Determination of Granting Appropriateness Credit at "Daruzzakah Rensing" Cooperative Using the Support Vector Machine (SVM) Algorithm

Yahya1(*) , Nurhidayati1

1Universitas Hamzanwadi Lombok, Indonesia

Abstract. Credit is the main product of savings and loan cooperatives to increase profitability. The greater the credit issued, the greater the benefits obtained by cooperatives. Each cooperative will package credit products in such a way as to attract the attention of every customer. However, cooperatives can find problems in the process of lending, such as the "Daruzzakah Rensing" Cooperative located in "Desa Rensing, Kecamatan Sakra Barat, Lombok Timur-NTB-Indonesia". The main products of the Cooperative "Daruzzakah Rensing" are savings and loans. In distributing credit, the cooperative always decides based on statistical data. This data is sometimes not useful if the supporting methods used to predict and classify the data are not appropriate. Therefore, this research requires a method that can classify and predict problematic and non-problematic customers. To answer this question, using the SVM (Support Vector Machine) algorithm to find out the level of accuracy in analyzing creditworthiness proposed by prospective debtors. The SVM algorithm is used to predict, classify, evaluate, and analyze credit. From the results of data processing carried out using the SVM algorithm (Support Vector Machine), it can be categorized as an excellent method, with an accuracy of 90.42% and AUC at 0.957. Accuracy of 90.42% means the SVM algorithm can provide decisions about feasible or not feasible in granting credit to customers who apply for loans.

Keywords: Credit, Cooperative, SVM (Support Vector Machine), AUC.

1. Introduction

The development of science and technology has produced many changes in all fields, especially in the economic field. The increasing necessities of life make the costs that must be born by each person more and more. Increasingly fierce free-market competition leads to positive synergies where there is involvement between economic actors in efforts to develop the economy. Investor competition has a very important role in influencing economic development.

Cooperatives as an economic movement that grows out of the community are non-governmental organizations, strength and participation of the community in determining the goals, objectives, and implementation [1]–[3].

The "Daruzzakah Rensing" Cooperative is a savings and loan cooperative whose business activities collect and distribute funds to its members at low-interest rates. This cooperative is also called cooperative credit where management is carried out independently and democratically, and its members join voluntarily. There is also a mention that savings and loan cooperatives are non-bank financial institutions that have business activities and accept deposits from members and provide money loans to members with low-interest.
Credit has several roles: to improve money efficiency, improve money circulation, improve the use and circulation of goods, become one of economic stability, increase income distribution and become a tool to improve international relations.

Problems in distributing credit cannot be completely avoided, so every financial institution must keep trying to prevent this from happening. Although credit activities aim to optimize revenue, they must control and minimize the risk of bad credit.

SVM has unique advantages and has been widely used in the field of cooperative, especially in dealing with classification problems[4]. The SVM method transforms the dataset to a high-dimensional and kernel-induced feature space and subsequently determines the separating hyperplane with the maximum distance to the closest points of the training set[5]. SVM model was constructed to distinguish all-α class from other three classes. All-α class are taken as one class alone, and the other three classes as a whole are taken as the other class which is denoted as others[6].

The support vector machine (SVM) is a popular classifier widely used in various machine learning applications. The original formulation of SVM was introduced as a linearly binary classifier. By introducing the kernel trick, linear SVM was developed into kernel SVM to process complicated linearly non-separable data[7].

Investigating the factors that influence the current pattern of whether or not credit payments are part of the credit analysis, where the process takes a long time and identifying data about problematic debtors is still very difficult, so determining creditworthiness is also difficult. Therefore, the application of the Support Vector Machine (SVM) algorithm in the "Daruzzakah Rensing" Cooperative to support the prediction of customers who qualify for a loan, reducing the risk in distributing to customers.

2. Method

2.1 Data Collection

To answer questions with objectives in this research, data was taken from the "Daruzzakah Rensing", Desa Sakra, Lombok Timur, Nusa Tenggara Barat. The data used were 254, 8 indicators used attributes in the data processing of this research.

2.2 Data Processing

A dataset is a database in memory. The dataset has all the characteristics, features and functions of an ordinary database. A dataset can have many tables, and tables can have relationships with one another (relationship). Tables in a dataset can have foreign keys and referential integrity.

A dataset is an object that represents data and its relationships in memory. The structure is similar to the data in the database. The dataset contains collections of table data. In this section, we conduct experiments on real graphs of various types and scales to validate the accuracy and efficiency of our proposed methods. First, we briefly introduce the datasets[8].

There are two kinds of dataset types:

2.2.1. Private Dataset

The dataset can be used to understand private for management behavior, interest in supplying raw material for energy production, and perceptions of bioenergy, etc. The dataset should be of interest to researchers and policy makers assessing the potential supply for energy and the motivations of any other resources and so private dataset is a dataset that can be taken from an organization that we make a place or object of research. As for examples such as agencies, hospitals, factories, service companies, etc[9].

2.2.2. Public Dataset

Public dataset is a dataset that can be taken from a public repository that has been agreed by the researchers. Examples: UCI Repository and ACM KDD, etc.

In this research, the data used private data. This private data is taken from Cooperative "Daruzzakah Rensing ", 254 data consisting of 8 indicators: Account Number, Realization, JKW, Loan Amount, Principal Remaining, Basil Remaining, Arrears, and Collectibility.

2.3 Data analysis
Functional data analysis is the analysis of data in the form of functions. Ideas from functional data analysis are used to represent the spectra with linear functions comprised of piecewise quadratic characteristics with a limited number of coefficients. Data collected from the "Daruzzakah Rensing" were analyzed quantitatively. In addition to descriptive statistics, data is processed using the Support Vector Machine Algorithm which is part of RapidMiner. In the process carried out using the Support Vector Machine Algorithm, attributes and labels are determined, to get results.

2.4. Experiment

To conduct experiments, conducted several stages to get maximum results, including:

a. Business understanding stage

Under the 2018 Cooperative of “Daruzzakah Rensing” Credit report, several credit problems greatly affect the stability of cooperatives. For this reason, the SVM (Support Vector Machine) algorithm model was developed to determine the creditworthiness to make the analysis more thorough. The credit data used is taken from the Cooperative “Daruzzakah Rensing” data in 2018. In this data there are 254 customers which 88 people are considered current, 3 people are substandard, and 163 people are declared bad. In the data, several customer data are used as attributes, namely customer account number, realization, period, loan amount, principal balance, basil balance, arrears, and collectibility.

b. Data understanding stage

In analysis of data with sequential structure, it is often necessary to aggregate or summarize data at a higher level of granularity than its lowest level of natural coarseness. At this stage, as much as 254 data and consisting of 8 attributes will be made several selections to produce the required data, the stages are:

- Data cleaning to clean up empty values or empty tuples.
- Data integration functions to unite different storage (archive) into one data. In this case for a year the data is separated monthly, then with this stage, monthly data will be put together in one year.
- Data reduction, only a few attributes are used, and unnecessary attributes will be deleted. The tuples in the dataset may duplicate or occur the same tuples, so to predict the number of tuples, the same tuples will be made in one tuple to represent the tuples as shown in the table below.

| Account Number | Realization | JKW Amount | Loan Remaining | Principal Remaining | Results | Arrears | Collectibility |
|----------------|-------------|------------|----------------|----------------------|---------|---------|----------------|
| 12.010.000.002 | 02 November | 10         | 500000         | 100000               | 30.000  | 100000  | Not feasible   |
| 12.010.000.003 | 06 November | 102        | 500000         | 215000               | 43.000  | 214999  | Not feasible   |
| 12.010.000.004 | 12 December | 2009       | 10             | 500000               | 100000  | 100000  | Not feasible   |
| 12.010.000.007 | 26 April   | 2010       | 119            | 1200000              | 588000  | 98.000  | 587999 Not feasible |
| 12.010.000.008 | 21 Mei 2010 | 14         | 1000000        | 720000               | 8.000   | 71999   | Not feasible   |
| 12.010.000.010 | 15 Juni 2010 | 72         | 300000         | 193000               | 46.000  | 193000  | Not feasible   |

c. Data preparation stage

In the modeling stage, this stage can also be called the learning stage because at this stage the training data is classified by the model, then produces several rules. The model used at this stage uses the SVM (Support Vector Machine) algorithm model. As explained earlier, several stages must be passed in producing predictions in the form of a kernel model (SVM).
3. Results and Discussion

3.1 The results of using K-Fold Validation 4.

This test is done by dividing 4 parts of 254 customer credit data to be tested. The 4 sections consist of 3 parts training data and 1 part testing data. Confusion matrix model will form a matrix consisting of true positive or positive tuples and true negative or negative tuples, then enter the testing data that has been prepared into the confusion matrix so that it gets results.

Table 2. Accuracy by K-Fold Validation

| Prediction Class | Observed Class | Not Feasible | Feasible |
|------------------|----------------|--------------|----------|
| Not Feasible     | 146            | 6            |
| Feasible         | 18             | 81           |

By testing the SVM algorithm (Support Vector Machine), there are 254 predictable data consisting of data mining and data testing obtaining false negative results on not feasible predictions, but the truth is feasible with a value of 81.82%, this means 206 predicted not feasible data but in the truth it is feasible to be given credit. So the error rate obtained is 19%, in this case 48 data are predicted to be fluent but in the truth it is not feasible to be given credit.

3.2 The results of using K-Fold Validation 7.

This test is done by dividing 7 parts of 254 customer credit data to be tested. The 7 sections consist of 6 parts training data and 1 part testing data. Confusion matrix model will form a matrix consisting of true positive or positive tuples and true negative or negative tuples, then enter the testing data that has been prepared into the confusion matrix so that it gets results.

Table 3. Accuracy by K-Fold Validation

| Predicate Class | Observed Class | Not Feasible | Feasible |
|-----------------|----------------|--------------|----------|
| Not Feasible    | 145            | 6            |
| Feasible        | 19             | 81           |
There is an ROC graph with AUC (Area Under Curve) value with accuracy results with K-Fold Validation 7 of 0.960% with an accuracy value of Excellent Classification.

![Fig 4. AUC with K-Fold Validation 7](image)

![Fig 5. Accuracy using K-Fold Validation 7](image)

### 3.3 Rekapitulation of the results.

The purpose of this research is to test the accuracy of predictive analysis of creditworthiness using the SVM (Support Vector Machine) Algorithm. The data analyzed are credit data for cooperative customers, which are all data that have been approved by the Cooperative “Daruzzakah Rensing”. From trials using k-fold validation 4 and 7, found the value of sensitivity, specificity, Ppv, and Npv, as shown as table below.

| Item      | K-Fold Validation 4 | K-Fold Validation 7 |
|-----------|---------------------|---------------------|
| Accuracy  | 90.42%              | 90.02%              |
| Sensitivity | 89.02%              | 88.41%              |
| Specificity | 93.10%              | 93.10%              |
| Ppv       | 96.05%              | 96.03%              |
| Npv       | 81.82%              | 81.00%              |

Based on the table above, it can be seen the accuracy value of the first trial with K-Fold Validation 4 is 90.42%, the accuracy value of the second trial with K-Fold Validation 7 is 90.02%, and the difference in accuracy between the two is 0.4%.

SVM has a good ability in solving data mining problems even with a limited sample. Experiments using the best SVM method produce an accuracy value of 90.42%. These results are obtained by the K-Fold Validation 4 method, where data is divided into 4 parts for training and testing. From this, it can be seen that the success of SVM is strongly influenced by choosing the right attribute. The more attributes and information used, it can reduce the level of accuracy that is higher.

### 4. Conclusion

From the implementation and discussion that have been carried out in this research, it can be concluded that in determining the feasibility of granting credit can be predicted and evaluated by utilizing data mining techniques using the SVM (Support Vector Machine) Algorithm. This method can analyze performing and non-performing loans, which are done by performing calculations by the performance of the SVM (Support Vector Machine) Algorithm to produce a model so that it is included in the excellent classification category, with proof of accuracy obtained at 90.42% and AUC at 0.957. From the two experiments conducted it can be seen that the success of SVM is greatly influenced by the selection of appropriate attributes. The more attributes and information used, it can increase the level of accuracy that is higher.

### 5. References

[1] W. Alves, P. Ferreira, and M. Araújo, “Journal of Co-operative Organization and Management Mining co-operatives : A model to establish a network for sustainability,” *J. Co-op. Organ. Manag.*, vol. 7, no. 1, pp. 51–63, 2019.

[2] E. Tremblay, A. Hupper, T. M. Waring, and S. George, “Journal of Co-operative Organization
and Management Co-operatives exhibit greater behavioral cooperation than comparable businesses: Experimental evidence,” *J. Co-op. Organ. Manag.*, vol. 7, no. 2, p. 100092, 2019.

[3] E. Tremblay, A. Hupper, and T. M. Waringa, “Cooperatives governing energy infrastructure: A case study of Berlin’s grid,” vol. 7, no. March, pp. 1–9, 2019.

[4] H. Sun, G. Lv, J. Mo, X. Lv, G. Du, and Y. Liu, “Optik Application of KPCA combined with SVM in Raman spectral discrimination,” *Opt. - Int. J. Light Electron Opt.*, vol. 184, no. February, pp. 214–219, 2019.

[5] D. Dou, W. Wu, J. Yang, and Y. Zhang, “Classification of coal and gangue under multiple surface conditions via machine vision and relief-SVM,” *Powder Technol.*, vol. 356, pp. 1024–1028, 2019.

[6] Y. Ge, S. Zhao, and X. Zhao, “Genomics A step-by-step classification algorithm of protein secondary structures based on double-layer SVM model,” *Genomics*, no. May, pp. 1–6, 2019.

[7] B. Xu, S. Shen, F. Shen, and J. Zhao, “Locally linear SVMs based on boundary anchor points encoding,” *Neural Networks*, vol. 117, pp. 274–284, 2019.

[8] J. Zhao, P. Wang, and J. C. S. Lui, “Optimizing node discovery on networks: Problem definitions, fast algorithms, and observations,” *Inf. Sci. (Ny)*., vol. 477, pp. 161–185, 2019.

[9] D. G. Hodges, B. P. Chapagain, K. L. Kline, and V. H. Dale, “Dataset of forest landowner survey to assess interest in supplying woody biomass in two Southeastern United States fuelsheds,” *Data Br.*, vol. 27, p. 104674, 2019.

[10] C. Aitken, Y. Chang, P. Buzzini, G. Zadora, and G. Massonnet, “The evaluation of evidence for microspectrophotometry data using functional data analysis,” *Forensic Sci. Int.*, vol. 305, p. 110007, 2019.

[11] R. Ibrahim, N. M. Jamal, R. I. Esmael, N. Zakuan, and N. Mohd, “ScienceDirect effectiveness b in Industry 4.0: c Trade-off Costing models for capacity optimization between used capacity and operational efficiency of business performance from the perspective manufacturing strategies: fit manufacturing and overall equipment manufacturing strategies: fit manufacturing and overall equipment understanding of business performance from the perspective of,” *Procedia Manuf.*, vol. 22, pp. 998–1006, 2018.

[12] M. W. Weaver, K. Crossan, H. B. Tan, and S. E. Paxton, “A systems approach to understanding the perspectives in the changing landscape of responsible business in Scotland,” *Eur. J. Oper. Res.*, vol. 268, no. 3, pp. 1149–1167, 2018.

[13] S. Nestorov, B. Jukić, N. Jukić, A. Sharma, and S. Rossi, “Generating insights through data preparation, visualization, and analysis: Framework for combining clustering and data visualization techniques for low-cardinality sequential data,” *Decis. Support Syst.*, vol. 125, no. July, p. 113119, 2019.

[14] N. Konstantinou and N. W. Paton, “Feedback driven improvement of data preparation pipelines,” *Inf. Syst.*, no. xxxx, p. 101480, 2019.

[15] O. D. Caicedo-salcedo, D. P. Vargas-delgadillo, L. Giraldo, and J. C. Moreno-piraj, “Data in brief Data of preparation and characterization of activated carbon using two activant agents and mango seed as precursor material,” vol. 27, 2019.