

**PENERAPAN ALGORITMA C5.0 DALAM MEMPREDIKSI
KETERLAMBATAN PEMBAYARAN BIAYA KULIAH
DI UNKRISWINA SUMBA**

Yanlychtar A. L. Mangngi¹, Alfrian C Talakua²
Universitas Kristen Wira Wacana Sumba

E-mail: yanmangngi040@gmail.com¹, alfriantalakua@unkriswina.ac.id²

Abstract

Wira Wacana Christian University (Unkriswina) Sumba is a private tertiary institution in East Sumba Regency, East Nusa Tenggara, which relies on tuition fee payments from its students for the continuity of its operations and development. In recent years, Unkriswina Sumba has faced increasing delays in paying tuition fees which has caused difficulties in financial management. This research aims to understand student behavior that causes late payments and develop a prediction system using the C5.0 algorithm to identify students who have the potential to be late in paying. The C5.0 algorithm was chosen because of its ability to handle complex data and is resistant to overfitting. Research data was obtained from forms distributed to students using indicators such as faculty, study program, class, parents' income and parents' education. The data collected was 16 samples. In this study, training data was 60% and testing data was 40%. From the classification results, an accuracy of 85.71% was obtained. It is hoped that the results of this research will help Unkriswina Sumba manage finances more effectively and prevent delays in paying tuition fees in the future.

Keywords — Data Mining, Decision Tree, Late Payment.

Abstrak

Universitas Kristen Wira Wacana (Unkriswina) Sumba merupakan perguruan tinggi swasta di Kabupaten Sumba Timur, Nusa Tenggara Timur, yang bergantung pada pembayaran biaya kuliah dari mahasiswanya untuk keberlangsungan operasional dan pengembangannya. Dalam beberapa tahun terakhir, Unkriswina Sumba menghadapi peningkatan keterlambatan pembayaran biaya kuliah yang menyebabkan kesulitan dalam pengelolaan keuangan. Penelitian ini bertujuan untuk memahami perilaku mahasiswa yang menyebabkan keterlambatan pembayaran dan mengembangkan sistem prediksi menggunakan algoritma C5.0 untuk mengidentifikasi mahasiswa yang berpotensi terlambat membayar. Algoritma C5.0 dipilih karena kemampuannya dalam menangani data kompleks dan tahan terhadap overfitting. Data penelitian diperoleh dari Form yang dibagikan pada mahasiswa dengan menggunakan indikator seperti fakultas, program studi, angkatan, penghasilan orang tua, serta pendidikan orang tua. Data yang terkumpul sebanyak 16 sampel. Pada penelitian ini data training 60% dan data testing 40%. Dari hasil klasifikasi diperoleh akurasi 85,71%. Hasil penelitian ini diharapkan dapat membantu Unkriswina Sumba dalam mengelola keuangan lebih efektif dan mencegah keterlambatan pembayaran biaya kuliah di masa depan

Kata Kunci — Data Mining, Decision Tree, Keterlambatan Pembayaran.

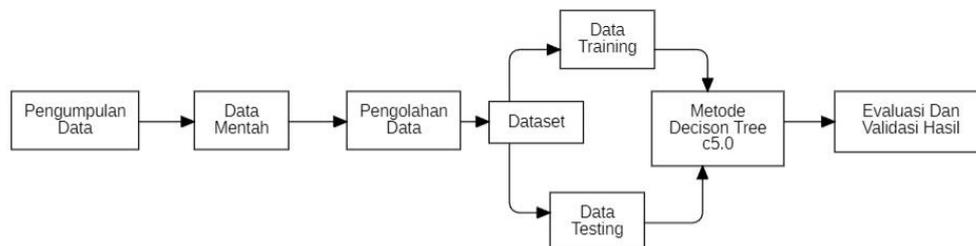
1. PENDAHULUAN

Universitas Kristen Wira Wacana Sumba merupakan salah satu perguruan tinggi swasta di Kabupaten Sumba Timur, Nusa Tenggara Timur. Sebagai perguruan tinggi swasta, Unkriswina Sumba bergantung pada pembayaran biaya kuliah dari mahasiswanya. Namun, dalam beberapa tahun terakhir Unkriswina Sumba mengalami peningkatan

keterlambatan pembayaran biaya kuliah. Hal ini menyebabkan kesulitan dalam pengelolaan keuangan yang bisa menghambat kemajuan universitas. Universitas Kristen Wira Wacana Sumba mengharuskan mahasiswa untuk membayar Sumbangan Pengembangan Pendidikan (SPP) berdasarkan ketetapan pembayaran yang telah ditentukan oleh Unkriswina Sumba. Pembayaran biaya kuliah tidak hanya merupakan dana tetap, namun mempunyai peranan yang sangat penting agar Unkriswina Sumba setiap tahunnya dapat berkembang. Tentu jika terjadi keterlambatan pembayaran biaya kuliah dapat merugikan pihak Universitas yang berpengaruh pada biaya operasional seperti penggajian hingga pembangunan yang tertunda. Keterlambatan pembayaran biaya kuliah dapat disebabkan oleh berbagai faktor, seperti kondisi ekonomi mahasiswa, dan kurangnya kesadaran akan pentingnya pembayaran tepat waktu. Oleh karena itu, penting untuk memahami perilaku mahasiswa yang menyebabkan keterlambatan pembayaran biaya kuliah dan mengembangkan sistem prediksi yang dapat membantu Unkriswina Sumba dalam mengidentifikasi mahasiswa yang berpotensi terlambat membayar. Algoritma C5.0 adalah versi yang lebih maju dari algoritma C4.5 karena meskipun proses dasarnya serupa, algoritma baru ini memiliki beberapa kelebihan dibandingkan yang lama. Hasil dari algoritma C5.0 adalah pohon keputusan atau aturan yang dibentuk berdasarkan nilai entropi atau gain [1]. Oleh karena itu, algoritma C5.0 menjadi pilihan yang tepat untuk diterapkan dalam memprediksi keterlambatan pembayaran biaya kuliah di Unkriswina Sumba. Tujuan penelitian ini adalah Membangun model prediksi keterlambatan pembayaran biaya kuliah di Unkriswina Sumba menggunakan algoritma C5.0 dan membangun model prediksi yang akurat dan dapat diandalkan untuk mengidentifikasi mahasiswa yang berpotensi terlambat membayar.

2. METODE

Berikut adalah kerangka pemikiran yang menggambarkan langkah-langkah yang akan diambil.



Gambar 1 Alur Penelitian

Tahap pertama adalah pengumpulan data, data dikumpulkan melalui form yang dibagikan pada mahasiswa unkriswina sumba. Data yang terkumpul berupa data nama, usia, jenis kelamin, alamat, semester, program studi, tahun masuk, IPK, jumlah sks yang diambil, pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan, sumber dana, pernah terlambat bayar atau tidak sebanyak 16 sampel. Tahap kedua adalah data mentah, pada tahap ini data yang terkumpul merupakan data yang belum diolah. Setelah Data mentah, data tersebut diteruskan ke pengolahan data. Pada tahap pengolahan data, data diolah menjadi dataset yang akan digunakan sebagai data training dan data testing. Pada tahap pengolahan data ditentukan atribut-atribut dan label. Data dibagi menjadi data training dan data testing, data training 60% dan data testing 40%. Pada tahap tersebut diterapkan metode decision tree. Pada tahap evaluasi dan validasi hasil, metrik evaluasi yang digunakan adalah confusion matriks, Confusion Matrix adalah metode yang digunakan untuk menghitung akurasi dalam konsep Data Mining. Evaluasi menggunakan

confusion matrix menghasilkan nilai akurasi, presisi, dan recall [2]. Confusion matrix adalah sebuah metode yang digunakan untuk menghitung akurasi dalam konsep data mining. Evaluasi menggunakan confusion matrix menghasilkan nilai akurasi, presisi, dan recall [3].

3. HASIL DAN PEMBAHASAN

Langkah pertama adalah pengumpulan data. Data dikumpulkan melalui formulir yang disebarakan kepada mahasiswa Unkriswina Sumba. Data yang terkumpul mencakup nama, usia, jenis kelamin, alamat, semester, program studi, tahun masuk, IPK, jumlah SKS yang diambil, pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan, sumber dana, serta riwayat keterlambatan pembayaran, dengan total 16 sampel. Pada tahap ini, dataset disiapkan dengan cermat. Setelah data dikumpulkan, dilakukan pengecekan ulang untuk memastikan keakuratannya. Langkah selanjutnya adalah mengolah data atau melakukan analisis klasifikasi dengan lebih efisien karena kesalahan data telah diminimalkan. Dataset yang disiapkan bisa mencakup berbagai jenis data, seperti data numerik, data kategorikal, dan data teks. Dengan dataset yang telah dipersiapkan dengan baik, proses klasifikasi akan menjadi lebih mudah dan efektif.

Tabel 1. Dataset

ID MAHASISWA	NAMA	JENIS KELAMIN	SEMESTER	PROGRAM STUDI	TAHUN MASUK	JUMLAH SKS YANG DIAMBIL	PEKERJAAN ORANG TUA	PENGHASILAN ORANG TUA	JUMLAH TANGGULANGAN ORANG TUA	SUMBER PENDANAAN PENDIDIKAN	PERNAH TERLAMBAT MEMBAYAR UANG KULIAH
MHS01	BRYAN	L	6	TIF	2021	23	Wiraswasta	2.500.000-4.000.000	5	Orang Tua	Ya
MHS02	IMANUEL	L	2	MAN	2023	18	Pensun PNS	2.500.000-4.000.000	5	Orang Tua	Tidak
MHS03	DELANO	L	8	TIF	2020	11	Petani	< 1.000.000	3	Orang Tua	Ya
MHS04	GLENFERIE	L	4	MAN	2022	20	Petani	1.000.000-2.500.000	5	Orang Tua	Tidak
MHS05	HILDEGARD	P	6	TIF	2021	19	Pegawai honorer	1.000.000-2.500.000	7	Orang Tua	Tidak
MHS06	FEBIYANTI	P	6	TIF	2021	19	Petani rumput laut	1.000.000-2.500.000	6	Beasiswa	Tidak
MHS07	DONI	L	2	AGB	2023	20	Petani	1.000.000-2.500.000	3	Orang Tua	Ya
MHS08	FEBIYANTI	P	6	TIF	2021	19	Petani rumput laut	1.000.000-2.500.000	6	Beasiswa	Tidak
MHS09	PIUS	L	6	TIF	2021	23	Guru	1.000.000-2.500.000	5	Orang Tua	Ya
MHS10	ADRIANUS	L	4	AGB	2022	20	PNS	1.000.000-2.500.000	1	Orang Tua	Ya
MHS11	ABNER	L	6	TIF	2021	21	Ibu rumah tangga	< 1.000.000	2	Beasiswa	Tidak
MHS12	JOYIS	L	6	MAN	2021	20	Tidak ada	1.000.000-2.500.000	6	Orang Tua	Ya
MHS13	MELKIANUS	L	2	AGB	2023	18	Petani	< 1.000.000	4	Orang Tua	Ya
MHS14	NONITA	P	6	MAN	2021	20	Usaha parang	< 1.000.000	4	Orang Tua	Tidak
MHS15	KRISTIAN	L	6	TIF	2021	20	Petani	1.000.000-2.500.000	5	Orang Tua	Ya
MHS16	IRVAN	L	10	AGB	2019	6	Petani	500.000-1.000.000	2	Orang Tua	Ya

Dari dataset diatas ada atribut ID MAHASISWA, NAMA, JENIS KELAMIN, SEMESTER, PROGRAM STUDI, TAHUN MASUK, JUMLAH SKS YANG DIAMBIL, PEKERJAAN ORANG TUA, PENGHASILAN ORANG TUA, JUMLAH TANGGULANGAN ORANG TUA, SUMBER PENDANAAN PENDIDIKAN DAN PERNAH TERLAMBAT MEMBAYAR UANG KULIAH. Yang menjadi label adalah atribut “PERNAH TERLAMBAT MEMBAYAR UANG KULIAH”.

Data terdiri dari 16 sampel, dengan 60% atau 9 sampel digunakan sebagai data latih, dan 40% atau 7 sampel digunakan sebagai data uji. Pembagian data dengan rasio 60:40, 80:20, atau 90:10 merupakan praktik umum dalam pembelajaran mesin untuk memisahkan dataset menjadi dua bagian, yaitu data latih dan data uji.

Tabel 2. Hasil Prediksi

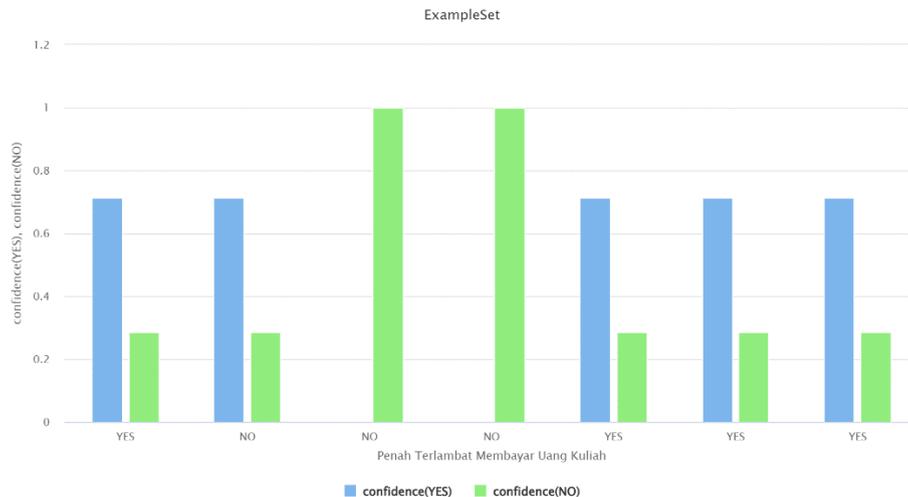
Penah Terlambat Membayar Uang Kuliah	prediction(Penah Terlambat Membayar Uang Kuliah)	confidence(YES)	confidence(NO)
YES	YES	0.714	0.286
NO	YES	0.714	0.286
NO	NO	0	1
NO	NO	0	1
YES	YES	0.714	0.286
YES	YES	0.714	0.286
YES	YES	0.714	0.286

Berdasarkan klasifikasi yang ditunjukkan pada tabel 2, kelas YES memiliki 4 sampel yang diprediksi sebagai YES juga sebanyak 4 sampel dan tidak ada sampel yang

diprediksi sebagai NO. Sementara itu, kelas NO yang terdiri dari 3 sampel, diprediksi sebagai NO sebanyak 2 sampel dan sebagai YES sebanyak 1 sampel. Oleh karena itu, tingkat akurasi prediksi dapat dilihat pada tabel di bawah ini.

accuracy: 85.71%

	true YES	true NO	class precision
pred. YES	4	1	80.00%
pred. NO	0	2	100.00%
class recall	100.00%	66.67%	



Gambar 2. Histogram confidence kelas YES dan kelas NO

Dari histogram di atas dijelaskan kemungkinan kelas YES diprediksi menjadi kelas NO atau kelas YES dan kelas NO kemungkinan diprediksi menjadi kelas YES atau kelas NO.

KESIMPULAN

Universitas Kristen Wira Wacana Sumba (Unkriswina Sumba) menghadapi tantangan dalam mengelola keuangan akibat meningkatnya keterlambatan pembayaran biaya kuliah oleh mahasiswa. Penelitian ini bertujuan untuk mengembangkan model prediksi keterlambatan pembayaran biaya kuliah menggunakan algoritma C5.0, yang terbukti efektif dalam menghasilkan pohon keputusan berdasarkan nilai entropi atau gain. Proses penelitian melibatkan beberapa tahap:

1. Pengumpulan Data: Data dikumpulkan melalui formulir yang disebarakan kepada mahasiswa Unkriswina Sumba. Data yang terkumpul mencakup nama, usia, jenis kelamin, alamat, semester, program studi, tahun masuk, IPK, jumlah SKS yang diambil, pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan, sumber dana, serta riwayat keterlambatan pembayaran, dengan total 16 sampel.
2. Pengolahan Data: Data dibagi menjadi data latih dan data uji dengan rasio 60:40, menghasilkan 9 sampel sebagai data latih dan 7 sampel sebagai data uji.
3. Penerapan Algoritma C5.0: Algoritma C5.0 diterapkan pada data latih untuk membentuk model prediksi keterlambatan pembayaran.
4. Evaluasi Model: Model dievaluasi menggunakan confusion matrix. Hasil klasifikasi menunjukkan bahwa:
 - Kelas YES memiliki 4 sampel yang diprediksi dengan benar sebagai YES (akurasi 100% untuk kelas YES).

- Kelas NO memiliki 3 sampel, di mana 2 sampel diprediksi dengan benar sebagai NO dan 1 sampel diprediksi salah sebagai YES (akurasi 66.67% untuk kelas NO).

5. Tingkat Akurasi Prediksi

- Akurasi keseluruhan: (4 sampel YES yang benar + 2 sampel NO yang benar) / 7 total sampel = $6/7 = 85.71\%$
- Presisi kelas YES: $4/4 = 100\%$
- Presisi kelas NO: $2/3 = 66.67\%$
- Recall kelas YES: $4/4 = 100\%$
- Recall kelas NO: $2/3 = 66.67\%$

Hasil penelitian menunjukkan bahwa model prediksi berbasis algoritma C5.0 dapat menjadi alat yang berguna bagi Unkriswina Sumba dalam mengidentifikasi mahasiswa yang berpotensi terlambat membayar biaya kuliah. Hal ini dapat membantu universitas dalam mengelola keterlambatan pembayaran dan meningkatkan efisiensi pengelolaan keuangan universitas.

REFERENCES

- [1] Ginting, V. S., Kusrini, K., & Taufiq, E. (2020). Implementasi algoritma c4. 5 untuk memprediksi keterlambatan pembayaran sumbangan pembangunan pendidikan sekolah menggunakan python. *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, 10(1), 36-44.
- [2] Muqorobin, M., Kusrini, K., & Luthfi, E. T. (2019). Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah. *Jurnal Ilmiah SINUS*, 17(1), 1-14.
- [3] Talakua, A. C. (2023). Penerapan Data Mining untuk memprediksi Nilai Akhir Matakuliah Menggunakan Metode K-Means Clustering. *JURNAL TEKNIK INFORMATIKA INOVATIF WIRA WACANA*, 2(3), 178-186.