Approval of Sharia Cooperative Customer Financing Using PSO-Based SVM Classification Algorithm

Nurajijah¹, Fachri Amsury², Irwansyah Saputra³, Friyadie⁴, Daning Nur Sulistyowati⁵, Bakhtiar Rifai⁶
Nusa Mandiri College of Information Management and Computer, Jatiwaringin Raya Street No 2, Cipinang Melayu Village, Makasar District, East Jakarta 13620
E-mail: nurajijah.nja@nusamandiri.ac.id

Abstract. Credit financing approval is an important task for financial institutions. The accuracy of decision making in accepting or rejecting credit applications must be precise and accurate. Data mining techniques can help credit approval with a smaller risk of error. Previous research has been carried out to classify customer loan history data with the highest accuracy obtained by the Support Vector Machine (SVM) algorithm. This study aims to improve the accuracy of previous research in classifying loan history data based on smooth and stuck credit labels to predict credible prospective customers. The algorithm used in managing the data is Support Vector Machine with Particle Swarm Optimization (PSO). The dataset used was 869 records from January 2015 to September 2018, labeled 291 records and 578 records smoothly. The results of the study proved that the Support Vector Machine algorithm based on Particle Swarm Optimization produced the best performance with an accuracy of 90.91% compared to SVM without PSO of 89.86%.

1. Introduction

Financing or credit approval has a risk that is the inability of customers to pay their credit obligations when due. For companies and financial institutions, managing credit risk has been considered the most important task [1] [2]. Banks hold huge volumes of customer behavior-related data from which they are unable to arrive at judgment if an applicant can be defaulter or not [3]. The banking industry applies data mining techniques that help compete in the market and provide appropriate products for customers with smaller risks [4]. Data mining takes knowledge using large amounts of data stored in a database, data warehouse, or information in a repository [5].

In a previous study [6] a classification of sharia cooperative customer loan history data was performed using the Naïve Bayes algorithm, Decision Tree, and Support Vector Machine (SVM) with the highest accuracy obtained using SVM. Support Vector Machine has been used [7] to classify sentiments to existing opinions so that it can be predicted in advance which ones have more positive and predictable sentiments as elected governors. The application of a genetic algorithm is proposed [8] as an optimal parameter value search algorithm so that it can improve the best classification accuracy in Support Vector Machine (SVM). Particle Swarm Optimization has been shown to increase the accuracy value in the classification of the Neural Network algorithm [9].
and improve the accuracy in the classification using the SVM algorithm [10]. In [11], it is explained that SVM has a higher resilience and generalization ability and a more stable classification accuracy compared to other algorithms.

Based on previous research, this study aims to improve the accuracy of the classification of sharia cooperative customer loan history data using the Support Vector Machine algorithm based on Particle Swarm Optimization to predict credible sharia cooperative prospective customers and obtain better accuracy results. The results of the classification can be used in helping to make good decisions for cooperative management in refusing or approving financing requests from customers.

![Figure 1. Research Framework](image)

2. Method

The flowchart in this study is illustrated in Figure 1. The research method used is the Cross-Standard Industry for Data Mining (CRISP-DM), this method as standardization of data mining approaches compiled by three people namely Daimler Chrysler (Daimler-Benz), SPSS (SPSS ISL), NC [12]. The stages in CRISP-DM include business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

This research uses a financing dataset obtained from one of the sharia cooperatives totaling 869 records during January 2015 to September 2018 labeled 291 records for stuck label and 578 records smooth label. The financing dataset is divided into 819 records (543 labeled smooth and 276 labeled stuck) as input classification using Rapidminer and 50 records (35 labeled smoothly and labeled 15 stuck) as testing. The financing dataset collected was then selected into 14 criteria (Table 1). The dataset is classified using the Support Vector Machine Algorithm. The modeling of the financing dataset classification into a smooth and traffic class using Rapidminer is shown in Figure 2. Evaluate model testing to get accurate model information. Evaluation and validation use several types of kernels and compare the validation values of 2-10 k-fold cross validations.

2.1. Experimental Results and Model Testing

2.1.1. Support Vector Machine Determining the value of research accuracy to do rapidminer testing using various types of kernel types and comparing validation values 2-10 by entering the value C = 0.0 and Epsilon = 0.0. Table 2 are the results of an
Table 1. Attribute data

| Attribute               | Type     | Information (M=Million) |
|-------------------------|----------|-------------------------|
| Gender                  | Binominal| Male, Female             |
| Age                     | Polinominal| <25, 25-50, >50         |
| Status                  | Polinominal| Married, Single, Widowed|
| Status of residence     | Binominal| Rent, Freehold           |
| The number of dependents| Polinominal| 0, 1, 2-3, >3           |
| Education               | Polinominal| Elementary, Middle School, High School, Diploma, Bachelor |
| Profession              | Polinominal| Traders, Employees       |
| Income                  | Polinominal| <3M, 3M-5M, >5M         |
| Flafond                 | Polinominal| <3M, 3M-5M, 5M-10M, >10M-50M, >50M |
| Loan type               | Polinominal| Musyarakah, Murabaha    |
| Period of time          | Polinominal| Short, Medium, Long     |
| Payment method          | Polinominal| Daily, Weekly, Monthly  |
| Guarantee               | Polinominal| Electronics, Motor Certificates, Car Certificates, Savings, Land Certificates |
| Label                   | Binominal| Smooth, stuck            |

experiment that has been carried out using several types of kernels. In table 2 the highest accuracy test results are the DOT kernel type found in K-Fold 9 with a value of 90.22%.

Table 2. Support Vector Machine Comparison of accuracy of kernel types DOT, Radial, Polynominal, Neural, and Multiquadric

| K-Fold | DOT  | Radial | Polynominal | Neural | Multiquadric |
|--------|------|--------|-------------|--------|--------------|
| 10     | 89.65%| 79.17% | 89.30%      | 53.51% | 66.51%       |
| 9      | 90.22%| 79.05% | 88.50%      | 55.18% | 66.52%       |
| 8      | 90.21%| 79.06% | 88.60%      | 54.43% | 66.51%       |
| 7      | 90.34%| 78.25% | 88.48%      | 57.31% | 66.51%       |
| 6      | 89.65%| 79.40% | 87.92%      | 59.14% | 66.51%       |
| 5      | 89.76%| 79.06% | 88.37%      | 56.27% | 66.51%       |
| 4      | 89.29%| 78.94% | 88.72%      | 56.17% | 66.51%       |
| 3      | 88.61%| 78.48% | 87.80%      | 59.49% | 66.51%       |
| 2      | 88.50%| 78.83% | 86.42%      | 59.49% | 66.51%       |

2.1.2. Support Vector Machine Based on Particle Swarm Optimization. Determining the value of research accuracy to do rapidminer trials using various types of kernel types and compare the validation values 2-10 by entering the value $C = 0.0$ and Population Size = 5. Table 3 are the results of an experiment that has been carried out using several types of kernels. In table 3 the highest accuracy test results are the DOT kernel type found in K-Fold 9 with a value of 90.22%.

3. Result and Discussion

Testing the model to predict sharia cooperative customer financing by using a financing dataset of 819 records is used as input for the classification of the Support Vector
Table 3. Support Vector Machine Based on Particle Swarm Optimization
Comparison of accuracy of kernel types DOT, Radial, Polynomial, Neural, and Multiquadric

| K-Fold | DOT     | Radial   | Polynomial | Neural    | Multiquadric |
|--------|---------|----------|------------|-----------|--------------|
| 10     | 91.03%  | 85.96%   | 90.80%     | 71.58%    | 66.51%       |
| 9.     | 91.48%  | 89.07%   | 90.68%     | 71.59%    | 66.52%       |
| 8      | 91.37%  | 88.96%   | 90.79%     | 75.02%    | 66.51%       |
| 7      | 91.37%  | 88.03%   | 90.91%     | 71.00%    | 66.51%       |
| 6      | 91.25%  | 89.41%   | 91.02%     | 80.32%    | 66.51%       |
| 5      | 90.79%  | 86.42%   | 90.36%     | 73.19%    | 66.51%       |
| 4      | 91.03%  | 89.19%   | 90.57%     | 76.75%    | 66.51%       |
| 3      | 90.79%  | 90.33%   | 90.45%     | 73.77%    | 66.51%       |
| 2      | 90.56%  | 84.70%   | 88.72%     | 78.36%    | 66.51%       |

Figure 2. The process of modeling a dataset

Machine algorithm with PSO using Rapidminer (Figure 2). PSO-based SVM and SVM algorithm classification to determine accuracy, precision, recall, and AUC values. Model validation uses 9 fold cross-validation.

Figure 3. AUC of SVM and SVM+PSO

The Support Vector Machine algorithm is very well implemented in this study with the results of the smooth class precision of 88.70%, the precision of the stuck class of 94.40%, the recall class of the smooth at 97.75%, the recall class of stuck at 75.26% (Table 5), AUC of 0.937 (Figure 3) and accuracy of 90.22% (Table 5). PSO-based SVM algorithm in the classification of cooperative customer financing produces
Table 4. Weight value of Support Vector Machine Based on Particle Swarm Optimization

| Attribute                | Weight | Attribute                | Weight |
|--------------------------|--------|--------------------------|--------|
| Payment=Monthly          | 0.0473 | Payment=Daily            | 0.6939 |
| Payment=Daily            | 0.6939 | Education=Diploma        | 0.0002 |
| Payment=Weekly           | 1      | Education=Bachelor       | 0.9922 |
| Flafond=10000001-3M      | 0      | Education=Elementary     | 0.0512 |
| Flafond=10000001-3M      | 1      | Education=High School    | 0      |
| Flafond=5000001-10M      | 0.0286 | Education=Middle School  | 1      |
| Flafond=<3M              | 0      | Income=3M-5M             | 0.9733 |
| Flafond=>5M              | 0.8176 | Income=<3M               | 0.4912 |
| Guarantee=Electronics    | 0.6029 | Income=>5M               | 0.1447 |
| Guarantee=Car Certificates| 0.0671 | Status=widowed           | 0.6411 |
| Guarantee=Motor Certificates | 0.6901 | Status=Single            | 1      |
| Guarantee=Land Certificates | 1     | Status=Married           | 0.9874 |
| Guarantee=Savings        | 0.5445 | Residence=Freehold       | 0.8786 |
| Gender=Male              | 0      | Residence=Rent           | 0.7685 |
| Gender=Female            | 0      | Age=25-50                | 0.8094 |
| Loan Type=Murabaha       | 0.2268 | Age=<25                  | 0      |
| Loan Type=musyarakah     | 0      | Age=>50                  | 0      |
| Dependents=0             | 0.8024 | Period of Time=Medium    | 1      |
| Dependents=1             | 0      | Payment Method=Weekly    | 1      |
| Dependents=2-3           | 0.0199 | Guarantee=Land Certificates | 1     |
| Dependents=>3            | 1      | Guarantee=Land Certificates | 1     |
| Profession=Employees     | 0.6297 |                         |        |

Table 5. Confusion matrix SVM algorithm

| True | True | Class |
|------|------|-------|
| Smooth | Stuck | Precision |
| Pred.Smooth | 565 | 72 | 88.70% |
| Pred.Stuck   | 13  | 219 | 94.40% |
| Class recall | 97.75% | 75.26% |

Table 6. Confusion matrix algorithm SVM + PSO

| True | True | Class |
|------|------|-------|
| Smooth | Stuck | Precision |
| Pred.Smooth | 561 | 57 | 90.78% |
| Pred.Stuck   | 17  | 234 | 93.23% |
| Class recall | 97.83% | 80.41% |

Table 7. Testing the SVM + PSO algorithm

| SVM | SVM + PSO |
|-----|-----------|
| Accuracy | 90.22% | 91.48% |
| AUC     | 0.937   | 0.942  |

precision class smooth of 90.78%, precision class stuck of 93.23%, recall class smooth of 97.78%, recall class stuck of 80.41% (Table 6), AUC of 0.942 (Figure 3).
The application of the Support Vector Machine algorithm produces an accuracy of 90.22%. The application of the PSO-based SVM algorithm in this study results in higher accuracy compared to without using PSO with an accuracy value of 91.48%.

Based on the results of the accuracy testing process of SVM and SVM based on PSO, it can be stated that the best performance is generated by the PSO based SVM algorithm. From this, the accuracy value for the SVM classification algorithm model is 90.22% and the accuracy value for the PSO based SVM classification algorithm is 91.48% with an accuracy difference of 1.26% can be seen in Table 7.


definition

PSO-based SVM for classification of sharia cooperative customer financing data into a traffic jam and smoothly produces the best performance with an accuracy of 91.48% and AUC 0.942 compared to SVM without PSO of 90.22%. The results of the classification can be used in helping to make the right decision for cooperative management in refusing or approving customer financing requests.

References

[1] Siswanto, Abdussomad, Gata W, Wardhani N K, Gata G and Prasetyo B H 2019 The Feasibility of Credit Using C4.5 Algorithm Based on Particle Swarm Optimization Prediction 2019 6th International Conference on Electrical Engineering Computer Science and Informatics (ECESI) Bandung Indonesia pp 416-421
[2] Liu N, Xia E J, and Yang L 2019 Research and Application Of Pso-Bp Neural Networks in Credit Risk Assessment in 2010 International Symposium on Computational Intelligence and Design Hangzhou China pp 105-106
[3] Sudhamathy G 2016 Credit Risk Analysis and Prediction Modelling of Bank Loans Using R International Journal of Engineering and Technology 8 1954-1966
[4] Sudhakar M and Reddy C V K 2016 Two Step Credit Risk Assessment Model for Retail Bank Loan Applications Using Decision Tree Data Mining Technique International Journal of Advanced Research in Computer Engineering and Technology vol 5 no 3 pp 705-718
[5] Jiawei H, Kamber M, Han J and Pei J 2012 Data Mining: Concepts and Techniques
[6] Nurajijah N and Riana D 2019 Algoritma Naive Bayes, Decision Tree dan SVM untuk Prediksi Penetapan Pembayaran Nasabah Koperasi Syariah Jurnal Teknologi dan Sistem Komputer vol 7 no 2 10.14710/jtsiskom.7.2.2019.77-82
[7] Kristiyanti D, Unam A H, Wahyudi M, Amin R and Marlinda L 2018 Comparison of SVM Naive Bayes Algorithm for Sentiment Analysis Toward West Java Governor Candidate Period 2018-2023 Based on Public Opinion on Twitter 1-6 10.1109/CITSM.2018.8674352.
[8] Manurung J, Mawengkang H, and Zamzami E 2017 Optimizing Support Vector Machine Parameters with Genetic Algorithm for Credit Risk Assessment Journal of Physics: Conference Series vol 930 no 1
[9] Badrul M, Friejadie, Akmaludin, Ningtyas D A, Sulistyowati D N and Nurajijah 2018 Comparison Of optimization Of Algorithm Particle Swarm optimization And Genetic Algorithm With Neural Network Algorithm For Legislative Election Result 1-7 10.1109/CITSM.2018.8674257.
[10] Kurniawati I and Parde H F 2018 Hybrid Method of Information Gain and Particle Swarm Optimization for Selection of Features of SVM-Based Sentiment Analysis 1-5 10.1109/ICTSI.2018.8695957.
[11] Yu H, Huang X, Hu X, and Cai H 2010 A Comparative Study on Data Mining Algorithms for Individual Credit Risk Evaluation in 2010 International Conference on Management of E-Commerce and E-Government Chengdu China Oct 2010 pp 35-38
[12] North M 2012 Data Mining for the Masses Global Text