

ANALYZING THE POSSIBILITY OF DELAYS OF SPP PAYMENTS WITH C4.5 ALGORITHM (CASE STUDY OF POLITEKNIK TEDC BANDUNG)

Tri Herdiawan Apandi¹⁾, Roby Bayu Maulana²⁾, Rian Piarna³⁾, Dwi Vernanda^{4)**}

^{1, 3, 4)} Information Systems Study Program
Politeknik Negeri Subang, West Java, Indonesia
www.polsub.ac.id

tritherdiawan@polsub.ac.id¹⁾; rianpiarna@polsub.ac.id³⁾; yogurt.nda@gmail.com⁴⁾

²⁾ Informatics Engineering Study Program
Politeknik TEDC Bandung, West Java, Indonesia
www.poltektedc.ac.id
bayuroby@gmail.com²⁾



The work is distributed under the Creative Commons Attribution-Non-Commercial International 4.0 License

Abstract—Payment of tuition as one of the sources of funds, plays an important role in the sustainability of the operations of higher education. The problem that arises is that students are not often late to make payments in a timely manner. One of the factors causing the many cases of late payment of tuition fees due to lack of policy and decisive action on the part of the campus when students are late in making payments, besides the factors of parents and students also have an influence on the delay. The purpose of this study is to classify students who are late and timely in making SPP payments using the C4.5 algorithm. From the total sample used then divided into 4 partitions, partition 1 for 90% training data and 10% testing data, partition 2 for 80% training data and 20% testing data, and partition 3 for 70% training data and 30% testing data, and partition 4 for 60% training data and 40% testing data. The classification results of the C4.5 algorithm are evaluated and validated with cross validation and confusion matrix to determine the accuracy of the C4.5 algorithm in predicting late SPP payments. Based on the comparison of the results of evaluations and validations conducted, it shows that data partition 2 has a better level of accuracy than the other partitions, which is 75%.

Keywords: Data Mining, Decision Tree (C4.5), SPP.

Intisari—Pembayaran uang kuliah sebagai salah satu sumber dana, memegang peranan penting dalam keberlangsungan operasional perguruan tinggi. Masalah yang timbul adalah mahasiswa tak jarang terlambat untuk melakukan pembayaran secara tepat waktu. Salah satu faktor penyebab banyaknya kasus keterlambatan pembayaran SPP

karena kurangnya kebijakan dan tindakan tegas dari pihak kampus ketika mahasiswa terlambat melakukan pembayaran, selain itu faktor orang tua dan mahasiswa juga memiliki pengaruh dalam keterlambatan tersebut. Tujuan penelitian ini adalah membuat klasifikasi mahasiswa yang terlambat dan tepat waktu dalam melakukan pembayaran SPP dengan menggunakan algoritma C4.5. Dari total sampel yang digunakan kemudian dibagi menjadi 4 partisi, partisi 1 untuk 90% data training dan 10% data testing, partisi 2 untuk 80% data training dan 20% data testing, dan partisi 3 untuk 70% data training dan 30% data testing, dan partisi 4 untuk 60% data training dan 40% data testing. Hasil klasifikasi dari algoritma C4.5 dievaluasi dan divalidasi dengan cross validation dan confusion matrix untuk mengetahui tingkat akurasi Algoritma C4.5 dalam memprediksi keterlambatan pembayaran SPP. Berdasarkan perbandingan hasil evaluasi dan validasi yang dilakukan menunjukkan bahwa data partisi 2 memiliki tingkat akurasi yang lebih baik dibanding partisi lainnya yaitu sebesar 75%.

Kata kunci : Data Mining, Decision Tree (C4.5), SPP.

INTRODUCTION

The source of funds plays an important role in the sustainability of operational activities of a college. One source of funding at private tertiary institutions is from paying tuition fees. So it cannot be denied at private universities that it is very important to pay tuition fees by students, this is because in the interests of the campus more funds are spent by the campus itself, so it should be the

tuition payment should be managed properly (Maulana, 2016).

TEDC Polytechnic as a provider of lecture facilities has a policy in making payments for Educational Development Donations (SPP). At the beginning of the semester, students must register for new students, and register for students who have previously been registered. Students can pay the semester fees in full, or can make installments with the first payment of 50% of the number of tuition fees per semester paid at the time of registration, and continued payments at 75% UTS and paid at the time of UAS. The campus certainly expects timely payment, because SPP payment money is used for campus infrastructure development. The problem that arises is that students often neglect to make payments on time. At the TEDC Polytechnic itself, there are still often a late payment of tuition fees (Maulana, 2016), although the number of delays is not much it increases every semester. Based on historical data, the TEDC Polytechnic's financial department can make predictions about the possibility of late payment by using data classification using data mining techniques. Data mining can be defined as a process of finding a pattern in data (Witten & Frank, 2005).

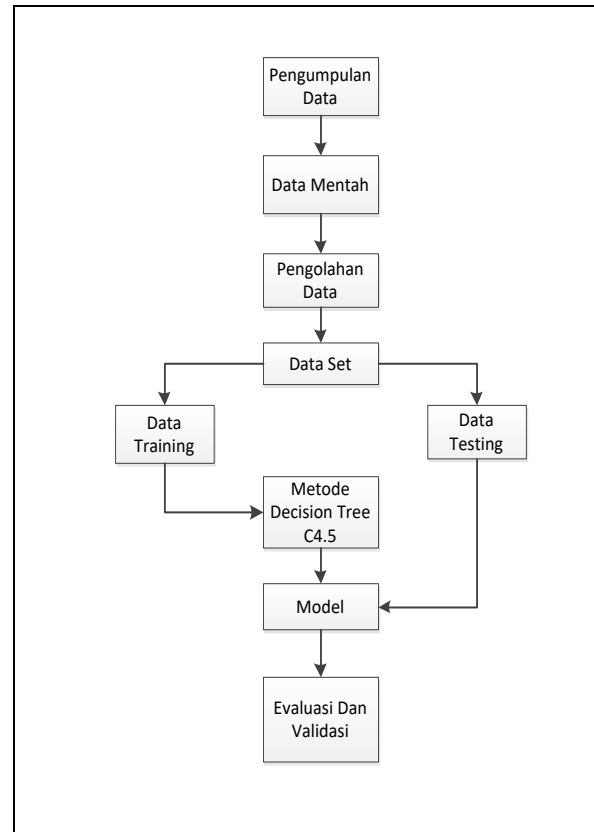
There are several techniques in data mining, one of which is classification. Classification is the process of finding a model (or function) that describes and distinguishes classes of data or concepts (Zhang, Williams, & Wang, 2018). Decision tree C4.5 is a classification algorithm. The learning and classification process using this method is very simple and fast. In general, classification using a decision tree has a good level of accuracy. But it depends on the data you have (Han, Pei, & Kamber, 2011).

Based on the description above, an analysis will be made to classify the late payment of tuition fees. In this study, data mining techniques are used to find patterns of student tuition payments that already exist, then serve as the basis for predicting payments in the next semester. This research will apply the C4.5 algorithm which is one of the decision tree methods that are easy to understand and can build decision trees quickly.

MATERIALS AND METHODS

A. Research Method Models

This Research Model Model was made as a travel flow to facilitate the writer in conducting final project research. The research method model can be seen in Figure 2 below.



Sumber: (T. H. Apandi, Maulana, Piarna, & Vernanda, 2019)

Figure 1 Model Research Methods

B. Initial Data Processing

Survey data in the form of questionnaires were distributed to students of the TEDC Bandung Polytechnic Informatics Engineering study program which consisted of several aspects or assessment factors. Research data processing will be processed using RapidMiner 5.2 to get the pattern of the decision tree and accuracy to analyze the possibility of late payment of SPP money. At this stage the variables will be determined, where these variables will produce an analysis to form the decision tree of the case. These variables were obtained from experts, from the results of interviews with financial staff obtained variable late payment of SPP.

C. Proposed Model

According to the Gartner Group, data mining is a process of finding meaningful relationships, patterns, and tendencies (T. Apandi, Piarna, & Vernanda, 2018). Proses dilakukan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005).

On the other hand, data mining is a term used to describe the discovery of knowledge in a

database. Data mining is a process that uses statistical, mathematical, artificial intelligence, and machine learning techniques to extract and identify useful information and related knowledge from various large databases (Kusrini & Luthfi, 2009).

The method proposed to predict student satisfaction with the performance of lecturers of Informatics Engineering study programs at TEDC Bandung Polytechnic is to use the decision tree method (C4.5). This method was chosen to manage data whose attributes and classes are not of definite value (Trabelsi, Elouedi, & Lefevre, 2019).

The stages of the decision tree method (C4.5) are as follows (Swastina, 2013):

1. Prepare training data.
2. Determine the roots of the tree.
3. Calculate Gain.

To choose the attribute as the root, based on the highest Gain value of the existing attributes. To calculate Gain you can use formula 1 as follows:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \dots\dots (1)$$

1. Repeat the second step until each branch is fulfilled. Meanwhile, for calculating Entropy values, you can use formula 2 as follows:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \dots\dots\dots (2)$$

2. The decision tree participation process will stop when all branches in node N get the same class.

RESULTS AND DISCUSSION

A. Test Results

The data used in this study amounted to 102 data obtained from questionnaires distributed to students of the Informatics Engineering study program in the 3rd, 5th and 7th semester of the regular class of the 2015/2016 TEDC Bandung Polytechnic. The data is used by the authors to be analyzed and processed to help provide information on the financial section of the TEDC polytechnic.

Of the total sample used then divided into 4 partitions, partition 1 for 90% training data and 10% testing data, partition 2 for 80% training data and 20% testing data, and partition 3 for 70% training data and 30% testing data, and partition 4 for 60% of training data and 40% of testing data.

Based on the sample data tested using the decision tree method with a testing data of 20%, an accuracy rate of 50% is obtained. As shown in Table 1 below.

Table 1 Performance Vector Data Partitions 2
Akurasi: 50%

	<i>True on time</i>	<i>True late</i>	<i>Class precision</i>
Pred. on time	9	6	60,00%
Pred. late	4	1	20,00%
<i>class recall</i>	69,23%	14,29%	

Sumber: (T. H. Apandi et al., 2019)

Based on sample data that has been tested using the decision tree (C4.5) method then it will be validated (Aguilar-Chinea, Rodriguez, & Exposito, 2019). Validation is done to get the best results so that it can be applied to predict the possibility of late payment of tuition fees at the Polytechnic of TEDC Bandung.

The results of the decision tree model partition 2 data validation using cross-validation obtained an accuracy rate of 75%. As stated in table 2 below.

Table 2 Performance Vector Data Partitions 2
Akurasi: 75%

	<i>True on time</i>	<i>True late</i>	<i>Class precision</i>
Pred. on time	12	4	75,00%
Pred. late	1	3	75,00%
<i>class recall</i>	92,31%	42,86%	

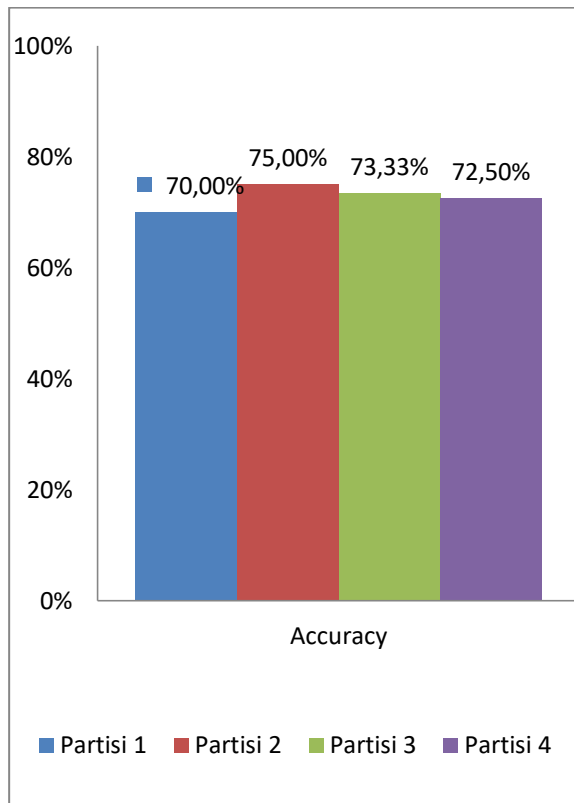
Sumber: (T. H. Apandi et al., 2019)

B. Comparison of Testing Results

Based on the whole data that has been tested, the authors get a comparison of the results between partition 1, partition 2, partition 3, and partition 4 to analysis the possibility of late payment of tuition fees at TEDC Bandung Polytechnic with the following results:

Comparison of accuracy between Partition 1, Partition 2, Partition 3 and Partition 4. A Comparison of accuracy can be seen in Figure 2 below.

From Figure 2 below shows that the partition 2 test data has a greater degree of accuracy than the other partition data that is equal to 75%. Based on the results of the above comparison the authors suggest using a decision tree pattern on partition 2 data to be used as a pattern to predict late payment of SPP.



Sumber: (T. H. Apandi et al., 2019)

Figure 2 Comparison of Data Accuracy

CONCLUSION

Based on the results of the research that has been done, several conclusions can be drawn, among others, from the results of the data set partition test found the highest level of accuracy in the data partition 2 is 80% training data and produces an accuracy of 75.00%. Thus the model obtained from partition 2 can be used as training data to predict new data. Academic service attributes have a considerable influence in all partitions tested, after an experiment using RapidMiner that the academic service attributes are at the top node. From this study, the authors also obtained a pattern of rules that can be used to predict students who are indicated on time or late when making tuition payments.

The author's suggestion for the continuation of this study is expected in subsequent studies, more records are used, making it possible to find patterns beyond expectations or characteristics of the possibility of late payment of SPP. It is expected that further research can be developed using similar types of data but with other data mining methods, such as clustering that can be used as a comparison

REFERENSI

- Aguilar-Chinea, R., Rodriguez, I. C., & Exposito, C. (2019). Using a decision tree algorithm to predict the robustness of a transshipment schedule. *Procedia Computer Science*, 149, 529–536. <https://doi.org/https://doi.org/10.1016/j.procs.2019.01.172>
- Apandi, T. H., Maulana, R. B., Piarna, R., & Vernanda, D. (2019). *Laporan Akhir Penelitian: Menganalisis Kemungkinan Keterlambatan Pembayaran Spp Dengan Algoritma C4.5 (Studi Kasus Politeknik TEDC Bandung)*. Subang.
- Apandi, T., Piarna, R., & Vernanda, D. (2018). Optimization of Feature Selection Using Genetic Algorithms to Increase Payment Delay Prediction Results (Subang Polytechnic State Case Study). In *The 1st International Conference on Computer Science and Engineering Technology Universitas Muria Kudus* (pp. 807–813). Kudus: EAI. Retrieved from <https://eudl.eu/doi/10.4108/eai.24-10-2018.2280507>
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). USA: Morgan Kaufmann.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining* (1st ed.). Yogyakarta: Andi Publisher. Retrieved from <http://andipublisher.com/produk-0907003050-algoritma-data-mining.html>
- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data mining*. New Jersey: John Willey & Sons, Inc.
- Maulana, R. B. (2016). Penerapan Data mining untuk Menganalisa Kemungkinan Keterlambatan Pembayaran SPP (Studi Kasus Politeknik TEDC Bandung).
- Swastina, L. (2013). Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. *Gema Aktualita*, Vol. 2(No. 1), 2–3.
- Trabelsi, A., Elouedi, Z., & Lefevre, E. (2019). Decision tree classifiers for evidential attribute values and class labels. *Fuzzy Sets and Systems*, 366, 46–62. <https://doi.org/https://doi.org/10.1016/j.fss.2018.11.006>

- Witten, & Frank, E. (2005). *Data mining Practical Machine Learning Tools and Techniques*. San Francisco: Elsevier (2nd ed., Vol. 2).
- Zhang, J., Williams, S., & Wang, H. (2018). Intelligent computing system based on pattern recognition and data mining algorithms. *Sustainable Computing: Informatics and Systems*, 20, 192–202. <https://doi.org/https://doi.org/10.1016/j.susc.2017.10.010>

