Optimization of Decision Tree with PSO on Sharia Cooperative Customer Funding

Eka Rahmawati\textsuperscript{1}\textsuperscript{*} and Candra Agustina\textsuperscript{2}
\textsuperscript{1}Sistem Informasi, Universitas Bina Sarana Informatika, Jakarta, Indonesia
\textsuperscript{2}Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika, Jakarta, Indonesia
E-mail: eka.eat@bsi.ac.id

Abstract. Credit is a service for Sharia Cooperatives to provide funding to customers who are paid in installments. The accuracy of the customer paying the payments is a determining factor for the smooth operation of the Sharia Cooperative. In addition to timeliness, the ability of customers to pay installments is also the most critical factor for credit returns. Bad credit is a significant threat for Sharia Cooperatives where customers cannot pay payments according to agreed agreements. That makes the prediction of the smooth operation of Sharia Cooperative customers needed. One way that can be done to make predictions is to apply data mining techniques. To make a prediction, several attributes of the customer are used, such as gender, age, status, residency status, number of dependents owned, education, employment, remittance, ceiling, type of loan, credit period, method of payment and collateral used. The research will use the J48 as one of the Decision Tree algorithms with particle swarm optimization techniques to improve algorithm performance. The accuracy of J48 with PSO is lower than without PSO.

1. Introduction
Islamic cooperatives become one of the providers of financial product services to improve welfare. The concept offered in sharia is done with a profit-sharing system so that it does not burden the members. Credit is a service that is used to help members to realize various objectives, such as for business interests or business development. In Indonesia, cooperative users continue to increase every year. A survey conducted by the Indonesian Central Statistics Agency showed that the number of cooperatives experienced a significant increase. Data were obtained from 34 provinces, with a total of 7612 organizations.

The provision of credit becomes the most significant cooperative effort, so the cooperative needs to provide an assessment of customers who apply for credit loans and ensure that customers can return credit that has been received\cite{1}. The main problem for credit service providers, especially in cooperatives, is the existence of bad credit. What happens when a customer cannot pay installments according to the amount agreed at the beginning. Lousy credit can hamper cooperative operations, so the prediction of prospective customers at the outset needs to be done. One way that can be done is with data mining techniques. The algorithm used is J48. To implement the algorithm, we need data for testing. This study uses data from Sharia cooperatives with 13 attributes. However, before measurement, the data will first be filtered with the participation swarm optimization technique.
2. Related Works

In a previous study, Agustina[2] had researched the analysis of time deposit customers. Customer data is processed using C45, Naïve Bayes, and Neural Network algorithms combined with Particle Swarm Optimization. The results obtained are as follows:

Table 1. Data Processing Results.

| Algorithms  | Accuracy |
|------------|----------|
| C45 + PSO  | 89.72    |
| NN + PSO   | 97.04    |
| NB + PSO   | 96.96    |

The research shows that PSO can be applied in the C45 algorithm, Neural Network, and Naïve Bayes. Two of the three algorithms have an accuracy above 95%. The study did not include the results of processing the algorithm before PSO was applied. Besides, data processing with the J48 algorithm has not been done yet. The data taken from abroad may be inappropriate if used in Indonesia. Research also conducted by Nurajija and Riana[3] with the results of T-Test Testing (Table 6) shows a significant difference between the classification of algorithm B for Naïve Bayes (77.29%) with C for Decision Tree (89.02%) and D for SVM (89.86%). However, the tests carried out did not use PSO. The J48 and PSO algorithms are used by Kaur and Sharma[4] to classify emails into spam and non-spam groups. The results obtained in this study are the accuracy level of 99.13%. This shows that the J48 algorithm can be combined with PSO in processing data sets with accuracy approaching 100%.
3. Materials and Method

3.1. Decision Tree

The decision tree is an algorithm that is widely used for classification problems, which consists of several nodes, namely, the tree’s root, internal node, and leaves. The concept of entropy is used to determine which attributes a tree will split (split)[5].

\[ \text{Entropy}(S) = - p1 \log_2 p1 - p2 \log_2 p2 \] (1)

Where: \( p1 \) is the proportion of sample or group 1 that will be paired with the portion of \( p2 \) the proportion of group 2. J48 is one of the decision tree induction algorithms that is part of the Java open-source implementation of the C4.5 WEKA data mining tool[6]. The classification steps are applied when the decision tree is built on the J48 algorithm. Trees will be made in a way to divide and conquer top-down recursively. Previous research has been conducted by Much Aziz Muslim, Aldi Nurzahputra, Budi Prasetyio about credit applications, which are used as decision support in determining credit requests. Validation in this study uses ten-fold cross-validation. While a confusion matrix measures the measurement of accuracy. The experimental results show there is an increase in the efficiency of the application of the C4.5 algorithm. The Syariah cooperative dataset is not used in this study. The study also has not used the PSO technique for attribute selection [7]. The shortcomings of the J48 algorithm are as follows[8]: 1. The possibility of an empty or valuable branch approaching Zero. 2. Getting insignificant value for classification. Getting too many results and details. This can be ruled out because many attributes produce trees with many branches as well. This can be anticipated by stopping the process before reaching the maximum value. Then it can also be done by pruning.

3.2. Particle Swarm Optimization

PSO performs activity modeling to find the best solution in a solution space as the activity of flying groups of particles in a solution space. The position of the particles in the solution space becomes a candidate for the solution containing the variables used in optimization. Each of these positions will have a relationship with a value called an objective value or fitness value. This value is calculated based on the actual function of the optimization problem to be solved[9]. From the existing attributes to determine predictions, it needs to be selected to determine the dominant qualities. Research has been conducted by Saruni Dwiasnati and Yudo Devianto using PSO in insurance customer predictions. PSO can increase accuracy value. This study has not used a database of Islamic cooperatives. The algorithm used for classification is C.45 [10].

3.3. Sharia Cooperative

Cooperative is an association of entities with a membership system that works together peacefully in doing business to expand or improve the welfare of its members [11]. Cooperatives lead to comfort and improve the welfare of each of its members. Cooperatives, in general, have the basic principles of the capitalist system, which bases on the interest system, which is known as syirkah ta’awuniyah, and becomes a sharia cooperative. Islamic cooperatives are intended to be free from banking with the principle of profit-loss (profit and loss sharing) and qard-hasan system (QHS) as one of the main principles in its business activities [12]. The cooperative as the party providing the funds is called a creditor. Then the customer receiving the credit (the debtor). The purpose of granting credit to customers in the cooperative [13] are as follows: 1. Assist the community in the availability of funds, especially for venture capital 2. Helping people avoid loan sharks who provide high-interest loans 3. Helping every member of the cooperative save money The granting of credit must, of course, be with the application of several requirements, which will then form a database consisting of several attributes. From each data collected can be classified to predict creditworthiness.
3.4. Research Methodology

- Data Understanding: Customer financing at Sharia Cooperatives leads to the risk of bad credit. Data that will be processed with data mining classification techniques are prepared by several attributes that will be used. The dataset comes from customer financing at Islamic cooperatives that provide credit services. Each customer data has been divided into two classes, namely identified as lousy credit and current credit. The amount collected consists of 869 records.

- Data Preparation: The initial preparation is done by selecting attributes. There are 14 attributes and one label in the customer financing dataset. The PSO technique does optimization. Attributes consisting of ID, gender, age, status, residency status, number of dependents, education, employment, income, ceiling, repayment period, and guarantee will be deducted based on the most critical attributes.

- Modelling: After the PSO is implemented in the dataset, the data will then be classified with the data mining algorithm. Testing will be carried out using J48. The test results will be presented in the form of figures and tables.

- Evaluation: The results of modeling are then evaluated through accuracy, AUC, and confusion matrix. The validation process is repeated, starting from the first part as test data to the tenth section, so that all data in the dataset is tested.

4. Results

The initial classification is done by applying the J48 algorithm without using PSO. The classification results for the Sharia Cooperative customer dataset show an accuracy rate of 88.38% and an AUC value of 0.868. The confusion matrix for this test is shown in Table 2.

Testing is done by selecting attributes using the PSO technique. Of the 13 characteristics, two attributes affect customer credit, namely income and balance. The most important attribute
on the smoothness of the customer’s credit is the ceiling. Then, the J48 algorithm is applied to classify customers who have credit at the Syariah Cooperative. Fig.3. Shows that customers with income below 3 million have the potential to be bad credit. However, it will depend on the amount of ceiling taken.

![Decision Tree Credit Analysis with PSO](image)

The accuracy value in testing using the J48 algorithm is 82.51%. The AUC value for this study is 0.797, which means the data used for testing is good. The test results with J48 can also be seen through the confusion matrix in Table 3.

![Table 3. Confussion Matrix Testing with PSO](image)

Based on previous research, PSO is effective for increasing accuracy. However, this did not occur in the application of J48 for Sharia cooperative customer data. After reviewing, the initial data uses 13 attributes, namely gender, age, status, residence status, number of dependents, education, employment, income, ceiling, type of loan, period, method of payment, and collateral. However, after PSO is applied there are only 2 attributes that can be used to determine the estimate of bad or current loans, namely income and ceiling. This makes the accuracy value decreases, although not significantly.

5. Conclusion
PSO can be used for attribute selection on the decision tree algorithm application for the classification of sharia cooperative customer data. The accuracy obtained is 82.51%, with an AUC value of 0.797. Before applying PSO, the accuracy value obtained in the study was 88.38%, and the AUC value was 0.868. After the application of PSO, there was a decrease of 5.87% for the accuracy value and a reduction of 0.071 for the AUC value.
References

[1] Amar S S and Nurfadila D 2016 Analisis Standar Operasional Prosedur Pemberian Kredit Pada Anggota Koperasi Pegawai Republik Indonesia (Kpri) Bhinneka Karya Pamkakn J. Akunt. dan Investasi 1 77–92

[2] Agustina C 2019 Analisa Nasabah Potensial Tabungan Deposito Berjangka Menggunakan Teknik Klasifikasi Data Mining J. Teknl. Inf. dan Terap. 5 105–12

[3] Nurajijah and Riana D 2019 Algoritma Naïve Bayes , Decision Tree dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah Naïve Bayes , Decision Tree , and SVM Algorithm for Classification of Sharia J. Teknl. dan Sist. Komput. 7 77–82

[4] Kaur H and Sharma A 2016 Novel Email Spam Classification using Integrated Particle Swarm Optimization and J48 Int. J. Comput. Appl. 149 23–7

[5] Fitri S 2014 Perbandingan Kinerja Algoritma Klasifikasi Naïve Bayesian , Lazy-Ibk , Zero-R , Dan Decision Tree- J48 Dasi 15 33–7

[6] Muslim M A, Nurzaliputra A and Prasetyo B 2018 Improving accuracy of C4.5 algorithm using split feature reduction model and bagging ensemble for credit card risk prediction 2018 Int. Conf. Inf. Commun. Technol. ICOIACCT 2018 2018-Janua 141–5

[7] Seif H 2016 Naive Bayes and J48 Classification Algorithms on Swahili Tweets: Performace Evaluation Int. J. Computer Sci. Inf. Secur. 14 1–4

[8] anaN N S and thri V G 2018 Performance and Classification Evaluation of J48 Algorithm and Kendall’s Based J48 Algorithm (KNJ48) Int. J. Comput. Trends Technol. 59 73–80

[9] Ana D, Wati R and Rochman Y A 2013 Model Penjadwalan Matakuliah Secara Otomatis Berbasis Algoritma Particle Swarm Optimization (PSO) 2 22–31

[10] Dwiasnati S and Devianto Y 2019 Optimasi Prediksi Keputusan Calon Nasabah Potensial Berbasis Particle Swarm Optimization J. Inform. 6 286–92

[11] Doloh R and Oktafi R 2018 Peranan Koperasi Dalam Meningkatkan Kesejahteraan Ekonomi Masyarakat Ditinjau Dari Hukum Ekonomi Islam Univ. Muhammadiyah Sidoarjo 2–6

[12] Aziz A 2018 Analisis Pengaruh Pengelolaan Manajemen Dan Kinerja Koperasi Syariah Dalam Menyalurkan Pembiayaan Produktif Di Sektor Usaha Kecil Survey di BMT-BMT Cirebon Inklusif (Jurnal Pengkaj. Penelt. Ekol. Dan Huk. Islam. 3 135–46

[13] Isa I G T and Hartawan G P 2017 Perancangan Aplikasi Koperasi Simpan Pinjam Berbasis Web (Studi Kasus Koperasi Mitra Setia) J. Ilm. Ilmu Ekon. (Jurnal Akuntansi, Pajak dan Manaj. 5 139–51