

MENGANALISIS KEMUNGKINAN KETERLAMBATAN PEMBAYARAN SPP DENGAN ALGORITMA C4.5 (STUDI KASUS POLITEKNIK TEDC BANDUNG)

Tri Herdiawan Apandi¹⁾, Roby Bayu Maulana²⁾, Rian Piarna³⁾, Dwi Vernanda⁴⁾**

^{1, 3, 4)}Program Studi Sistem Informasi, Politeknik Negeri Subang
www.polsub.ac.id

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

²⁾Program Studi Teknik Informatika, Politeknik TEDC Bandung
www.poltektedc.ac.id
bayuroby@gmail.com²⁾



Ciptaan disebarluaskan di bawah Lisensi Creative Commons Atribusi-NonKomersial 4.0 Internasional.

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.

PENDAHULUAN

Sumber dana memegang peranan penting dalam keberlangsungan kegiatan operasional suatu perguruan tinggi. Salah satu sumber dana pada perguruan tinggi swasta yaitu dari pembayaran uang kuliah. Sehingga tidak bisa dipungkiri pada perguruan tinggi swasta sangat penting pembayaran uang kuliah oleh

mahasiswa, hal ini dikarenakan untuk kepentingan kampus lebih banyak dikeluarkan dana oleh pihak kampus itu sendiri, maka sudah semestinya pembayaran uang kuliah harus dikelola dengan baik (Maulana, 2016).

Politeknik TEDC sebagai penyedia fasilitas perkuliahan memiliki kebijakan dalam melakukan pembayaran *Sumbangan Pembinaan Pendidikan* (SPP). Setiap awal semester mahasiswa wajib melakukan registrasi bagi mahasiswa baru, dan heregistrasi bagi mahasiswa yang sebelumnya sudah tercatat. Mahasiswa dapat membayar uang semester secara penuh, atau dapat melakukan cicilan dengan bayaran pertama 50% dari jumlah uang SPP per-semesternya dibayar pada saat melakukan heregistrasi, dan dilanjutkan pembayaran pada saat UTS sebesar 75% dan dilunasi pada saat UAS. Pihak kampus tentu mengharapkan pembayaran tepat waktu, karena uang pembayaran SPP digunakan untuk pembangunan infrastruktur kampus. Masalah yang timbul adalah mahasiswa tak jarang mengabaikan untuk melakukan pembayaran secara tepat waktu. Di Politeknik TEDC sendiri masih sering terjadi keterlambatan pembayaran SPP (Maulana, 2016), walaupun jumlah keterlambatan tidak banyak namun meningkat tiap semester nya. Berdasarkan data historis bagian keuangan Politeknik TEDC dapat dibuat rekomendasi prediksi kemungkinan keterlambatan pembayaran yaitu menggunakan klasifikasi data dengan menggunakan teknik *data mining*. *Data mining* dapat didefinisikan sebagai suatu proses untuk menemukan suatu pola di dalam data (Witten & Frank, 2005).

Terdapat beberapa teknik dalam *data mining*, salah satunya yaitu klasifikasi. Klasifikasi adalah proses menemukan model (atau fungsi) yang menggambarkan dan membedakan kelas data atau konsep (Zhang, Williams, & Wang, 2018). *Decision tree* C4.5 merupakan salah satu algoritma klasifikasi. Pembelajaran dan proses pengklasifikasian menggunakan metode ini sangat sederhana dan cepat. Secara umum pengklasifikasian menggunakan *decision tree* memiliki tingkat akurasi yang baik. Namun tergantung data yang dimiliki (Han, Pei, & Kamber, 2011).

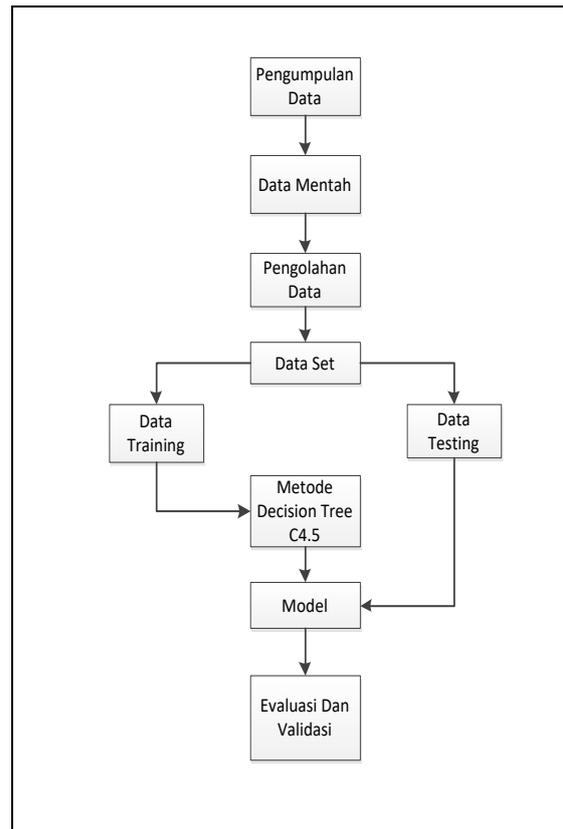
Berdasarkan uraian di atas akan dibuat analisis untuk mengklasifikasi keterlambatan pembayaran uang SPP. Pada penelitian ini digunakan teknik *data mining* untuk menemukan pola pembayaran SPP mahasiswa yang sudah ada, kemudian dijadikan dasar untuk memprediksi pembayaran pada semester berikutnya. Penelitian ini akan menerapkan algoritma C4.5 yang merupakan salah satu metode *decision tree* yang mudah dimengerti

serta dapat membangun pohon keputusan dengan cepat.

BAHAN DAN METODE

A. Model Metode Penelitian

Model Metode Penelitian ini dibuat sebagai alur perjalanan untuk memudahkan penulis dalam melakukan penelitian tugas akhir. Model metode penelitian dapat dilihat pada gambar 2 berikut ini.



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

Gambar 1 Model Metode Penelitian

B. Pengolahan Data Awal

Data survei dalam bentuk kuesioner yang dibagikan kepada mahasiswa program studi teknik informatika Politeknik TEDC Bandung yang terdiri dari beberapa aspek atau faktor penilaian. Pengolahan data penelitian akan diolah menggunakan RapidMiner 5.2 untuk mendapatkan pola dari *decision tree* dan akurasi untuk menganalisa kemungkinan keterlambatan pembayaran uang SPP. Pada tahap ini akan dilakukan penentuan variabel-variabel, dimana variabel-variabel tersebut akan menghasilkan sebuah analisa untuk membentuk pohon keputusan dari kasus. Variabel-variabel ini didapat dari ahli, dari hasil wawancara dengan

staff keuangan didapatlah variabel keterlambatan pembayaran SPP.

C. Model Yang Diusulkan

Menurut Gartner Group *data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan (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).

Di sisi lain *data mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Kusrini & Luthfi, 2009).

Metode yang diusulkan untuk memprediksi kepuasan mahasiswa terhadap kinerja dosen program studi Teknik Informatika di Politeknik TEDC Bandung adalah dengan menggunakan metode *decision tree* (C4.5). Metode ini dipilih untuk mengelola data yang atribut dan kelasnya tidak bernilai pasti (Trabelsi, Elouedi, & Lefevre, 2019).

Tahapan metode *decision tree* (C4.5) adalah sebagai berikut (Swastina, 2013):

1. Menyiapkan *data training*.
2. Menentukan akar dari pohon.
3. Hitung *Gain*.

Untuk memilih atribut sebagai akar, didasarkan pada nilai *Gain* tertinggi dari atribut-atribut yang ada. Untuk menghitung *Gain* dapat menggunakan rumus 2.1 sebagai berikut:

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

4. Ulangi langkah kedua hingga setiap cabang terpenuhi.

Sementara itu, untuk penghitungan nilai *Entropy* dapat menggunakan rumus 2.2 sebagai berikut:

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

5. Proses partisipasi pohon keputusan akan berhenti saat semua cabang dalam *node* N mendapat kelas yang sama.

HASIL DAN PEMBAHASAN

A. Hasil Pengujian

Data yang digunakan dalam penelitian ini berjumlah 102 data yang didapatkan dari kuesioner yang disebarikan kepada mahasiswa program studi Teknik Informatika semester 3, 5 dan 7 kelas reguler tahun ajaran 2015/2016 Politeknik TEDC Bandung. Data tersebut digunakan oleh penulis untuk dianalisa dan diolah untuk membantu memberikan informasi pada bagian keuangan politeknik TEDC.

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*.

Berdasarkan data sampel yang diuji menggunakan metode *decision tree* dengan data *testing* sebesar 20% didapatkan tingkat akurasi sebesar 50%. Seperti terlihat pada tabel 1 berikut ini.

Tabel 1 *Performance Vector* Data Partisi 2
Akurasi: 50%

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

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

Berdasarkan data sampel yang telah diuji dengan menggunakan metode *decision tree* (C4.5) kemudian akan divalidasi (Aguilar-Chinea, Rodriguez, & Exposito, 2019). Validasi dilakukan untuk mendapatkan hasil yang terbaik agar dapat diterapkan untuk memprediksi kemungkinan keterlambatan pembayaran SPP di Politeknik TEDC Bandung.

Hasil validasi model *decision tree* data partisi 2 dengan menggunakan *cross validation* didapatkan tingkat akurasi sebesar 75%. Seperti tercantum pada tabel 2 berikut ini.

Tabel 2 *Performance Vector* Data Partisi 2

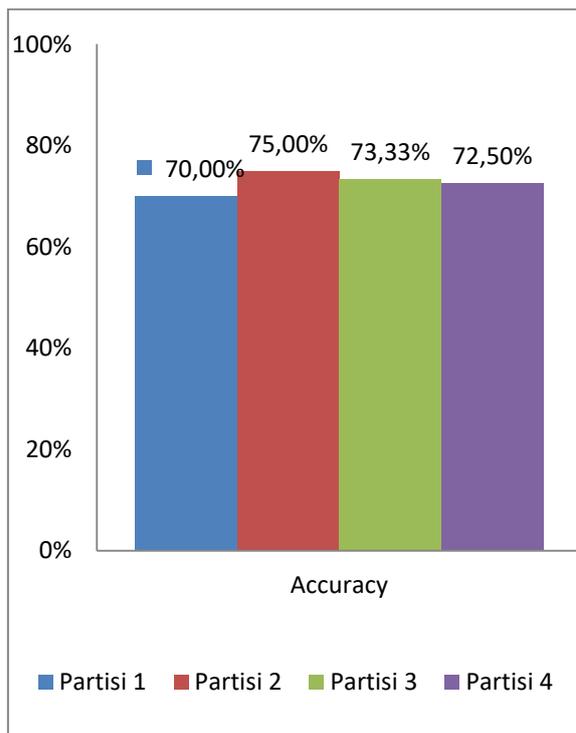
	<i>True</i> tepat waktu	<i>True</i> terlambat	<i>Class</i> <i>precision</i>
Pred. tepat waktu	12	4	75,00%
Pred. terlambat	1	3	75,00%
<i>class recall</i>	92,31%	42,86%	

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

B. Perbandingan Hasil Pengujian

Berdasarkan keseluruhan data yang telah diuji, penulis mendapatkan perbandingan hasil antara partisi 1, partisi 2, partisi 3, dan partisi 4 untuk menganalisa kemungkinan keterlambatan pembayaran SPP di Politeknik TEDC Bandung dengan hasil sebagai berikut:

1. Perbandingan akurasi antara Partisi 1, Partisi 2, Partisi 3, dan Partisi 4. Perbandingan akurasi dapat dilihat pada gambar 4.7 berikut ini.



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

Gambar 2 Perbandingan Akurasi Data

Dari gambar 2 diatas menunjukkan bahwa pengujian data partisi 2 memiliki tingkat akurasi yang lebih besar dibandingkan data partisi lainnya yaitu sebesar 75%. Berdasarkan hasil perbandingan di atas penulis menyarankan menggunakan pola pohon keputusan pada data partisi 2 untuk dijadikan pola memprediksi keterlambatan pembayaran SPP.

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan maka dapat diambil beberapa kesimpulan, antara lain dari hasil uji coba partisi data set ditemukan tingkat akurasi tertinggi pada data partisi 2 sebanyak 80% *data training* dan menghasilkan akurasi sebesar 75,00%. Dengan demikian model yang didapatkan dari partisi 2 ini dapat digunakan sebagai *data training* untuk memprediksi data yang baru. Atribut pelayanan akademik memiliki pengaruh yang cukup besar dalam semua partisi yang diuji, setelah dilakukan percobaan dengan menggunakan RapidMiner bahwa atribut pelayanan akademik berada pada *node* paling atas. Dari penelitian ini penulis juga memperoleh pola aturan yang dapat digunakan untuk memprediksi mahasiswa yang terindikasi tepat waktu atau terlambat ketika melakukan pembayaran SPP.

Adapun saran penulis terhadap kelanjutan dari penelitian ini adalah diharapkan pada penelitian selanjutnya, *record* yang digunakan lebih banyak, sehingga memungkinkan untuk menemukan pola diluar perkiraan atau karakteristik kemungkinan keterlambatan pembayaran SPP. Diharapkan pada penelitian selanjutnya dapat dikembangkan dengan menggunakan jenis data serupa namun dengan metode *data mining* yang lain, seperti *clustering* yang dapat digunakan sebagai perbandingan.

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.uscom.2017.10.010>

