

OPTIMASI CHI SQUARE DAN PERBAIKAN TEKNIK PRUNNING UNTUK PENINGKATAN AKURASI ALGORITMA C4.5 DALAM MODEL KASUS PREDIKSI KETERLAMBATAN BIAYA KULIAH

Anton Saputra¹, Taghfirul Azhima Yoga Siswa²

^{1,2}Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur

Email: saputraanton490@gmail.com¹, tay758@umkt.ac.id²

Abstrak

Universitas Muhammadiyah Kalimantan Timur menerapkan biaya kuliah yang mempunyai peranan penting dalam meningkatkan mutu pendidikan dan pembangunan infrastruktur, namun masih banyak mahasiswa yang terlambat dalam melakukan pembayaran biaya kuliah yang dapat berpengaruh pada biaya operasional. Penelitian ini bertujuan menentukan fitur, implementasi algoritma C4.5, dan mengevaluasi kinerja algoritma C4.5 dengan membagi 90% data training dan 10% data testing menggunakan confusion matrix. Pada penelitian ini dilakukan penerapan seleksi fitur menggunakan metode chi square dan teknik pruning dalam meningkatkan akurasi algoritma C4.5. Data yang digunakan berasal dari bagian Biro Administrasi Keuangan dan Biro Administrasi Akademik dengan jumlah 12.408. Pengujian algoritma C4.5 dilakukan tanpa menggunakan seleksi fitur chi square dan teknik pruning memperoleh nilai akurasi sebesar 61,40%. Sedangkan pengujian algoritma C4.5 dengan seleksi fitur chi square dan teknik pruning memperoleh nilai akurasi sebesar 65,53%. Dengan demikian adanya seleksi fitur chi square dan teknik pruning dapat meningkatkan akurasi algoritma C4.5 dalam prediksi keterlambatan pembayaran biaya kuliah.

Kata Kunci: Algoritma C4.5, Biaya Kuliah, Prediksi, Seleksi Fitur

Abstract

Muhammadiyah University of East Kalimantan applies tuition fees which have an important role in improving the quality of education and infrastructure development, but there are still many students who are late in paying tuition fees which can affect operational costs. This study aims to determine the features, implement the C4.5 algorithm, and evaluate the performance of the C4.5 algorithm by dividing 90% of the training data and 10% of the testing data using a confusion matrix. In this study, the application of feature selection using the chi square method and pruning technique was carried out in increasing the accuracy of the C4.5 algorithm. The data used comes from the Financial Administration Bureau and the Academic Administration Bureau with a total of 12,408. Testing the C4.5 algorithm was carried out without using the chi square feature selection and pruning technique to obtain an accuracy value of 61.40%. While testing the C4.5 algorithm with chi square feature selection and pruning technique obtained an accuracy value of 65.53%. Thus the chi square feature selection and pruning technique can increase the accuracy of the C4.5 algorithm in predicting late payment of tuition fees.

KeyWords : C4.5 Algorithm, Feature Selection, Prediction, Tuition Fee

I. PENDAHULUAN

Universitas Muhammadiyah Kalimantan Timur (UMKT) merupakan salah satu perguruan swasta yang mengharuskan mahasiswa untuk membayar Sumbangan Pembangunan Pendidikan (SPP) berdasarkan ketentuan pembayaran yang telah ditentukan. Biaya kuliah merupakan sumber dana tetap yang mempunyai peranan penting untuk meningkatkan mutu pendidikan dan pembangunan infrastruktur. Tentu jika terjadi keterlambatan pembayaran biaya kuliah dapat merugikan pihak UMKT yang berpengaruh pada biaya operasional seperti penggajian hingga pembangunan infrastruktur yang akan tertunda. Berdasarkan hal tersebut dibutuhkan sebuah data analisis yang dapat digunakan untuk prediksi keterlambatan pembayaran biaya kuliah, karena dengan adanya prediksi diharapkan dapat membantu pihak UMKT khususnya pada bagian Biro Administrasi Keuangan untuk dapat manajemen keuangan dengan lebih baik lagi dan sebagai rekomendasi kepada pihak UMKT dalam melakukan evaluasi agar dapat meminimalisir mahasiswa yang terlambat dalam melakukan pembayaran biaya kuliah.

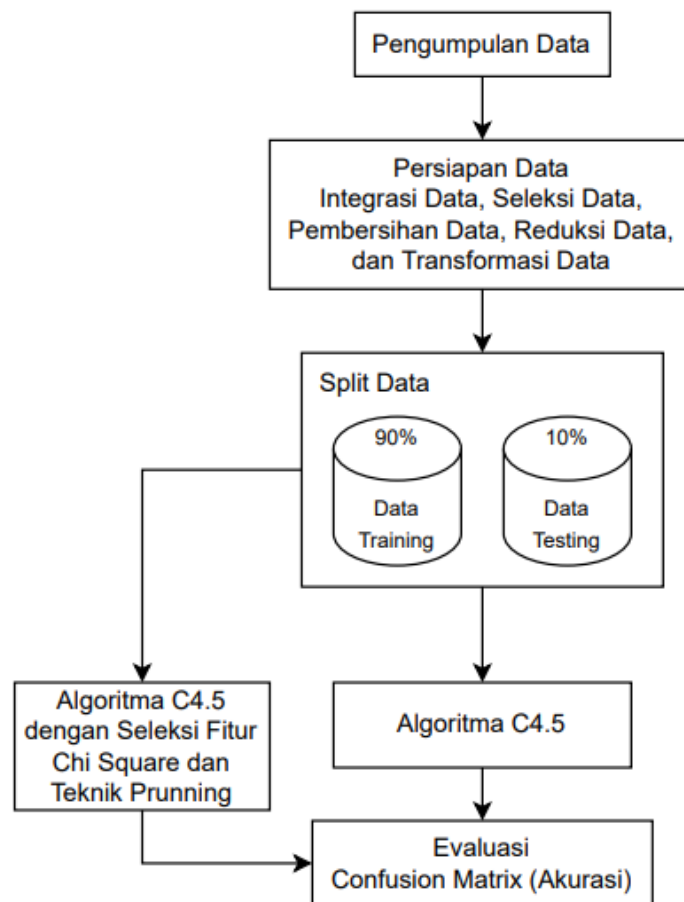
Penelitian sebelumnya telah dilakukan mengenai prediksi keterlambatan pembayaran sumbangan pendidikan sekolah menggunakan algoritma C4.5. Hasil dari penelitian menunjukkan bahwa dengan variabel penghasilan orangtua, tanggungan keluarga, pendidikan ayah, pendidikan ibu, umur ibu, dan keterangan memperoleh nilai akurasi sebesar 73% [1]. Penelitian lain melakukan analisis keterlambatan pembayaran sumbangan pembangunan pendidikan menggunakan algoritma C4.5. Pada penelitian ini dilakukan pengujian data yang berasal dari kuisioner mahasiswa program studi Teknik Informatika semester 3, 5 dan 7 kelas reguler ajaran 2015/2016 memperoleh nilai akurasi paling tinggi sebesar 75% [2]. Penelitian Desiani,dkk melakukan perbandingan metode klasifikasi dari algoritma C4.5 dengan algoritma *Naïve Bayes*. Hasil dari penelitian ini menunjukkan bahwa kinerja algoritma C4.5 lebih baik dibandingkan kinerja algoritma *Naïve Bayes* [3]. Penelitian Karyadiputra telah membandingkan atribut yang digunakan dengan seleksi fitur *chi square*. Hasil yang diperoleh bahwa dengan adanya seleksi fitur *chi square* mampu meningkatkan akurasi dengan menghapus atribut yang tidak relevan [4]. Selanjutnya untuk algoritma C4.5 dilakukan

dengan teknik *prunning* menghasilkan pohon keputusan menjadi lebih kecil dan hasil dari teknik *prunning* mampu meningkatkan akurasi [5].

Berdasarkan beberapa penelitian yang telah dipaparkan terdapat perbedaan dalam penelitian ini menggunakan algoritma C4.5 dengan seleksi fitur *chi square* dan teknik *prunning* untuk meningkatkan akurasi. Fitur yang tidak relevan sebaiknya dihilangkan karena tidak semua fitur memiliki hubungan yang jelas dengan atribut label atau target, maka hasil dari seleksi fitur dapat mencegah terjadinya *overfitting* yang dapat meningkatkan akurasi dan data yang digunakan menjadi lebih sedikit sehingga waktu pelatihan menjadi lebih cepat [6]. Teknik *prunning* digunakan dalam *decision tree* karena memiliki cabang yang terlalu panjang dapat mengakibatkan *overfitting*, maka dilakukan pemangkasan dan *node-node* di tingkat atas dijadikan sebagai *leaf node*, sehingga waktu pelatihan menjadi lebih cepat dan hasil yang diberikan menjadi lebih akurat [7]. Dari hal tersebut seleksi fitur dapat dilakukan salah satunya menggunakan *chi square* untuk meningkatkan akurasi algoritma C4.5 yang telah dibuktikan pada penelitian [4]. Selanjutnya penggunaan teknik *prunning* mampu meningkatkan akurasi algoritma C4.5 telah terbukti dalam penelitian [5]. Sehingga pada penelitian ini untuk peningkatan akurasi algoritma C4.5 dilakukan kombinasi seleksi fitur *chi square* dan teknik *prunning* untuk menggunakan atribut yang lebih sedikit, namun mempunyai hubungan yang jelas terhadap atribut label atau target dan membuat pohon keputusan menjadi lebih kecil.

II. METODE PENELITIAN

Pada penelitian ini berfokus pada prediksi keterlambatan pembayaran biaya kuliah dengan peneparan algoritma C4.5, kemudian dilakukan peningkatan akurasi menggunakan seleksi fitur *chi square* dan teknik *prunning*. Adapun tahapan penelitian ini dapat dilihat pada Gambar 1.



Gambar 1: Tahapan Penelitian

A. Pengumpulan Data

Pengumpulan data dilakukan melalui observasi pada bagian Biro Administrasi Keuangan, adapun data yang diperoleh berjumlah 39.644 data mahasiswa yang membayar biaya kuliah di Universitas Kalimantan Timur tahun 2019 s/d 2021 yang terdapat atribut nim, nama, dan keterangan. Kemudian untuk melengkapi data mahasiswa yang membayar biaya kuliah, maka data juga didapatkan pada bagian Biro Administrasi Akademik berjumlah 10.959 data diri mahasiswa dari angkatan 2017 s/d 2021 terdapat atribut nim, nama, fakultas, prodi, angkatan, gender, penghasilan ayah, penghasilan ibu, pendidikan ayah, dan pendidikan ibu.

B. Persiapan Data

Persiapan data dilakukan untuk memperbaiki masalah yang terdapat pada dataset dengan melakukan integrasi data, seleksi data, pembersihan data, reduksi data, dan transformasi data.

- 1) Integrasi Data Integrasi data dilakukan dengan menggabungkan data yang bersumber dari bagian Biro Administrasi Keuangan menggunakan atribut nim untuk mengambil data pada bagian Biro Administrasi Akademik sehingga menghasilkan sebuah data yang baru. Adapun hasil dari integrasi data memperoleh 11 atribut yang dapat dilihat pada Tabel I.

Tabel I: Itegrasi Data

No	Atribut	Jenis Data
1	Nim	Karakter
2	Nama	Karakter
3	Fakultas	Kategorikal
4	Prodi	Kategorikal
5	Angkatan	Kategorikal
6	Gender	Kategorikal
7	Penghasilan Ayah	Kategorikal
8	Penghasilan Ibu	Kategorikal
9	Pendidikan Ayah	Kategorikal
10	Pendidikan Ibu	Kategorikal
11	Keterangan	Kategorikal

- 2) Seleksi Data Seleksi data dilakukan dengan membuang atribut yang tidak di perlukan seperti nim dan nama karena atribut ini hanya digunakan pada saat melakukan proses penggabungan data.
- 3) Pembersihan Data Pembersihan data dilakukan pada dataset dengan menghapus data yang tidak lengkap dan data yang error karena sebelum dianalisis harus bersih agar dapat memberikan hasil yang akurat.
- 4) Reduksi Data Reduksi data dilakukan pada dataset karena jumlah kelas target yang dimiliki tidak seimbang, maka perlu melakukan proses balancing (penyeimbangan) dengan jumlah kelas target yang paling banyak akan dibuang sebagian sehingga dataset menjadi lebih kecil dengan tetap menjaga integritas data asli.
- 5) Transformasi Data Transformasi data perlu dilakukan pada algoritma *machine learning* karena membutuhkan data numerik untuk digunakan sebagai data *training* [7]. Pada tahap ini mengubah data yang bertipe kategorikal ditransformasi menjadi data yang bertipe numerik, kecuali atribut keterangan karena digunakan sebagai label atau target, sehingga hasil dari transformasi data dapat berjalan dengan baik pada saat memasuki tahap permodelan. Adapun dari transformasi data dapat dilihat pada Tabel II.

Tabel II: Transformasi Data

No.	Attribut	Keterangan Nilai
1	Fakultas	0. Ekonomi Bisnis dan Politik
		1. Farmasi
		2. Hukum
		3. Ilmu Keperawatan
		4. Keguruan dan Ilmu Pendidikan
		5. Kesehatan Masyarakat
		6. Psikologi
2	Prodi	7. Sain dan Teknologi
		0. Farmasi
		1. Hubungan Internasional
		2. Hukum
		3. Keperawatan
		4. Kesehatan Masyarakat
		5. Manajemen
3	Angkatan	6. Ners
		7. Pendidikan Bahasa Inggris
		8. Pendidikan Olah Raga
		9. Psikologi
		10. Teknik Informatika
		11. Teknik Mesin
		12. Teknik Sipil
		0. 2017

		1. 2018
		2. 2019
		3. 2020
4	Gender	0. L 1. P
5	Penghasilan Ayah	0. Kurang dari Rp.500,000 1. Rp. 500,000 – Rp. 999,999 2. Rp. 1,000,000 – Rp. 1.999,999 3. Rp. 2,000,000 – Rp. 4,999,999 4. Rp. 5,000,000 – Rp. 20,000,000 5. Lebih dari Rp. 20,000,000
6	Penghasilan Ibu	0. Kurang dari Rp.500,000 1. Rp. 500,000 – Rp. 999,999 2. Rp. 1,000,000 – Rp. 1.999,999 3. Rp. 2,000,000 – Rp. 4,999,999 4. Rp. 5,000,000 – Rp. 20,000,000 5. Lebih dari Rp. 20,000,000

C. Split Data

Data yang digunakan setelah melakukan proses persiapan data berjumlah 12.408 data yang akan dibagi menjadi 2 bagian yaitu 90% data training berjumlah 11.168 data untuk membentuk model algoritma C4.5, kemudian 10% data testing berjumlah 1.241 data untuk mengevaluasi hasil kinerja dari model algoritma C4.5 yang telah terbentuk sebelumnya.

D. Chi Square

Chi square memiliki ukuran statistik yang dapat digunakan untuk menentukan fitur yang lebih baik menjadi dependen maupun independen, selain itu *chi square* menggunakan frekuensi sebagai dasar perhitungan di berbagai statistik untuk melakukan uji hipotesis dalam beberapa hal [8]. *Chi square* termasuk salah satu seleksi fitur *supervised* yang digunakan untuk menghapus fitur, namun tidak mengurangi hasil akurasi yang didapatkan [9]. Adapun untuk perhitungan *chi square* menggunakan persamaan (1) sebagai berikut.

$$x^2(t, c) = \frac{N(AD - CB)^2}{(A + C)(B + D)(A + B)(C + D)} \quad (1)$$

Keterangan :

$x^2(t, c)$: Nilai *chi-square* term t disetiap kategori c.

t : Kata

c : Kategori

N : Jumlah dokumen latih

A : Total dokumen pada kategori c yang memiliki term t

B : Total dokumen di kategori selain c yang memiliki term t

C : Total dokumen pada kategori c yang tidak memiliki term t

D : Total dokumen dikategori selain c yang tidak memiliki term t

E. Algoritma C4.5

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan berdasarkan kriteria pembentuk keputusan [10]. Algoritma C4.5 mirip seperti pohon dengan memiliki node internal (bukan daun) untuk deskripsi atribut pada tiap-tiap cabang menggambarkan hasil atribut yang diuji dan tiap daun menggambarkan kelas [11]. Algoritma C4.5 dalam membangun pohon keputusan dilakukan tahap perhitungan sebagai berikut [12] : noitemsep

- 1) Mempersiapkan data training dengan mengambil data history yang pernah terjadi sebelumnya.
- 2) Menghitung nilai entropy menggunakan persamaan (2).

$$Entropy(i) = - \sum_{i=1}^p \frac{n_i}{n} \log_2 \left(\frac{n_i}{n} \right) \quad (2)$$

Keterangan :

p : Jumlah partisi atribut

n_i : Proporsi n_i terhadap n

n : Jumlah kasus dalam n

- 3) Menentukan akar dari pohon dengan menghitung nilai gain yang tertinggi pada masing-masing atribut menggunakan persamaan (3).

$$Gain = - \sum_{j=1}^m f(i, j) \cdot 2f [(i \quad j)] \quad (3)$$

Keterangan :

i : Himpunan kasus

m : Jumlah partisi i

$f(i, j)$: Proporsi j terhadap i

- 4) Menghitung split information menggunakan persamaan (4)

$$Split Information = - \sum_{t=1}^c \frac{s1}{s} \log_2 \frac{s1}{s} \quad (4)$$

Keterangan :

S_1 sampai S^2 : c subset yang dihasilkan dari pemecahan S dengan menggunakan atribut A yang mempunyai sebanyak nilai c .

- 5) Menghitung gain ratio menggunakan persamaan (5).

$$Gainratio(S, A) = \frac{Gain(S, A)}{Splitinformation(S, A)} \quad (5)$$

- 6) Ulangi langkah ke-2 hingga semua record terpartisi.

- 7) Proses partisi pohon keputusan akan berhenti saat:

- a) Semua tupel dalam record dalam simpul m mendapatkan kelas yang sama.
- b) Tidak ada atribut dalam record yang dipartisi lagi.
- c) Tidak ada record dalam cabang yang kosong.

F. Confusion Matrix

Confusion matrix merupakan teknik yang digunakan untuk mengetahui seberapa akurat model klasifikasi menggunakan tabel *confusion matrix* pada Tabel III sebagai berikut [6]:

Tabel III: confusion matrix

	Actual=yes	Actual=No
Predicted=Yes	FP	FN
Predicted=No	FN	TN

Pengukuran akurasi dapat dilakukan menggunakan persamaan (6) sebagai berikut :

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (6)$$

Keterangan :

TP : True Positive

FP : False Positive

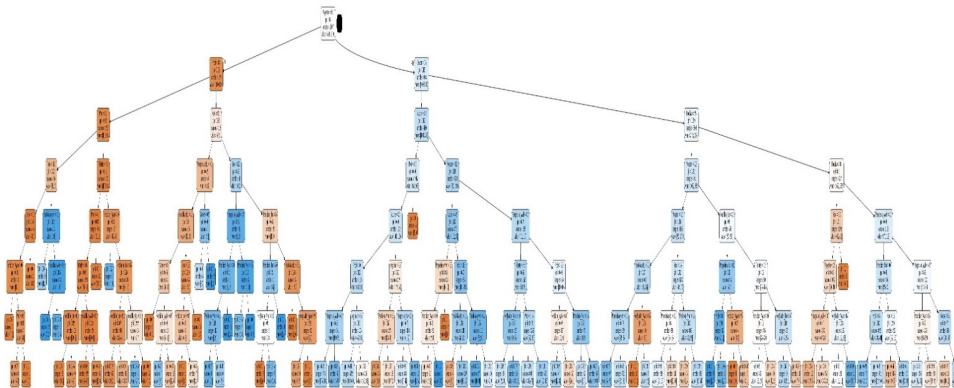
TN : True Negative

FN : False Negative

III. HASIL

A. Hasil Algoritma C4.5 Tanpa Seleksi Fitur dan Teknik Pruning

Hasil yang diperoleh menggunakan atribut fakultas, prodi, angkatan, gender, penghasilan ayah, penghasilan ibu, pendidikan ayah, dan pendidikan ibu menggunakan algoritma C4.5 dengan pemrograman python memperoleh pohon keputusan yang dapat dilihat pada Gambar 2.



Gambar 2: Visualisasi Pohon Keputusan Tanpa Seleksi Fitur dan Teknik Pruning

Gambar 2 merupakan hasil dari pohon keputusan yang telah terbentuk, adapun aturan pohon keputusan dengan penjelasan sebagai berikut:

- 1) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 1.50 , Gender ≤ 0.50 , Pendidikan Ibu ≤ 9.00 , maka mahasiswa tepat.
- 2) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 1.50 , Gender ≤ 0.50 , Pendidikan Ibu ≤ 9.00 , Prodi ≤ 0.50 , maka mahasiswa tepat.
- 3) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 1.50 , Gender ≤ 0.50 , Pendidikan Ibu ≤ 9.00 , Prodi ≤ 0.50 , maka mahasiswa tepat.
- 4) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 1.50 , Gender ≤ 0.50 , maka mahasiswa tepat.
- 5) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 1.50 , Pendidikan Ayah ≤ 5.00 , maka mahasiswa terlambat.
- 6) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 1.50 , Pendidikan Ayah ≤ 5.00 , Pendidikan Ibu ≤ 9.00 , maka mahasiswa terlambat.
- 7) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 1.50 , Pendidikan Ayah ≤ 5.00 , Pendidikan Ibu ≤ 9.00 , maka mahasiswa terlambat.
- 8) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , Prodi ≤ 4.50 , Pendidikan Ibu ≤ 6.50 , maka mahasiswa tepat.
- 9) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , Prodi ≤ 4.50 , Pendidikan Ibu ≤ 6.50 , maka mahasiswa tepat.
- 10) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , Prodi ≤ 4.50 , maka mahasiswa tepat.
- 11) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , Pendidikan Ayah ≤ 1.00 , maka mahasiswa terlambat.
- 12) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , Pendidikan Ayah ≤ 1.00 , penghasilan Ibu ≤ 4.50 , pendidikan Ibu ≤ 9.00 , maka mahasiswa tepat.
- 13) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , Pendidikan Ayah ≤ 1.00 , penghasilan Ibu ≤ 4.50 , Pendidikan Ibu ≤ 9.00 , maka mahasiswa tepat.
- 14) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , Pendidikan Ayah ≤ 1.00 , penghasilan Ibu ≤ 4.50 , Pendidikan Ibu ≤ 6.50 , maka mahasiswa terlambat.
- 15) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , Pendidikan Ayah ≤ 1.00 , penghasilan Ibu ≤ 4.50 , Pendidikan Ibu ≤ 6.50 , maka mahasiswa tepat.
- 16) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , Pendidikan Ibu ≤ 7.50 , Gender ≤ 0.50 , maka mahasiswa tepat.
- 17) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , Pendidikan Ibu ≤ 7.50 , Gender ≤ 0.50 , Pendidikan Ayah ≤ 6.50 , maka mahasiswa terlambat.
- 18) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , Pendidikan Ibu ≤ 7.50 , Gender ≤ 0.50 , Pendidikan Ayah ≤ 6.50 , maka mahasiswa tepat.
- 19) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , Pendidikan Ibu ≤ 7.50 , Fakultas ≤ 5.00 , Pendidikan Ibu ≤ 9.00 , maka mahasiswa tepat.
- 20) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , Pendidikan Ibu ≤ 7.50 , Fakultas ≤ 5.00 , Pendidikan Ibu ≤ 9.00 , maka mahasiswa tepat.
- 21) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , Pendidikan Ibu ≤ 7.50 , Fakultas ≤ 5.00 , maka mahasiswa tepat.
- 22) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , Prodi ≤ 10.50 , maka mahasiswa terlambat.
- 23) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , Prodi ≤ 10.50 , maka mahasiswa terlambat.

- 24) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\mu = 12.50$, Penghasilan Ayah $\mu = 1.00$, Pendidikan Ayah $\mu = 9.00$, Pendidikan Ayah $\sigma = 7.50$, maka mahasiswa terlambat.
- 25) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\mu = 12.50$, Penghasilan Ayah $\mu = 1.00$, Pendidikan Ayah $\mu = 9.00$, Pendidikan Ayah $\sigma = 7.50$, maka mahasiswa terlambat.
- 26) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\mu = 12.50$, Penghasilan Ayah $\mu = 1.00$, Pendidikan Ayah $\sigma = 9.00$, maka mahasiswa terlambat.
- 27) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\mu = 12.50$, Penghasilan Ayah $\sigma = 1.00$, Penghasilan Ayah $\mu = 3.50$, maka mahasiswa terlambat.
- 28) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\mu = 12.50$, Penghasilan Ayah $\sigma = 1.00$, Penghasilan Ayah $\sigma = 3.50$, maka mahasiswa terlambat.
- 29) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\sigma = 12.50$, Pendidikan Ibu $\mu = 6.50$, Pendidikan Ibu $\mu = 5.00$, Penghasilan Ibu $\mu = 1.00$, maka mahasiswa terlambat.
- 30) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\sigma = 12.50$, Pendidikan Ibu $\mu = 6.50$, Pendidikan Ibu $\mu = 5.00$, Penghasilan Ibu $\sigma = 1.00$, maka mahasiswa terlambat.
- 31) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\sigma = 12.50$, Pendidikan Ibu $\mu = 6.50$, Pendidikan Ibu $\sigma = 5.00$, maka mahasiswa terlambat.
- 32) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\sigma = 12.50$, Pendidikan Ibu $\sigma = 6.50$, Pendidikan Ayah $\mu = 7.00$, maka mahasiswa tepat.
- 33) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\sigma = 12.50$, Pendidikan Ibu $\sigma = 6.50$, Pendidikan Ayah $\sigma = 7.00$, Penghasilan Ayah $\mu = 2.50$, maka mahasiswa tepat.
- 34) Jika Angkatan $\mu = 0.50$, Prodi $\sigma = 8.50$, Prodi $\sigma = 11.50$, Prodi $\sigma = 12.50$, Pendidikan Ibu $\sigma = 6.50$, Pendidikan Ayah $\sigma = 7.00$, Penghasilan Ayah $\sigma = 2.50$, maka mahasiswa tepat.
- 35) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\mu = 3.50$, Prodi $\mu = 6.50$, Fakultas $\mu = 2.50$, Prodi $\mu = 4.00$, Penghasilan Ayah $\mu = 4.50$, maka mahasiswa terlambat.
- 36) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\mu = 3.50$, Prodi $\mu = 6.50$, Fakultas $\mu = 2.50$, Prodi $\mu = 4.00$, Penghasilan Ayah $\sigma = 4.50$, maka mahasiswa terlambat.
- 37) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\mu = 3.50$, Prodi $\mu = 6.50$, Fakultas $\mu = 2.50$, Prodi $\sigma = 4.00$, Angkatan $\mu = 2.50$, maka mahasiswa terlambat.
- 38) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\mu = 3.50$, Prodi $\mu = 6.50$, Fakultas $\mu = 2.50$, Prodi $\sigma = 4.00$, Angkatan $\sigma = 2.50$, maka mahasiswa tepat.
- 39) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\mu = 3.50$, Prodi $\mu = 6.50$, Fakultas $\sigma = 2.50$, Angkatan $\mu = 2.50$, Pendidikan Ayah $\mu = 3.50$, maka mahasiswa tepat.
- 40) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\mu = 3.50$, Prodi $\mu = 6.50$, Fakultas $\sigma = 2.50$, Angkatan $\mu = 2.50$, Pendidikan Ayah $\sigma = 3.50$, maka mahasiswa tepat.
- 41) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\mu = 3.50$, Prodi $\mu = 6.50$, Fakultas $\sigma = 2.50$, Angkatan $\sigma = 2.50$, Penghasilan Ibu $\mu = 2.50$, maka mahasiswa terlambat.
- 42) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\mu = 3.50$, Prodi $\mu = 6.50$, Fakultas $\sigma = 2.50$, Angkatan $\sigma = 2.50$, Penghasilan Ibu $\sigma = 2.50$, maka mahasiswa terlambat.
- 43) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\mu = 3.50$, Prodi $\sigma = 6.50$, maka mahasiswa tepat.
- 44) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\sigma = 3.50$, angkatan $\mu = 1.50$, Fakultas $\mu = 4.50$, Penghasilan Ayah $\mu = 3.50$, Pendidikan Ayah $\mu = 9.00$, maka mahasiswa tepat.
- 45) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\sigma = 3.50$, angkatan $\mu = 1.50$, Fakultas $\mu = 4.50$, Penghasilan Ayah $\mu = 3.50$, Pendidikan Ayah $\sigma = 9.00$, maka mahasiswa terlambat.
- 46) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\sigma = 3.50$, angkatan $\mu = 1.50$, Fakultas $\mu = 4.50$, Penghasilan Ayah $\sigma = 3.50$, maka terlambat tepat.
- 47) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\sigma = 3.50$, angkatan $\mu = 1.50$, Fakultas $\sigma = 4.50$, Prodi $\mu = 11.50$, Pendidikan Ibu $\mu = 0.50$, maka mahasiswa tepat.
- 48) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\sigma = 3.50$, angkatan $\mu = 1.50$, Fakultas $\sigma = 4.50$, Prodi $\mu = 11.50$, Pendidikan Ibu $\sigma = 0.50$, maka mahasiswa terlambat.
- 49) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\sigma = 3.50$, angkatan $\mu = 1.50$, Fakultas $\sigma = 4.50$, Prodi $\sigma = 11.50$, Pendidikan Ayah $\mu = 7.50$, maka mahasiswa terlambat.
- 50) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\sigma = 3.50$, angkatan $\mu = 1.50$, Fakultas $\sigma = 4.50$, Prodi $\sigma = 11.50$, Pendidikan Ayah $\sigma = 7.50$, maka mahasiswa terlambat.
- 51) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\sigma = 3.50$, angkatan $\sigma = 1.50$, Penghasilan Ayah $\mu = 3.50$, Fakultas $\mu = 4.50$, Penghasilan Ayah $\mu = 1.00$, maka mahasiswa terlambat.
- 52) Angkatan $\sigma = 0.50$, Gender $\mu = 0.50$, Fakultas $\sigma = 3.50$, angkatan $\sigma = 1.50$, Penghasilan Ayah $\mu = 3.50$, Fakultas $\mu = 4.50$, Penghasilan Ayah $\sigma = 1.00$, maka mahasiswa terlambat.

- 53) Angkatan ζ 0.50, Gender η = 0.50, Fakultas ζ 3.50, angkatan ζ 1.50, Penghasilan Ayah η = 3.50, Fakultas ζ 4.50, Prodi η = 11.50, maka mahasiswa terlambat.
- 54) Angkatan ζ 0.50, Gender η = 0.50, Fakultas ζ 3.50, angkatan ζ 1.50, Penghasilan Ayah η = 3.50, Fakultas ζ 4.50, Prodi ζ 11.50, maka mahasiswa terlambat.
- 55) Angkatan ζ 0.50, Gender η = 0.50, Fakultas ζ 3.50, angkatan ζ 1.50, Penghasilan Ayah ζ 3.50, Prodi η = 12.50, Pendidikan Ayah η = 5.00, maka mahasiswa terlambat.
- 56) Angkatan ζ 0.50, Gender η = 0.50, Fakultas ζ 3.50, angkatan ζ 1.50, Penghasilan Ayah ζ 3.50, Prodi η = 12.50, Pendidikan Ayah ζ 5.00, maka mahasiswa tepat.
- 57) Angkatan ζ 0.50, Gender η = 0.50, Fakultas ζ 3.50, angkatan ζ 1.50, Penghasilan Ayah ζ 3.50, Prodi ζ 12.50, Pendidikan Ayah η = 6.50, maka mahasiswa terlambat.
- 58) Angkatan ζ 0.50, Gender η = 0.50, Fakultas ζ 3.50, angkatan ζ 1.50, Penghasilan Ayah ζ 3.50, Prodi ζ 12.50, Pendidikan Ayah ζ 6.50, maka mahasiswa terlambat.
- 59) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan η = 2.50, Angkatan η = 1.50, Pendidikan Ibu η = 9.00, Penghasilan Ayah η = 0.50, maka mahasiswa terlambat.
- 60) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan η = 2.50, Angkatan η = 1.50, Pendidikan Ibu η = 9.00, Penghasilan Ayah η = 0.50, maka mahasiswa terlambat.
- 61) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan η = 2.50, Angkatan η = 1.50, Pendidikan Ibu η = 9.00, Penghasilan Ayah ζ 0.50, maka mahasiswa terlambat.
- 62) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan η = 2.50, Angkatan η = 1.50, Pendidikan Ibu ζ 9.00, Penghasilan Ayah η = 3.50, maka mahasiswa tepat.
- 63) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan η = 2.50, Angkatan η = 1.50, Pendidikan Ibu ζ 9.00, Penghasilan Ayah ζ 3.50, maka mahasiswa terlambat.
- 64) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan η = 2.50, Angkatan ζ 1.50, Prodi η = 1.50, Penghasilan Ibu η = 1.50, maka mahasiswa terlambat.
- 65) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan η = 2.50, Angkatan ζ 1.50, Prodi η = 1.50, Penghasilan Ibu ζ 1.50, maka mahasiswa tepat.
- 66) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan η = 2.50, Angkatan ζ 1.50, Prodi ζ 1.50, Pendidikan Ayah η = 2.50, maka mahasiswa tepat.
- 67) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan η = 2.50, Angkatan ζ 1.50, Prodi ζ 1.50, Pendidikan Ayah ζ 2.50, maka mahasiswa terlambat.
- 68) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan ζ 2.50, Pendidikan Ibu η = 2.50, Penghasilan Ibu η = 1.00, Prodi η = 3.50, maka mahasiswa terlambat.
- 69) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan ζ 2.50, Pendidikan Ibu η = 2.50, Penghasilan Ibu η = 1.00, Prodi ζ 3.50, maka mahasiswa terlambat.
- 70) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan ζ 2.50, Pendidikan Ibu η = 2.50, Penghasilan Ibu ζ 1.00, Penghasilan Ayah η = 2.50, maka mahasiswa tepat.
- 71) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan ζ 2.50, Pendidikan Ibu η = 2.50, Penghasilan Ibu ζ 1.00, Penghasilan Ayah ζ 2.50, maka mahasiswa tepat.
- 72) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan ζ 2.50, Pendidikan Ibu ζ 2.50, Fakultas η = 1.50, Pendidikan Ayah η = 9.00, maka mahasiswa terlambat.
- 73) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan ζ 2.50, Pendidikan Ibu ζ 2.50, Fakultas η = 1.50, Pendidikan Ayah ζ 9.00, maka mahasiswa tepat.
- 74) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan ζ 2.50, Pendidikan Ibu ζ 2.50, Fakultas ζ 1.50, Penghasilan Ibu η = 4.50, maka mahasiswa terlambat.
- 75) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas η = 2.50, Angkatan ζ 2.50, Pendidikan Ibu ζ 2.50, Fakultas ζ 1.50, Penghasilan Ibu ζ 4.50, maka mahasiswa tepat.
- 76) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas η = 3.50, Prodi η = 5.00, Angkatan η = 2.50, Penghasilan Ayah η = 1.00, maka mahasiswa tepat.
- 77) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas η = 3.50, Prodi η = 5.00, Angkatan η = 2.50, Penghasilan Ayah ζ 1.00, maka mahasiswa tepat.
- 78) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas η = 3.50, Prodi η = 5.00, Angkatan ζ 2.50, Penghasilan Ayah η = 4.50, maka mahasiswa terlambat.
- 79) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas η = 3.50, Prodi η = 5.00, Angkatan ζ 2.50, Penghasilan Ayah ζ 4.50, maka mahasiswa terlambat.
- 80) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas η = 3.50, Prodi ζ 5.00, maka mahasiswa tepat.
- 81) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas ζ 3.50, Penghasilan Ayah η = 2.50, Pendidikan Ibu η = 9.00, Prodi η = 4.50, maka mahasiswa terlambat.

- 82) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas ζ 3.50, Penghasilan Ayah η = 2.50, Pendidikan Ibu η = 9.00, Prodi ζ 4.50, maka mahasiswa terlambat.
- 83) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas ζ 3.50, Penghasilan Ayah η = 2.50, Pendidikan Ibu ζ 9.00, Prodi η = 6.50, maka mahasiswa tepat.
- 84) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas ζ 3.50, Penghasilan Ayah η = 2.50, Pendidikan Ibu ζ 9.00, Prodi ζ 6.50, maka mahasiswa telambat.
- 85) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas ζ 3.50, Penghasilan Ayah ζ 2.50, Penghasilan Ayah η = 4.50, Pendidikan Ayah η = 3.50, maka mahasiswa terlambat.
- 86) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas ζ 3.50, Penghasilan Ayah ζ 2.50, Penghasilan Ayah η = 4.50, Pendidikan Ayah ζ 3.50, maka mahasiswa tepat.
- 87) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas ζ 3.50, Penghasilan Ayah ζ 2.50, Penghasilan Ayah ζ 4.50, Fakultas η = 4.50, maka mahasiwa tepat.
- 88) Angkatan ζ 0.50, Gender ζ 0.50, Fakultas ζ 2.50, Fakultas ζ 3.50, Penghasilan Ayah ζ 2.50, Penghasilan Ayah ζ 4.50, Fakultas ζ 4.50, maka mahasiwa terlambat.

Untuk mengetahui kinerja dari model yang telah terbentuk dengan melihat nilai akurasi menggunakan confusion matrix. Adapun hasil tabel confusin matrix dapat dilihat pada Tabel IV.

Tabel IV: Hasil confusion matrix

	Actual=yes	Actual=No
Predicted=Yes	393	223
Predicted=No	256	369

Tabel IV merupakan hasil confusion matrix dengan penjelasan sebagai berikut.

- 1) True Positive (TP) dimana 393 mahasiswa diprediksi tepat membayar biaya yang memang sebenarnya mahasiswa tersebut tepat membayar biaya kuliah.
- 2) True Negative (TN) dimana 369 mahasiswa diprediksi terlambat membayar biaya kuliah yang memang sebenarnya mahasiswa tersebut terlambat membayar biaya kuliah.
- 3) False Positive (FP) dimana 223 mahasiswa diprediksi tepat membayar biaya kuliah, ternyata mahasiswa tersebut terlambat membayar biaya kuliah.
- 4) False Negative (FN) dimana 256 mahasiswa diprediksi terlambat membayar biaya kuliah, ternyata mahasiswa tersebut tepat dalam membayar biaya kuliah.

Berdasarkan hasil confusion matrix dapat dilakukan perhitungan akurasi dengan menggunakan persamaan (5).

$$\begin{aligned}
 accuracy &= \frac{393 + 369}{393 + 369 + 223 + 256} \times 100\% \\
 &= \frac{752}{1241} \times 100\% \\
 &= 62,40\%
 \end{aligned}$$

Hasil diatas menunjukkan bahwa model algoritma C4.5 untuk prediksi keterlambatan pembayaran biaya kuliah memperoleh akurasi sebesar 61,40%. Selanjutnya akan dilakukan peningkatan akurasi algoritma C4.5 dengan melakukan seleksi fitur *chi square* dan teknik *prunning*.

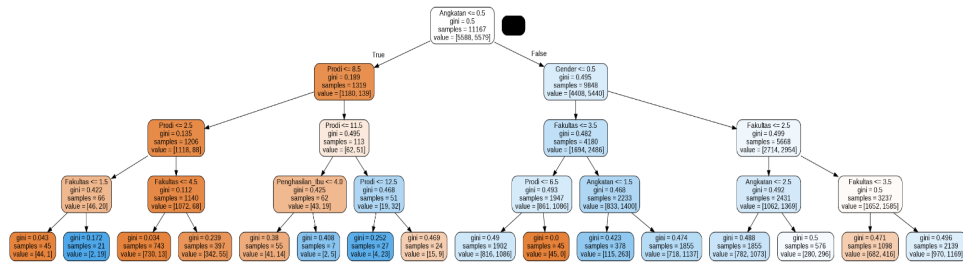
B. Hasil Algoritma C4.5 Menggunakan Seleksi Fitur dan Teknik Prunning

Hasil dari seleksi fitur *chi square* diperoleh melalui perhitungan menggunakan persamaan (1) yang di implementasikan kedalam pemrograman python, sehingga menghasilkan nilai *chi square* di masing-masing atribut yang dapat dilihat pada Tabel V. Berdasarkan hasil nilai *chi square* maka diambil 5 atribut terbaik seperti angkatan, prodi, fakultas, gender, dan penghasilan ibu yang akan digunakan untuk permodelan algoritma C4.5.

Tabel V: Hasil *chi Square*

No.	Atribut	Nilai Chi Squire
1	Angkatan	200.984
2	Prodi	236.022
3	Fakultas	70.598
4	Gender	39.506
5	Penghasilan Ibu	14.472
6	Penghasilan Ayah	3.111
7	Pendidikan Ayah	0.530
8	Pendidikan Ibu	0.030863

Hasil yang diperoleh dari teknik pruning dengan menentukan jumlah kedalaman pohon keputusan dapat dilihat pada Gambar 3 sebagai berikut:



Gambar 3: Visualisasi Pohon Keputusan Menggunakan Seleksi Fitur dan Teknik Pruning

Berdasarkan Gambar 3 merupakan visualiasi pohon keputusan yang dihasilkan melalui pemrograman python, adapun aturan keputusan yang telah terbentuk dengan penjelasan sebagai berikut:

- 1) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 1.50 , maka mahasiswa tepat.
- 2) Jika Angkatan ≤ 0.50 , prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 1.50 , maka mahasiswa terlambat.
- 3) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , maka mahasiwa tepat.
- 4) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 2.50 , Fakultas ≤ 4.50 , maka mahasiswa tepat.
- 5) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , maka mahasiswa tepat.
- 6) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Penghasilan Ibu ≤ 4.00 , maka mahasiswa terlambat.
- 7) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Prodi ≤ 12.50 , maka mahasiswa terlambat.
- 8) Jika Angkatan ≤ 0.50 , Prodi ≤ 8.50 , Prodi ≤ 11.50 , Prodi ≤ 12.50 , maka mahasiswa tepat.
- 9) Jika Angkatan ≤ 0.50 , Prodi ≤ 0.50 , Fakultas ≤ 3.50 , Prodi ≤ 6.50 , maka mahasiswa terlambat.
- 10) Jika Angkatan ≤ 0.50 , Gender ≤ 0.50 , Fakultas ≤ 3.50 , Prodi ≤ 6.50 , maka mahasiswa tepat.
- 11) Jika Angkatan ≤ 0.50 , Gender ≤ 0.50 , Fakultas ≤ 3.50 , Angkatan ≤ 1.50 , maka mahasiswa terlambat.
- 12) Jika Angkatan ≤ 0.50 , Gender ≤ 0.50 , Fakultas ≤ 3.50 , Angkatan ≤ 1.50 , maka mahasiswa terlambat.
- 13) Jika Angkatan ≤ 0.50 , Gender ≤ 0.50 , Fakultas ≤ 2.50 , Angkatan ≤ 2.50 , maka mahasiswa terlambat.
- 14) Jika Angkatan ≤ 0.50 , Gender ≤ 0.50 , Fakultas ≤ 2.50 , Angkatan ≤ 2.50 , maka mahasiswa terlambat.
- 15) Jika Angkatan ≤ 0.50 , Gender ≤ 0.50 , Fakultas ≤ 2.50 , Fakultas ≤ 3.50 , maka mahasiswa tepat.
- 16) Jika Angkatan ≤ 0.50 , Gender ≤ 0.50 , Fakultas ≤ 2.50 , Fakultas ≤ 3.50 , maka mahasiswa terlambat.

Untuk mengetahui performa atau kinerja dari model telah terbentuk dalam melakukan prediksi keterlambatan pembayaran biaya kuliah dengan melihat nilai akurasi menggunakan confusion matrix. Untuk tabel confusion matrix dapat dilihat pada Tabel VI.

Tabel VI: Confusion matrix

	Actual=yes	Actual=No
Predicted=Yes	226	390
Predicted=No	40	585

Tabel VI merupakan Hasil confusion matrix dengan penjelasan sebagai berikut.

- 1) *True Positive* (TP) dimana 226 mahasiswa diprediksi tepat membayar biaya yang memang sebenarnya mahasiswa tersebut tepat membayar biaya kuliah.
- 2) *True Negative* (TN) dimana 585 mahasiswa diprediksi terlambat membayar biaya kuliah yang memang sebenarnya mahasiswa tersebut terlambat membayar biaya kuliah.
- 3) *False Positive* (FP) dimana 390 mahasiswa diprediksi tepat membayar biaya kuliah, ternyata mahasiswa tersebut terlambat membayar biaya kuliah.
- 4) *False Negative* (FN) dimana 40 mahasiswa diprediksi terlambat membayar biaya kuliah, ternyata mahasiswa tersebut tepat dalam membayar biaya kuliah.

Berdasarkan hasil *confusion matrix* dapat dilakukan perhitungan akurasi dengan menggunakan persamaan (5).

$$\begin{aligned}
 \text{Akurasi} &= \frac{225 + 585}{226 + 585 + 390 + 40} \times 100\% \\
 &= \frac{811}{1241} 100\% \\
 &= 65,63\%
 \end{aligned}$$

Hasil dari algoritma C4.5 menggunakan seleksi fitur *chi square* dan teknik *pruning* memperoleh nilai akurasi sebesar 65,53%

IV. PEMBAHASAN

Pengujian yang telah dilakukan menunjukkan bahwa dengan adanya seleksi fitur chi square dan teknik *prunning* mampu meningkatkan nilai akurasi dari 61,40% menjadi 65,53% dengan jumlah peningkatan sebesar 4,13%. Maka untuk penggunaan seleksi fitur *chi square* dapat digunakan untuk meningkatkan akurasi, selain itu menggunakan semua fitur tanpa diseleksi dari beberapa fitur tersebut dapat membuat akurasi model menjadi lebih buruk. Kemudian teknik *prunning* dapat digunakan untuk meningkatkan akurasi dengan mencegah terjadinya *overfitting*, karena jika tingkatan pohon keputusan terlalu panjang dapat menyebabkan hasil tidak akurat ketika dihadapkan dengan data baru. Berikut ini adalah hasil dari perbandingan nilai akurasi yang dapat dilihat pada Tabel VII.

Tabel VII: Perbandingan hasil akurasi

	Algoritma C4.5	Algoritma C4.5 + Chi Square + Prunning
Akurasi	61,40%	65,53%

V. SIMPULAN

Prediksi keterlambatan pembayaran biaya kuliah dengan implementasi algoritma C4.5 menggunakan pemrograman python dapat dilakukan dengan memperoleh nilai akurasi tertinggi pada seleksi fitur *chi square* dan teknik *prunning* sebesar 65,53% untuk atribut yang digunakan diantaranya fakultas, prodi, angkatan, gender, dan penghasilan ibu. Fitur untuk menentukan keterlambatan pembayaran biaya kuliah sangat berpengaruh terhadap akurasi yang didapatkan, sehingga untuk memperoleh hasil yang lebih baik lagi menggunakan seleksi fitur lainnya seperti *forward selection* maupun *particle swarm optimization* dan menggunakan metode klasifikasi *data mining* lainnya seperti *Naïve Bayes*, *Random Forest*, dan *K-Nearest Neighbors*

PUSTAKA

- [1] V. S. Ginting, K. Kusriani, dan E. Taufiq, "Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python," *Inspir. J. Teknol. Inf. dan Komun.*, vol. 10, no. 1, hal. 36–44, 2020.
- [2] T. H. Apandi, R. B. Maulana, R. Piarna, dan D. Vernanda, "Menganalisis Kemungkinan Keterlambatan Pembayaran Spp Dengan Algoritma C4.5 (Studi Kasus Politeknik Tede Bandung)," *J. Techno Nusa Mandiri*, vol. 16, no. 2, hal. 93–98, 2019.
- [3] A. Desiani, S. Yahdin, D. Rodiah, U. Sriwijaya, P. Korespondensi, dan N. Bayes, "Prediksi Tingkat Indeks Prestasi Kumulatif Akademik Mahasiswa dengan Menggunakan Teknik Data Mining," vol. 7, no. 6, hal. 1237–1244, 2020.
- [4] E. Karyadiputra dan Z. Zaenuddin, "Penerapan Algoritma Decision Tree C4.5 Berbasis Seleksi Atribut Chi Squared Untuk Klasifikasi Tingkat Pengetahuan Ibu Dalam Pemberian Asi Eksklusif Pada Bayi," *Technol. J. Ilm.*, vol. 11, no. 1, hal. 7, 2020.
- [5] I. Iskandar, L. Hiryanto, dan J. Hendryli, "Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5 dengan Teknik Pruning," *J. Ilmu Komput. dan Sist. Inf.*, vol. 6, no. 1, hal. 64, 2018.
- [6] I. Daqiqil Id, *Machine Learning: Teori, Studi Kasus dan Implementasi Menggunakan Python*. Riau: UR PRESS, 2021.
- [7] D. Kurniawan, *Pengenalan Machine Learning dengan Python*. Jakarta: PT Elex Media Komputindo, 2020.
- [8] Y. P. Utomo, *Analisis Chi Square Untuk Penelitian Sosial (dan Medik)*. Yogyakarta: PANDIVA BUKU, 2020.
- [9] O. Somantri dan D. Apriliani, "Support Vector Machine Berbasis Feature Selection Untuk Sentiment Analysis Kepuasan Pelanggan Terhadap Pelayanan Warung dan Restoran Kuliner Kota Tegal," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 5, hal. 537, 2018.
- [10] D. Nofriansyah, *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Yogyakarta: Deepublish, 2017.
- [11] A. Rohman, *Prediksi Penyakit Jantung Menggunakan Algoritma C4.5 Berbasis ADABOOST*. Klaten: Lakeisha, 2021.
- [12] A. Rufiyanto, M. Rochcmam, dan A. Rohman, *Penerapan Algoritma C4.5 Untuk Prediksi Kepuasan Mahasiswa Tahun 2020*. Yogyakarta: Deepublish, 2021.