been selected, then the f1-score results can be seen in the following table-6.

Tabel 6. F1-Score PSO-based SVM classification

SVM Kernel	Credit Privat Data Set (%)	Credit German Data Set (%)
Linear	89.10	76.92
RBF	90.19	82.35
Polinomial	82.66	78.65

The F1 score for linear SVM was obtained from the Linear SVM classification using the parameter value c = 0.0336, and resulted in an f1-score of 89.10%. For the f1-score SVM kernel RBF obtained from the SVM kernel RBF classification using the parameter values c = 8.9540 and = 3.5291, and produces an f1-score of 90.19%. While the f1-score SVM kernel polynomial is obtained from the SVM kernel polynomial classification using the parameter values c = 5.5674 and d = 5.34 and produces an f1-score of 82.66%. In general, in this study, the RBFi kernel was able to produce higher accuracy and better f1-score values than linear and polynomial kernels. However, it is not certain that the RBF kernel is the best kernel because the selection of the kernel depends on the data used. The execution time of each classification procession with optimization and the execution time without optimization is presented in Table-7.

Tabel 7. SVM classification Execution Time with PSO optimization

Set Data	Kernel	SVM + PSO Time (s)	SVM + PSO Number of Iteration	SVM Time (s)
Credit Privat Dataset	Linear	159.60	13	4.1423
	RB	236.73	35	4.4217
	Polinomial	454.24	11	24.3865
Credit German Data Set	Linear	398.2	13	4.4693
	RB	262.14	15	2.5334
	Polinomial	419.87	11	23.5967

The calculation of execution time for each iteration process has different execution times, i depending on the number of instances and parameter values for each kernel. i In addition to being influenced by the number of instances, i execution time is also influenced by the magnitude of the parameter values. Also time the execution.

6. CONCLUSION

Each kernel in the SVM method requires parameters that can provide maximum results in data classification. A technique is needed to find parameter values in each kernel, where in this study, SVM was combined with PSO optimization techniques in finding parameter values that could provide optimal results. Although achieving optimal accuracy values requires a longer execution time than the trial and error SVM parameter search technique, combining SVM and PSO to find parameter values that provide optimal accuracy has resulted in a more systematic technique. SVM classification with optimization using PSO can produce better accuracy than classification using SVM without optimization, namely the determination of parameters randomly. Credit data classification accuracy increased to 92.31%.

REFERENCES

[1] M.A. Mukid, T. Widiharih, A. Rusgiono, and A. Prahutama, “Credit Scoring Analysis Using Weighted K Nearest Neighbor,” *Journal of Physics: Conference Series* 1025, 2018.

[2] H. Leidiyana, “Penerapan Algoritma K-Nearest Neighbor untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor,” *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, Vol. 1, Issue: 1, pp: 65-76, 2013.

[3] M.F. Pasha, M.D. Sikder, A. Rana, M.S. Lidya, R. Purba, & R. Budiarto, “Experimenting Diabetic Retinopathy Classification Using Retinal Images,” *Data Science: Journal of Computing and Applied Informatics*, vol. 5, no. 1, pp. 28-38, 2021.

[4] J. Abellán and J.G. Castellano, “A Comparative Study on Base Classifiers in Ensemble Methods For Credit Scoring,” *Expert Systems with Applications*, Vol. 73, pp: 1-10, 2017.

[5] B. Baesens, T.V. Gestel, S. Viaene, and M. Stepanova, “Benchmarking State-Of-The-Art Classification Algorithms for Credit Scoring,” *Journal of the Operational Research Society*, Vol. 54, Issue: 6, pp: 627-635, 2003.

[6] Sudianto, P. Wahyuningtias, H. W. Utami, U. A. Raihan, H. N. Hanifah, and Y. N. Adanson, “COMPARISON OF RANDOM FOREST AND SUPPORT VECTOR MACHINE METHODS ON TWITTER SENTIMENT ANALYSIS (CASE STUDY: INTERNET SELEBGRAM RACHEL VENNYA ESCAPE FROM QUARANTINE)”, *J. Tek. Inform. (JUTIF)*, vol. 3, no. 1, pp. 141-145, Feb. 2022.

[7] T. Moro, G. Tinsley, A. Bianco, G. Marcolin, Q.F Pacelli, G. Battaglia, A. Palma, P. Gentil, M. Neri, and A. Paoli, “Effects of eight weeks of time-restricted feeding (16/8) on basal metabolism, maximal strength, body composition, inflammation, and cardiovascular risk factors in resistance-trained males,” *Journal of Translational Medicine*, Vol 14, Issue: 1, pp: 1-10, 2016.

[8] H.A. Alakaa, L.O. Oyedele, H.A. Owolabi, V. Kumar, O. Saheed, S.O. Ajayi, O.O. Akinadef, and M. Bilal, “Systematic Review of Bankruptcy Prediction Models: Towards A Framework For Tool Selection,” *Expert Systems with Applications*, Vol. 94, pp: 164-

184, 2018.

[9] S. Yenaeng, S. Saelee, and W. Samai, "Automatic Medical Case Study Essay Scoring by Support Vector Machine and Genetic Algorithms," *International Journal of Information and Education Technology*, Vol. 4, Issue 2, pp: 132-137, 2014.

[10] C. Huang and C. Wang, "A GA-Based Feature Selection and Parameters Optimizationfor Support Vector Machines," *Expert Systems with Applications*, Vol. 31, Issue 2, pp: 231-240, 2006.

[11] A.T. Abolhassani and M. Yaghoobi, "Stock Price Forecasting Using PSO SVM," *Advanced Computer Theory and Engineering*, pp. 352-356, 2010.

[12] R. Habibi, "SVM Performance Optimization Using PSO for Breast Cancer Classification," *Budapest International Research in Exact Science*, Vol. 3, No. 1, 2021.

[13] Y. Tang, and J. Zhou, "The Performance of PSO-SVM in Inflation Forecasting," *International Conference on Service Systems and Service Management (ICSSSM) IEEE*, 2015.

[14] P. Danenas & G. Garsva, "Credit Risk Evaluation Using SVM-Based Classifier," *Lecture Notes in Business Information Processing*, pp: 7–12, 2010.

[15] P. Li and H. Xiao, "An Improved Quantum-Behaved Particle Swarm Optimization Algorithm," *Applied Intelligence*, Vol. 40, Issue 3, pp: 479–496, 2013.

[16] Ivandari, T. Chasanah, S. Binabar, and M. Karomi, "Data Attribute Selection with Information Gain to Improve Credit Approval Classification Performance using K-Nearest Neighbor Algorithm," *International Journal of Islamic Business and Economics (IJIBEC)*, Vol. 13, 2017.

[17] L. Gafarova, "Usage of Artificial Neural Network and Support Vector Machine model for classification of Credit Scores," Tesis, Khazar University, Azerbaijan, 2017.

[18] S. Maldonado, C. Bravo, J. López, and J. Péreza, "Integrated Framework For Profit-Based Feature Selection And SVM Classification In Credit Scoring," *Decision Support Systems*, Vol. 104, pp: 113-121, 2017.

[19] Hartono, O.S. Sitompul, Tulus, and E.B. Nababan, "Biased Support Vector Machine and Weighted-Smote In Handling Class Imbalance Problem," *International Journal of Advances in Intelligent Informatics*, Vol. 4, no. 1, pp: 21–27, 2018.

[20] L. Yu, R. Zhou, L. Tang, and R. Chen, "A DBN-Based Resampling SVM Ensemble Learning Paradigm for Credit Classification With Imbalanced Data," *Applied Soft Computing*, Vol. 69, pp: 192-202, 2018.

[21] J. Han, W. Jiang, C. Dai, and H. Ma, "The Design of Diabetic Retinopathy Classifier Based on Parameter Optimization SVM," *International Conference on Intelligent Informatics and Biomedical Sciences*, Vol. 3, pp: 52–58, 2018.

[22] Y. Xie, Y. L. Murphey, and D.S. Kochhar, "SVM Parameter Optimization Using Swarm Intelligence for Learning from Big Data," *International Conference on Computational Collective Intelligence (ICCCI): Lecture Notes in Computer Science*, Vol. 11055, pp: 469-478, 2018.

[23] E. Prasetyo, "Data Mining – Mengolah Data Menjadi Informasi Menggunakan Matlab," Edisi 1, Penerbit Andi, Yogyakarta, 2014.

[24] H. Al-Azies, D. Trishnanti, and E. Mustikawati, "Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI)," *IPTEK Journal of Proceedings Series*, No. 6, ISSN (2354-6026) 53, The 1st International Conference on Global Development - ICODEV, November 19th, 2019, Rectorate Building, ITS Campus, Sukolilo, Surabaya, Indonesia, 2019.