

PENERAPAN ALGORITMA C4.5 DALAM MEMPREDIKSI KETERLAMBATAN PEMBAYARAN PEMBAYARAN SUMBANGAN PEMBINAAN PENDIDIKAN

Mutimatul Mubarokah¹⁾, Danang Aditya Nugraha²⁾, Amak Yunus³⁾

Universitas PGRI Kanjuruhan Malang

mutimatulmubarokah@gmail.com

Abstrak

Keberlangsungan kegiatan operasional dalam suatu sekolah swasta yang sangat memiliki peran penting adalah sumber dana. Salah satu yang menjadi sumber dana dalam suatu sekolah swasta adalah pembayaran SPP. Oleh karena itu sangatlah penting suatu sekolah mengelola keuangan sekolah dengan baik. Namun akan menjadi masalah yang cukup besar bagi lembaga sekolah apabila terjadi keterlambatan pembayaran SPP oleh murid. Metode yang diusulkan pada penelitian ini adalah eksperimen dan pengujian Dalam perhitungan algoritma C4.5 proses pembagian dataset dibagi menjadi 2 subset yaitu data training dan data testing. Penerapan Algoritma C4.5 dalam penelitian ini diterapkan untuk membentuk pohon keputusan dan untuk mengklasifikasi pembayaran SPP dengan melakukan pengujian akurasi menggunakan Confusion Matrix. Pada klasifikasi algoritma C4.5 data akan dibagi menjadi data training dan testing dengan masing-masing presentasi berbeda yaitu 60% dan 40%, 70% dan 30%, 80% dan 20%. Hasil akurasi yang didapatkan dari uji coba data training dan testing sebanyak 60% dan 40% menghasilkan akurasi sebesar 72,5 %. Hasil akurasi yang didapatkan dari uji coba data training sebanyak 70% dan testing 30% menghasilkan akurasi sebesar 76,6%. Sedangkan hasil akurasi yang didapatkan dari training 80% dan testing 20% menghasilkan akurasi sebesar 90 %. Dapat dilihat pada Berdasarkan hasil dari penelitian ini dapat disimpulkan bahwa: Hasil dari klasifikasi menggunakan metode Algoritma C4.5 dapat disimpulkan bahwa metode algoritma C4.5 cocok untuk pengklasifikasian data Pembayaran SPP dengan hasil akurasi sebesar 90 %.

Kata Kunci : Prediksi; Algoritma C4.5; python dan Confusion Matrix

Abstract

The continuity of operational activities in a private school which plays a very important role is the source of funds. One of the sources of funds in a private school is tuition payments. Therefore, it is very important for a school to manage school finances well. However, it will be quite a big problem for school institutions if there is a delay in paying tuition fees by students. The method proposed in this research is experimentation and testing. In calculating the C4.5 algorithm, the process of dividing the dataset is divided into 2 subsets, namely training data and testing data. Training data for model formation and testing data for model testing. In this scenario the aim is to test the classification model in different dataset conditions. Application of the C4.5 Algorithm The C4.5 algorithm in this research is applied to form a decision tree and to classify SPP payments by testing accuracy using the Confusion Matrix. In the C4.5 algorithm classification, data will be divided into training and testing data with each different presentation, namely 60% and 40%, 70% and 30%, 80% and 20%. The accuracy results obtained from the training and testing data trials were 60% and 40% resulting in an accuracy of 72.5%.

The accuracy results obtained from training data trials were 70% and 30% testing resulted in an accuracy of 76.6%. Meanwhile, the accuracy results obtained from 80% training and 20% testing resulted in an accuracy of 90%. It can be seen in Based on the results of this research, it can be concluded that: The results of the classification using the C4.5 Algorithm method can be concluded that the C4.5 algorithm method is suitable for classifying SPP Payment data with an accuracy result of 90%.

Keywords : Prediction; Algorithma C4.5; python and confusion matrix

1. PENDAHULUAN

Keberlangsungan kegiatan operasional dalam suatu sekolah swasta yang sangat memiliki peran penting adalah sumber dana. Salah satu yang menjadi sumber dana dalam suatu sekolah swasta adalah pembayaran SPP. Oleh karena itu sangatlah penting suatu sekolah mengelola keuangan sekolah dengan baik. Namun akan menjadi masalah yang cukup besar bagi lembaga sekolah apabila terjadi keterlambatan pembayaran SPP oleh murid. Salah satunya adalah di MTs Nurul Huda Berdasarkan masalah diatas perlu dilakukan penelitian untuk memprediksi keterlambatan pembayaran SPP yang dilakukan oleh murid. Namun dalam pengelolaan keuangan masih terbilang minim teknologi di karenakan kurangnya menggali informasi manajemen keuangan .Dan juga bisa dari faktor penghasilan keuangan orang tua siswa.

Berdasarkan dari masalah diatas perlu dilakukan sebuah penelitian untuk mengetahui penyebab apa yang membuat siswa terlambat dalam pembayaran SPP. Salah satu metode yang digunakan adalah metode algoritma C 4.5 yang merupakan metode yang ada didalam data mining.

Algoritma yang dipakai oleh peneliti adalah analisis algoritma C4.5 karena dalam penelitian sebelumnya yang dilakukan oleh Anisa Fitri Sari (2020) dalam memprediksi keterlambatan pembayaran SPP yang paling tepat adalah menggunakan algoritma C4.5. algoritma ini tingkat akurasinya lebih tinggi dibanding dengan menggunakan metode Naive Bayes Clasifier dengan tingkat akurasi algoritma C4.5 sebesar 96,5% sedangkan tingkat akurasi Naive Bayes Clasifier sebesar 94%.

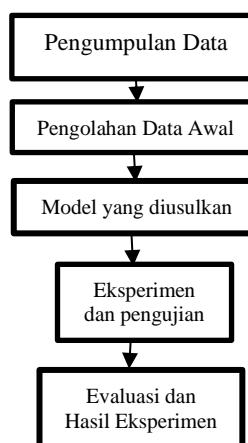
2. METODE / ALGORITMA

Dalam Penelitian ini langkah langkah yang diambil peneliti adalah di mulai dengan mengumpulkan data sampai pengujian Evaluasi dan hasil. Metode yang digunakan adalah Kuantitatif dan data yang digunakan adalah data primer.

Pada tahap pengumpulan data peneliti mengambil data peserta didik dari Staff TU dan diolah sehingga mendapatkan tingkat akurasi.

Dalam tahap ini juga peneliti akan membuat diagram alir. Tahap penelitian akan dimulai dengan pengumpulan data, pengolahan data awal, model yang diusulkan menggunakan algoritma C4.5, eksperimen dan pengujian serta evaluasi dan hasil eksperimen menggunakan confusion matrix.

Berikut struktur tahapan penelitian yang akan digunakan :



Gambar 1 Alur Penelitian**3. HASIL DAN PEMBAHASAN****Tabel 1 Perhitungan Nilai Gain dan Entropy**

Node Akar		Jumlah	Tepat	Terlambat	Entropy	Gain
Total		100	49	51	0,999711	
Penghasilan	500000	57	19	38	0,918296	0,660538
	1000000	28	18	10	0,940286	0,365881
	1500000	11	10	1	0,439497	0,028862
	2000000	3	1	2	0,918296	0,056411
						0,056411
Tanggungan	Sedikit	56	30	26	0,996317	0,614348
	Cukup	39	17	22	0,988111	0,294698
	Banyak	5	2	3	0	0,294698
Pekerjaan ayah						0,130536
	Wiraswasta	71	39	32	0,992977	0,889516
	Guru	1	1	0	0	0,096509
	Buruh Harian	16	1	15	0,33729	0,04028
	Lainnya	12	8	4	0,918296	0,902616
Pekerjaan Ibu						0,05338
	Ibu Rumah Tangga	85	41	44	0,999101	0,999711
	Buruh	3	0	3	0	-0,02669
	Wiraswasta	10	6	4	0,970951	0,01041
	Guru	2	2	0	0	0,939711
Pendidikan Ayah						0,07041
	SD	93	45	48	0,999249	0,939711
	SMP	6	3	3	1	0,249808
	SMA	0	0	0	0	0,868071
	S1	1	1	0	0	0,890368
Usia Ayah						0,249808
	Dewasa	75	38	37	0,999872	-0,05067
	Lansia	14	5	9	0,940286	0,830368
	Manula	11	6	5	0,99403	0,999711
Pendidikan Ibu						0,080966
	SD	92	44	48	0,998636	0,939711
	SMP	6	3	3	1	0,240091
	SMA	0	0	0	0	0,868071
	S1	2	2	0	0	0,899711
Umur Ibu						0,108451
	Dewasa	76	39	37	0,9995	0,768071
	Lansia	14	5	9	0,940286	0,899711
	Manula	10	5	5	1	0,762236
P1						0,563604
	Ya	52	5	47	0,456684	0,801079
	Tidak	48	44	4	0,413817	0,729668
P2						0,532305
	Ya	53	6	47	0,509516	0,802349
	Tidak	47	43	4	0,419921	0,668375

P3	Ya	56	8	48	0,591673	0,841707
	Tidak	44	41	3	0,359102	0,999711
P4	Ya	51	0	51	0	0,999711
	Tidak	49	49	0	0	0,999711

Hasil Confusion Matrix

Dari beberapa 3 percobaan yang peneliti gunakan hasil tertinggi terdapat pada data training 80% dan data testing 20 %.

Tabel 2 Hasil Perhitungan Confussion Matrix Menggunakan Training 80%

	True TEPAT (False)	True TERLAMBAT(True)
Pred Tepat	7	0
Prediksi Terlambat	2	11

Keterangan :

Hasil yang didapat dari pengujian data training 80 % dan data testing 20%

Akurasi = 90%

Precision = 100%

Recall = 78 %

Penerapan Algoritma C4.5

Algoritma C4.5 dalam penelitian ini diterapkan untuk membentuk pohon keputusan dan untuk mengklasifikasi pembayaran SPP dengan melakukan pengujian akurasi menggunakan Confusion Matrix. Pada klasifikasi algoritma C4.5 data akan dibagi menjadi data training dan testing dengan masing-masing presentasi berbeda yaitu 60% dan 40%, 70% dan 30%, 80% dan 20%.

Tabel 3 Hasil Akurasi Algoritma C4.5

Metode	Akurasi		
	Training 60 %	Training 70 %	Training 80%
Algoritma C4.5	72,5 %	76,6 %	90 %

Algoritma C4.5



Gambar 2 Grafik Perbandingan Akurasi

4. KESIMPULAN

Berdasarkan hasil dari penelitian ini dapat disimpulkan bahwa:

1. Hasil dari klasifikasi menggunakan metode Algoritma C4.5 dapat disimpulkan bahwa metode algoritma C4.5 cocok untuk pengklasifikasian data Pembayaran SPP dengan hasil akurasi sebesar 90 %.
2. Pengklasifikasian secara kombinasi menggunakan algoritma C4.5 bahwa belum tentu jumlah data training semakin tinggi belum tentu menghasilkan akurasi lebih tinggi.

Adapun saran untuk penelitian selanjutnya guna mengembangkan metode algoritma C4.5 agar lebih baik dalam mengklasifikasi.

1. Jumlah data yang digunakan lebih banyak agar mendapatkan hasil akurasi yang baik, karena semakin banyak data yang dibutuhkan maka hasil akurasi yang didapatkan juga akan lebih baik.
2. Penelitian selanjutnya pengklasifikasian seperti pada penelitian ini dapat dilakukan oleh penelitian selanjutnya menggunakan software lainnya seperti Rapid Miner.
3. Penelitian selanjutnya bisa menggunakan Metode KNN, Metode CNN, Metode Naive Bayes untuk mendapatkan tingkat akurasi yang lebih tinggi.
4. Mencoba melakukan penelitian yang lebih mendalam pada atribut yang akan digunakan sehingga mengoptimalkan tingkat akurasi yang didapatkan

5. REFERENSI

- [1] Putri & Waspada – 2018 “Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika” Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika
- [2] Muqorobin et al. 2019 “Optimasi Metode Naïve Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan Sekolah”. Jurnal Ilmiah Sinus (JIS),17(01).
- [3] Ginting et al. 2020 “Penerapan Algoritma C4.5 Dalam Memprediksi Keterlambatan Pembayaran Uang Sekolah Menggunakan Python” Inspiration: Jurnal Teknologi Informasi dan Komunikasi. 4(01).
- [4] Apandi et al. 2019 “Menganalisis Kemungkinan Keterlambatan Pembayaran Spp Dengan Algoritma C4.5 (Studi Kasus Politeknik Tedc Bandung). Jurnal Techno Nusa 16(2) : 93
- [5] Mardi .2017. “Data Mining : Klasifikasi Menggunakan Algoritma C4.5” Edik Informatika.
- [6] Santosa, B., 2007. Data Mining Terapan Surabaya : Graha Ilmu
- [7] Ginting et al. 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.
- [8] Choirul Anam,Heri Budi Santoso.,2018 “Perbandingan Kinerja Algoritma C4.5 dan Naive Bayes untuk Klasifikasi Penerima Beasiswa”.Vol.8 No.1
- [9] Asmaul Husnah Nasrulloh, 2018 ” Penerapan Metode C4.5 Untuk Klasifikasi Mahasiswa Berpotensi Drop Out”. Jurnal Ilmiah Vol 10 No.2
- [10] Norzam Yahya, Arief Jananto.,2019 ” Komparasi Kinerja Algoritma C.45 Dan Naive Bayes Untuk Prediksi Kegiatan Penerimaanmahasiswa Baru (Studi Kasus : Universitas Stikubank Semarang”