

Analisis Klasifikasi UKT Mahasiswa Berdasarkan Tingkat Penghasilan Orang Tua Menggunakan Algoritma C4.5

Naufal Shabir Ramadhan^{1),*}, Wulan Nadia Nursafitri¹⁾

¹⁾ Teknologi Informasi dan Komputer, Sistem Informasi, Politeknik Negeri Subang.

* wulannadianursafitri@gmail.com

Abstract

UKT (Single Tuition Fee) is a payment system for State University tuition fees where each PTN has its own regulations. To make it easier for PTNs to determine UKT prices, a data cluster technique called classification is needed. Classification is a systematic arrangement into groups or groups according to established rules or standards. With classification, it will make it easier for PTNs to know how data should be grouped so that they can get a fairly relative picture of students' socio-economic conditions. In the research conducted, the data mining classification algorithm used was the c4.5 algorithm. This algorithm is a classification of data with a relatively large number of attributes. To prove the effectiveness of the c4.5 algorithm, a classification validation test was carried out. The research results show that the c4.5 algorithm can classify students into groups with appropriate UKT based on the characteristics of their parents' income. This shows that the c4.5 algorithm can be used to make it easier for PTNs to determine UKT prices. Apart from that, the results of this research can also help in efforts to increase the accessibility of higher education for students from various economic backgrounds.

Keywords: UKT (Single Tuition Fee); Classification; Data Mining; C4.5 Algorithm; Socio-economic Conditions.

Abstrak

UKT (Biaya Pendidikan Tunggal) merupakan sistem pembayaran biaya pendidikan perguruan tinggi negeri dan setiap PTN mempunyai peraturan tersendiri. Untuk membantu PTN dalam menentukan harga UKT diperlukan suatu teknik clustering data yang disebut klasifikasi. Klasifikasi adalah pengelompokan sesuatu secara sistematis ke dalam kelompok atau kelas menurut aturan atau kriteria yang telah ditetapkan. Kategorisasi membantu PTN memahami cara mengelompokkan data dan memberikan pemahaman relatif mengenai status sosial ekonomi siswa. Penelitian yang dilakukan menggunakan algoritma c4.5 sebagai algoritma klasifikasi data mining. Algoritma ini mengklasifikasikan data dengan jumlah atribut yang relatif banyak. Uji validasi klasifikasi dilakukan untuk menunjukkan efektivitas algoritma c4.5. Hasil penelitian menunjukkan bahwa algoritma c4.5 mampu mengelompokkan siswa ke dalam kelompok dengan UKT yang sesuai berdasarkan karakteristik pendapatan orang tuanya. Hal ini menunjukkan bahwa dengan menggunakan algoritma c4.5, PTN dapat dengan mudah mengetahui harga UKT. Lebih lanjut, hasil penelitian ini juga akan membantu meningkatkan akses terhadap pendidikan tinggi bagi siswa dari latar belakang ekonomi yang berbeda.

Kata kunci: UKT (Uang Kuliah Tunggal); Klasifikasi; Data Mining; Algoritma C4.5; Kondisi Sosial Ekonomi.

Submitted: 7 November 2023

Published: 19 Februari 2024

@ 2023 Inventor

PENDAHULUAN

Kemajuan teknologi informasi di berbagai bidang turut mempengaruhi Revolusi Industri 4.0. Hampir seluruh aktivitas kini dilakukan secara digital dan otomatis. Sistem digitalisasi dan otomasi ini juga merambah ke sistem pemerintahan, memudahkan bahkan menggantikan pekerjaan pegawai negeri. Selain itu, digitalisasi layanan mahasiswa baru di perguruan tinggi telah melahirkan database berskala besar (big data). Biaya Pendidikan Seragam (UKT) merupakan sistem pembiayaan perkuliahan yang dibayarkan oleh mahasiswa pada perguruan tinggi negeri di bawah Departemen Riset, Teknologi, dan Pendidikan Tinggi (Rokhman, 2017). Pendidikan tinggi merupakan salah satu aspek terpenting dalam pembangunan suatu Negara (Ghazali et al., 2018). Pendidikan merupakan salah satu aspek penting dalam pengembangan sumber daya manusia. Sebab pendidikan merupakan alat yang membantu membebaskan masyarakat tidak hanya dari keterbelakangan namun juga dari kebodohan dan kemiskinan. Universitas Nasional Indonesia (PTN)

berperan penting dalam memberikan pendidikan tinggi yang berkualitas kepada masyarakat setempat. Pasal 1 ayat (3) Peraturan Menteri Pendidikan dan Kebudayaan menyebutkan bahwa lump sum atau UKT adalah bagian dari biaya pendidikan perseorangan yang dibayarkan oleh setiap peserta didik sesuai dengan potensi keuangannya (Irianto, 2017). Pendidikan Tinggi Nomor 12 Tahun 2012 Pasal 88 menetapkan bahwa menteri berwenang menetapkan standar satuan biaya operasional. Dalam konteks UKT, tingkat penghasilan orang tua mahasiswa sering menjadi faktor penentu dalam menentukan besaran UKT yang harus mereka bayar (Pranatawijaya et al., 2018). Oleh karena itu, sangat penting bagi PTN untuk memiliki cara yang efisien dan adil dalam menentukan tarif UKT yang sesuai bagi setiap mahasiswa. Salah satu teknik yang dapat digunakan untuk mengatasi masalah ini adalah teknik pengelompokan data atau klasifikasi. Dalam beberapa tahun terakhir, PTN di Indonesia telah mengalami peningkatan signifikan dalam jumlah mahasiswa yang mendaftar (Marleny et al., 2015). Besarnya UKT yang harus di bayar oleh masing-masing mahasiswa mengalami kesulitan karena harus memperhatikan pertimbangan kondisi perekonomian orang tua/ wali mahasiswa dalam pengelompokan. Oleh karena itu, diperlukan teknik pengambilan keputusan. Teknik tersebut berguna untuk mengelompokkan besarnya UKT sesuai dengan kemampuan ekonomi masing-masing mahasiswa (Larytasari et al., 2019). Oleh karena itu pemerintah melalui Menteri Pendidikan dan Kebudayaan (Kemendikbud) pada tanggal 23 Mei 2013 telah mengeluarkan ketetapan mengenai besarnya Uang Tunggal Kuliah pada perguruan tinggi negeri (PTN) di lingkungan kementerian pendidikan dan kebudayaan (Kemendikbud) (Muchsin & Sudarma, 2015).

Klasifikasi adalah teknik yang dapat digunakan untuk mengelompokkan data ke dalam kategori atau kelas yang berbeda berdasarkan karakteristik tertentu (Nurjanah & Rifai, 2023). Algoritma C4.5 merupakan program yang memberi kontribusi satu set data berlabel dan menghasilkan pohon keputusan sebagai keluaran. Pohon keputusan tindak lanjut ini kemudian diverifikasi terhadap data uji berlabel yang tidak terlihat untuk menghitung generalisasinya. C4.5 adalah program yang digunakan untuk menghasilkan peraturan taksonomi dengan menggunakan pohon keputusan dari sekumpulan data yang diberikan (Rofik & Prehanto, 2020).

Tujuan utama dari penelitian ini adalah untuk melakukan analisis klasifikasi UKT mahasiswa berdasarkan tingkat penghasilan orang tua menggunakan algoritma data mining. Dengan melakukan ini, penelitian ini bertujuan untuk mencapai beberapa tujuan yang lebih spesifik diantaranya: (1) Mengembangkan model klasifikasi yang dapat mengelompokkan mahasiswa ke dalam kelompok tarif UKT yang sesuai berdasarkan karakteristik penghasilan orang tua mereka. (2) Menerapkan algoritma data mining, khususnya algoritma C4.5, untuk menghasilkan model klasifikasi yang akurat. (3) Melakukan uji validasi klasifikasi untuk memastikan keakuratan dan ketepatan model klasifikasi yang telah dikembangkan. (4) Menganalisis hasil klasifikasi untuk memberikan wawasan yang lebih baik tentang distribusi tarif UKT di antara mahasiswa berdasarkan tingkat penghasilan orang tua.

Penelitian ini memiliki signifikansi yang besar dalam konteks pendidikan tinggi di Indonesia. Untuk membantu penetapan kategori yang tepat pada setiap mahasiswa, dapat digunakan pendekatan Data Mining dengan algoritma tertentu. Dimana Data Mining merupakan suatu proses untuk menemukan informasi yang bermanfaat dari sekumpulan data yang besar. Ada beberapa faktor yang mempengaruhi penetapan UKT tersebut, antara lain kondisi orang tua, pendidikan orang tua, pekerjaan orang tua, penghasilan orang tua, jumlah tanggungan, dan kepemilikan rumah (Karim & Sentinuwo, 2017). Dengan mengembangkan model klasifikasi merupakan tatanan yang sangat penting dalam menambang data komunitas. Klasifikasi adalah teknik penambangan data prediktif yang menggunakan hasil yang diketahui dari kumpulan data yang berbeda untuk membuat prediksi tentang data nilai (Prasetya et al., 2022). Selain itu, penelitian ini juga berkontribusi pada upaya meningkatkan aksesibilitas pendidikan tinggi bagi semua individu, sehingga pendidikan tinggi tetap menjadi pintu gerbang untuk mencapai masa depan yang lebih cerah. Penelitian ini membahas pentingnya klasifikasi UKT mahasiswa berdasarkan tingkat penghasilan orang tua menggunakan algoritma data mining. Dengan mengembangkan model klasifikasi yang efisien dan akurat, PTN dapat meningkatkan efisiensi dan akurasi proses penentuan UKT. Ini akan memastikan bahwa biaya pendidikan tinggi lebih terjangkau bagi semua mahasiswa, tanpa mengorbankan kualitas pendidikan.

METODE

Salah satu algoritma yang dapat digunakan untuk membuat pohon keputusan (*decision tree*) adalah algoritma C4.5 ini. Algoritma ini merupakan algoritma yang sangat populer yang digunakan oleh banyak peneliti di dunia, hal ini dijelaskan oleh Xindong Wu dan Vipin Kumar dalam bukunya yang berjudul *The*

Top Ten Algorithms in Data Mining. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang di ciptakan oleh J.Rose Quinlan (Wu et al., 2008). Di bawah ini adalah langkah-langkah umum dalam menggunakan metode penelitian C4.5 yaitu sebagai berikut: (1) **Pengumpulan Data**: Tahap pertama dalam penelitian dengan C4.5 adalah mengumpulkan data yang relevan untuk analisis. Data ini harus mencakup atribut atau fitur yang akan digunakan untuk membuat pohon keputusan, serta marker atau kelas yang akan diprediksi. (2) **Preprocessing Data**: Data seringkali perlu diolah sebelum digunakan dengan C4.5. Ini termasuk mengatasi data yang hilang, menghapus atau mengganti outlier, dan mengubah format data jika diperlukan. (3) **Pembuatan Pohon Keputusan**: C4.5 membangun pohon keputusan dengan menggunakan algoritma rekursif. Algoritma ini bekerja dengan memilih atribut terbaik untuk membagi data pada setiap tahap dan membuat simpul keputusan berdasarkan atribut tersebut. (4) **Pemilihan Atribut Terbaik**: C4.5 menggunakan metrik seperti Gain Ratio atau Gini Index untuk menentukan atribut terbaik untuk membagi data pada setiap simpul. Atribut yang memberikan pemisahan yang paling informatif akan dipilih. (5) **Pemotongan Pohon**: Setelah pohon keputusan dibangun, terkadang pohon tersebut bisa menjadi terlalu kompleks, yang dapat menyebabkan overfitting. C4.5 melakukan pemotongan pohon dengan menghapus cabang-cabang yang tidak signifikan atau menggantinya dengan simpul daun. (6) **Validasi dan Evaluasi**: Setelah pohon keputusan dibangun, Anda perlu menguji kinerjanya menggunakan data pengujian yang belum pernah dilihat sebelumnya. Anda dapat menggunakan metrik seperti akurasi, presisi, recall, atau F1-score untuk mengevaluasi kinerja model. (7) **Optimisasi dan Tuning**: Terkadang, Anda perlu melakukan optimisasi parameter atau tuning model untuk meningkatkan kinerjanya. (8) **Implementasi**: Setelah Anda puas dengan model pohon keputusan yang dibangun, Anda dapat mengimplementasikannya dalam aplikasi praktis untuk mengambil keputusan berdasarkan data baru.

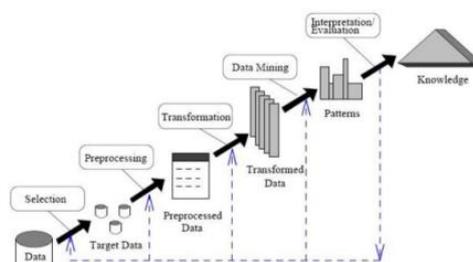
Pengumpulan data pada penelitian ini mengandalkan data sekunder. Data yang digunakan adalah dataset UKT mahasiswa berdasarkan pekerjaan orang tua yang nantinya akan dilakukan decision tree sebagai suatu keputusan. Dataset UKT mahasiswa dapat dirincikan seperti pada gambar 1.

Tempat Tinggal	Pekerjaan Orang Tua	Penghasilan Orang Tua	Kendaraan	Kelayakan Keringanan UKT	PENGELUARAN ORANG TUA
0 PNS		1000000		1 0	4000000
0 TNI/POLRI		800000		2 1	3500000
1 Petani		4000000		0 0	1500000
1 Dokter		3000000		1 0	10000000
0 Buruh		2000000		1 1	1000000
1 Ibu Rumah Tangga		500000		0 0	2500000
0 PNS		9000000		2 1	4000000
1 Wiraswasta		600000		1 1	3000000
1 Dosen		700000		1 1	4000000
0 Ibu Rumah Tangga		300000		1 0	1500000
0 TNI/POLRI		800000		2 1	3500000
1 Petani		400000		0 0	1500000
1 Buruh		200000		1 0	1000000
0 Dosen		700000		1 1	3000000
0 Ibu Rumah Tangga		300000		1 1	1500000
1 PNS		1000000		2 1	4000000
1 Wiraswasta		600000		1 1	3000000
0 Petani		400000		0 0	1000000
0 Buruh		200000		1 1	1000000

Gambar 1. Dataset UKT mahasiswa berdasarkan pekerjaan orang tua

Metode pengumpulan data ini adalah menggunakan kaggle. Kaggle adalah komunitas ilmu data terbesar di dunia dengan alat dan sumber daya canggih untuk membantu Anda mencapai tujuan ilmu data Anda (kaggle).

Dalam Tahapan Penelitian, Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran *computer (machine learning)* untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis (Eska, 2018). Metode penelitian yang digunakan penulis dalam Menerapkan Algoritma C4.5 untuk klasifikasi pembayaran UKT mahasiswa berdasarkan pekerjaan orang tua adalah metode dengan proses *knowledge Discovery in Databases (KDD)*: Berikut tahapan-tahapan yang akan dilakukan dalam penelitian ini.



Gambar 2. Alur Penelitian

Berdasarkan alur penelian pada gambar 2 yang meliputi beberapa Tahap: (1) Seleksi Data (*Selection*), Tahapan ini memilih atribut mana yang akan digunakan dalam proses data mining dengan menggunakan atribut yang dipilih. Keenam atribut tersebut diantaranya adalah tempat tinggal, pekerjaan orang tua, pendapatan orang tua, mobil, hak keringanan UKT, dan pengeluaran orang tua. (2) Pemilihan Data (*Preprocessing/Cleaning*), Preprocessing adalah proses menghilangkan data duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan data seperti kesalahan ketik dan data yang hilang. Operator *replace missing values*. Akan digunakan untuk menghapus kumpulan data tanpa nilai (hilang). (3) Transformasi (*Transformation*), Pada tahap ini adalah tentang mengkonversi data yang belum memiliki entitas yang jelas menjadi bentuk data yang valid atau siap untuk data mining. (4) *Data Mining*, Pada tahap ini data akan diolah menggunakan algoritma C4.5. Proses data mining pembayaran UKT mahasiswa berdasarkan pekerjaan orang tua akan menggunakan tool Rapid Miner. (5) Interpretasi/Evaluasi (*Interpretation/Evaluation*), Pada tahap interpretasi/evaluasi adalah proses menghasilkan hasil yang dapat dipahami dari proses pemodelan informasi data mining (Nurjanah & Rifai, 2023).

HASIL DAN PEMBAHASAN

Ada beberapa tahapan dalam membuat sebuah pohon keputusan dalam algoritma C4.5 yaitu: (1) Mempersiapkan data *training*. Data *training* biasanya diambil dari data historis yang sudah pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu. (2) Menghitung akar pohon. Akar akan diambil dari atribut yang akan dipilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi akan menjadi akar pertama. Sebelum menghitung nilai gain dari atribut, hitung dulu nilai entropy (Ardiansyah, 2019).

Proses seleksi pemberian uang kuliah tunggal memerlukan sebuah sistem yang dapat membantu penilaian kriteria-kriteria yang ada. Adapun kriteria-kriteria tersebut sebagai berikut, Tempat tinggal, Pekerjaan Orangtua, Penghasilan Orangtua, Kendaraan, Kelayakan keringanan UKT, dan Pengeluaran Orangtua. Adapun hasil dan pembahasan dataset dengan menggunakan metode algoritma C4.5 pada analisis klasifikasi UKT mahasiswa berdasarkan tingkat penghasilan orang tua menggunakan algoritma data mining.

Pada Perhitungan Manual Menggunakan Excel dengan Algoritma C4.5, penelitian ini digunakan software Rapidminer. Untuk perhitungan manual menggunakan excel dengan kumpulan data 101 record dan untuk pengujian. Berikut langkah-langkah proses pohon keputusan dengan perhitungan manual algoritma C4.5 menggunakan data latih terkait pembayaran UKT. Data atribut pada penelitian meliputi 6 atribut seperti disajikan ada gambar 3.

Tempat Tinggal	Pekerjaan Orang Tua	Penghasilan Orang Tua	Kendaraan	Kelayakan Keringanan UKT	PENGELUARAN ORANG TUA
0 PNS		10000000	1 0		4000000
0 TNI/POLRI		8000000	2 1		3500000
1 Petani		4000000	0 0		1500000
1 Dokter		30000000	1 0		10000000
0 Buruh		2000000	1 1		1000000
1 Ibu Rumah Tangga		5000000	0 0		2500000
0 PNS		9000000	2 1		4000000
1 Wiraswasta		6000000	1 1		3000000
1 Dosen		7000000	1 1		4000000
0 Ibu Rumah Tangga		3000000	1 0		1500000
0 TNI/POLRI		8000000	2 1		3500000
1 Petani		4000000	0 0		1500000
1 Buruh		2000000	1 0		1000000
0 Dosen		7000000	1 1		3000000
0 Ibu Rumah Tangga		3000000	1 1		1500000
1 PNS		10000000	2 1		4000000
1 Wiraswasta		6000000	1 1		3000000
0 Petani		4000000	0 0		1000000
0 Buruh		2000000	1 1		1000000

Gambar 3. Pekerjaan orang tua

Data yang digunakan adalah dataset UKT mahasiswa berdasarkan pekerjaan orang tua yang nantinya akan dilakukan decision tree sebagai suatu keputusan. Atribut atribut yang ada pada penelitian terdiri dari 6 diantaranya: (1) Tempat Tinggal, (2) Pekerjaan Orang Tua, (3) Penghasilan Orang Tua, (4) Kendaraan, (5) Kelayakan Keringanan UKT. (6) Pengeluaram Orang Tua.

Pada preprocessing data, Data yang telah diolah tadi selanjutnya dilakukan perubahan format yang sebelumnya txt di ubah menjadi csv sehingga lebih terstruktur seperti yang terditampilkan pada gambar 3. Pada gambar 3 yaitu memuat Tingkat penghasilan Orang Tua yang dimana diakumulasikan dari: Tempat

Tinggal, Pekerjaan Orang Tua, Penghasilan Orang Tua, Kendaraan, Kelayakan Keringanan UKT dan Pengeluaran Orang Tua.

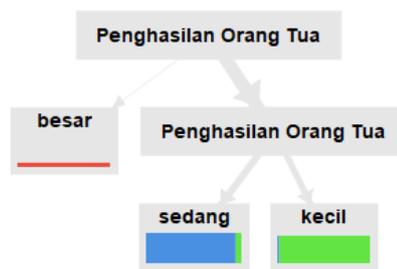
Ketika Membuat pohon keputusan, Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal. Dimana metode pohon keputusan mengubah fakta yang sangat besar menjadi poton keputusan yang dapat mempresentasikan aturan-aturan. Aturan tersebut dapat dengan mudah dipahami ataupun dijelaskan dengan ukang karena memiliki bahasa alami yang dapat dimengerti. Pohon keputusan juga berguna untuk menjelajahi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Pohon keputusan memiliki fleksibilitas (mudah digunakan) yang membuatnya sangat menarik, terutama karena menyajikan keuntungan dari proses pengubahan konsep menjadi gambar yang telah disajikan sangat mudah dimengerti dimana cabang-cabang dari pohon merangkum klasifikasi (Putri et al., 2021). Metode ini berfungsi untuk mengubah fakta menjadi pohon keputusan yang mempresentasikan aturan yang dapat mudah dimengerti dengan bahasa alami (Harryanto & Hansun, 2017). Data yang telah di ubah format sebelumnya selanjutnya di lakukan pemrosesan menggunakan aplikasi rapid miner untuk di uji untuk melakukan pengklasifikasi dataset yang telah ada tadi.

Row No.	status	Tempat Ting...	Pekerjaan O...	Penghasilan...	Kendaraan	Kelayakan K...	PENGELUAR...
1	sedang	0	PNS	10000000	1	0	4000000
2	sedang	0	TNI/POLRI	8000000	2	1	3500000
3	kecil	1	Petani	4000000	0	0	1500000
4	besar	1	Dokter	30000000	1	0	10000000
5	kecil	0	Buruh	2000000	1	1	1000000
6	kecil	1	Ibu Rumah T...	5000000	0	0	2500000
7	sedang	0	PNS	9000000	2	1	4000000
8	sedang	1	Wiraswasta	6000000	1	1	3000000
9	sedang	1	Dosen	7000000	1	1	4000000
10	kecil	0	Ibu Rumah T...	3000000	1	0	1500000
11	sedang	0	TNI/POLRI	8000000	2	1	3500000
12	kecil	1	Petani	4000000	0	0	1500000
13	kecil	1	Buruh	2000000	1	0	1000000
14	sedang	0	Dosen	7000000	1	1	3000000
15	kecil	0	Ibu Rumah T...	3000000	1	1	1500000

ExampleSet (100 examples, 1 special attribute, 6 regular attributes)

Gambar 4. Proses pemberian label pada dataset

Pada gambar 4 menjelaskan proses pemilihan label pada klasifikasi, saat mengimport data di RapidMiner dan perlu memilih jenis atribut "label" untuk kolom yang akan di klasifikasikan. Alur pembuatan decision tree seperti pada gambar 5.



Gambar 5. Pembuatan decision tree

Ketika pemilihan atribut terbaik, Data yang telah diberi label selanjutnya melakukan pemilihan atribut terbaik berikut penghitungan atributnya, sehingga atribut terbaik dapat di sajikan dalam data gambar 6.

Pekerjaan Orang ... Category	Penghasilan Orang Tua Number
PNS	10000000
TNI/POLRI	8000000
Petani	4000000
Dokter	30000000
Buruh	2000000
Ibu Rumah Tangga	5000000
PNS	9000000
Wiraswasta	6000000
Dosen	7000000

Gambar 6. Pemilihan atribut terbaik yaitu penghasilan orang tua dan statusnya

Label status UKT besar (5 data), status UKT sedang (48 data), dan status UKT kecil (47 data). Kemudian, kita akan menghitung *Information Gain* untuk atribut “Pengeluaran Orang Tua” dan “Status” (Status UKT) dari dataset yang diberikan, langkah-langkah yang dilakukan adalah sebagai berikut: Langkah pertama yaitu menghitung Entropi awal dengan rumus berikut,

$$S = -(1025 \cdot \log_2(1025) + 10248 \cdot \log_2(10248) + 10247 \cdot \log_2(10247)) \dots\dots\dots(1)$$

Menghitung Entropi awal (Entropi S) dari dataset dengan jumlah total data (N) sebesar 102 dan dengan masing-masing kategori status UKT yaitu Status UKT Besar (5 data), Status UKT Sedang (48 data), dan Status UKT Kecil (47 data). Kemudian, langkah kedua adalah menghitung entropi setelah pemisahan untuk atribut “Pengeluaran Orang Tua” dengan dua kategori, yaitu “Rendah” dan “Tinggi”. Entropi dihitung untuk masing-masing kategori setelah pemisahan tersebut dilakukan.

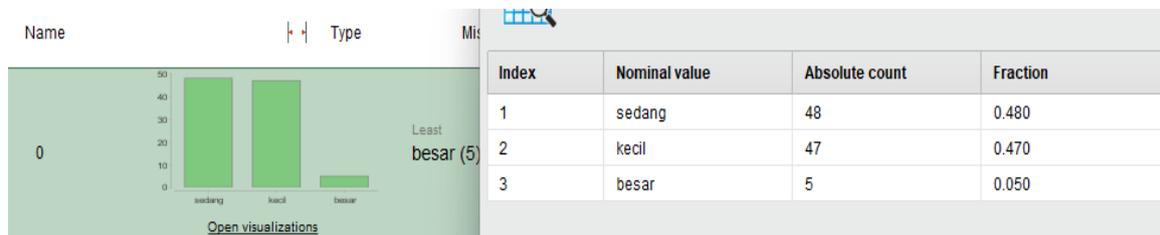
Selanjutnya, untuk atribut “Pengeluaran Orang Tua”, *Information Gain* dihitung dengan mengurangi entropi awal (S) dengan jumlah berbobot dari entropi setelah pemisahan berdasarkan kategori “Rendah” dan “Tinggi”, yang kemudian diberi bobot sesuai dengan jumlah data pada masing-masing kategori.

Langkah terakhir, untuk atribut “Status” (Status UKT) yang memiliki tiga kategori yaitu “Kecil”, “Sedang”, dan “Besar”, *Information Gain* dihitung dengan cara yang serupa. Entropi dihitung setelah pemisahan untuk masing-masing kategori “Kecil”, “Sedang”, dan “Besar”, kemudian *information gain* dihitung dengan mengurangi entropi awal (S) dengan jumlah berbobot dari entropi setelah pemisahan berdasarkan ketiga kategori tersebut, yang kemudian diberi bobot sesuai dengan jumlah data pada masing-masing kategori.

Dengan melakukan perhitungan seperti di atas, mendapatkan *Information Gain* untuk kedua atribut, "Pengeluaran Orang Tua" dan "Status" (Status UKT), dalam konteks dataset yang telah diberikan. Atribut yang memiliki *Information Gain* yang lebih tinggi akan lebih informatif dalam memisahkan data.

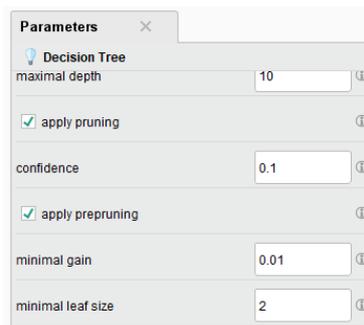
Pemotongan pohon pada algoritma C4.5 merupakan strategi penting untuk mencegah overfitting, di mana model terlalu cocok dengan data latih sehingga kehilangan kemampuannya.

Ketika melakukan Validasi dan Evaluasi, untuk menghitung nilai akurasi dan nilai *Decision tree* yang telah dibangun dilakukan pengujian *stittistic matrix*, untuk mengevaluasi kinerja dari *decision tree* yang telah dibuat.



Gambar 7. *Statistic matrix*

Kemudian melakukan Optimisasi dan tuning, Optimisasi model dengan tuning parameter *decision tree* membantu mencari nilai terbaik untuk parameter seperti kedalaman pohon, jumlah sampel di leaf node, dan kriteria pemisahan. Teknik seperti Grid Search atau Random Search digunakan untuk menemukan kombinasi parameter terbaik yang meningkatkan performa model, mengurangi overfitting, dan meningkatkan keakuratan prediksi. Melakukan optimisasi parameter atau tuning model untuk meningkatkan kinerjanya.



Gambar 8. Parameter decision tree

Pada tahap Implementasi, Pastikan data dipersiapkan dengan baik, pilih model decision tree yang sesuai, latih model dengan data training, evaluasi kinerja dengan data testing, gunakan model untuk membuat prediksi pada data baru, analisis hasil, dan jika diperlukan, lakukan penyesuaian model untuk meningkatkan performa. Implementasi menggunakan aplikasi sederhana menggunakan data baru yang telah didapat.



Gambar 9. Hasil Decision Tree

Gambar 9 menjelaskan jika penghasilan orang tua lebih dari dua puluh juta termasuk kedalam kategori besar dan pada decision tree ini ada lima orang tua yang berpenghasilan lebih dari dua puluh juta. Penghasilan orang tua kurang dari dua puluh juta dengan jumlah nol. Penghasilan orang tua lebih dari lima juta termasuk kedalam kategori sedang ada empat 47 dan termasuk kecil ada tiga orang. Penghasilan orang tua kurang dari lima juta ada dengan kategori sedang ada satu orang dan kategori kecil ada empat puluh empat orang.

Pada penelitian ini dijelaskan bahwa prosedur analisis pemeringkatan UKT (Uang Kuliah Tunggal) mahasiswa berdasarkan tingkat pendapatan orang tuanya dengan menggunakan algoritma C4.5. Tujuan utama penelitian ini adalah untuk mendukung pengambilan keputusan mahasiswa terkait Uang Kuliah Tunggal dengan menggunakan algoritma data mining dan pohon keputusan (decision tree).

Proses penelitian diawali dengan pengumpulan data yang mencakup sejumlah atribut, seperti tempat tinggal, pekerjaan orang tua, pendapatan orang tua, sarana transportasi, dan kelayakan mendapatkan bantuan pengobatan, tunjangan Uang Kuliah Tunggal, dan pengeluaran orang tua. Data ini kemudian diolah dan dibersihkan untuk memastikan validitas dan keakuratannya. Data tersebut kemudian dikonversi ke format Comma Separated Values (CSV) untuk memudahkan analisis.

Setelah data siap, langkah selanjutnya adalah membuat pohon keputusan dengan menggunakan algoritma C4.5. Pohon keputusan ini membantu mengklasifikasikan mahasiswa ke dalam kelompok UKT yang sesuai berdasarkan atribut yang dianalisis. Dalam proses ini, memilih properti terbaik sangatlah penting. Dalam contoh ini, pendapatan orang tua ternyata menjadi atribut yang paling informatif dalam pengambilan keputusan.

Setelah membuat pohon keputusan, pengujian dilakukan dengan menggunakan data pengujian. Hasil pengujian dievaluasi menggunakan statistik seperti akurasi, presisi, recall, dan skor F1 untuk mengukur performa model (Ghazali et al., 2023). Hasil evaluasi ini membantu mengevaluasi seberapa akurat model dapat mengklasifikasikan siswa. Selain itu, model pohon keputusan dapat dioptimalkan dengan menyesuaikan parameter atau penyetulan. Tujuannya adalah untuk meningkatkan performa dari model dengan mencoba konfigurasi yang berbeda tentunya dengan lainnya. Dengan optimalisasi yang tepat, model dapat menjadi lebih efektif dalam mengklasifikasikan data pada siswa.

Setelah model diuji dan dioptimalkan, maka langkah selanjutnya adalah penerapan. Model pohon keputusan yang optimal dapat diimplementasikan dalam aplikasi sederhana. Hal ini memungkinkan pengguna untuk memasukkan data baru dan menerima prediksi UKT berdasarkan karakteristik orang tuanya.

Tabel 1. Hasil klasifikasi berdasarkan karakteristik

Penghasilan Orang Tua	Penghasilan Orang Tua	Hasil	Jumlah
> 20.000.000	-	Besar	5
≤ 20.000.000	> 5.500.000	Sedang	47
≤ 20.000.000	≤ 5.500.000	Kecil	44

Jadi, berdasarkan metode kuantitatif, kita dapat menyimpulkan bahwa: Jumlah kasus dengan penghasilan besar adalah 5. Jumlah kasus dengan penghasilan sedang adalah 47. Jumlah kasus dengan penghasilan kecil adalah 44. Harap dicatat bahwa hasil ini didasarkan pada informasi yang diberikan dan asumsi bahwa "sedang," "kecil," dan "besar" adalah kategori atau label yang diberikan untuk menggambarkan suatu hasil atau kondisi.

SIMPULAN

Kesimpulan dari penelitian ini adalah algoritma C4.5 sangat efektif digunakan untuk mengelompokkan mahasiswa ke dalam kelompok UKT berdasarkan pendapatan orang tua. Pendapatan orang tua menjadi faktor terpenting dalam keputusan ini. Hasil penelitian ini dapat digunakan untuk mendukung pengambilan keputusan mengenai UKT mahasiswa dan meningkatkan pengelolaan UKT di perguruan tinggi. Selain itu, penting untuk mempertimbangkan aspek etika dan privasi saat menggunakan data siswa dalam penelitian ini. Dengan lebih memahami faktor-faktor yang mempengaruhi keputusan UKT, lembaga pendidikan dapat lebih memberikan dukungan kepada siswa yang membutuhkan.

DAFTAR PUSTAKA

- Ardiansyah, D. (2019). Algoritma c4. 5 untuk klasifikasi calon peserta lomba cerdas cermat siswa smp dengan menggunakan aplikasi rapid miner. *Jurnal Inkofar*, 1(2).
- Eska, J. (2018). *Penerapan Data Mining Untuk Prediksi Penjualan Wallpaper Menggunakan Algoritma C4. 5*.
- Ghazali, M., Fitriati, I., & Purnamasari, R. (2018). Support Vector Regression for Modeling Effect of Education Rate on Life Expectancy Rate in Indonesia. *IJECA (International Journal of Education and Curriculum Application)*, 64–69.
- Ghazali, M., Zohri, M., Lestari, J., Purnamasari, R., & Fitriati, I. (2023). Transformasi kemampuan analisis statistik mahasiswa UIN Mataram melalui pelatihan software JASP. *ABSYARA: Jurnal Pengabdian Pada Masyarakat*, 4(2), 235–247.
- Harryanto, F. F., & Hansun, S. (2017). Penerapan Algoritma C4. 5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 3(2), 95–103.
- Irianto, H. A. (2017). *Pendidikan sebagai investasi dalam pembangunan suatu bangsa*. Kencana.
- Karim, B., & Sentinuwo, S. (2017). Penentuan Besaran Uang Kuliah Tunggal untuk Mahasiswa Baru di Universitas Sam Ratulangi Menggunakan Data Mining. *Jurnal Teknik Informatika*, 11(1).
- Larytasari, L. A., Susanti, Y., & Respatiwan, R. (2019). PENENTUAN UKT MAHASISWA UNS DENGAN ALGORITMA ITERATIVE DICHOTOMISER THREE DAN CLASSIFICATION VERSION 4.5. *Seminar & Conference Proceedings of UMT*.
- Marleny, F. D., Junaidi, H. M., & Mambang, M. (2015). PENERAPAN K-MEANS CLUSTER UNTUK PENGARUH KECERDASAN EMOSI DAN STRES TERHADAP PRESTASI BELAJAR MAHASISWA. *SEMNASTEKNOMEDIA ONLINE*, 3(1), 1–2.
- Muchsin, A. K., & Sudarma, M. (2015). Penerapan Fuzzy C-Means Untuk Penentuan Besar Uang Kuliah Tunggal Mahasiswa Baru. *Lontar Komput. J. Ilm. Teknol. Inf*, 6(3), 175.
- Nurjanah, A., & Rifai, A. (2023). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Kelayakan Status Penduduk Miskin Di Desa Susukan Tonggoh. *Jurnal Wahana Informatika*, 2(1), 164–176.
- Pranatawijaya, V. H., Putra, P. B. A. A., & Sari, N. N. K. (2018). Pengembangan Perangkat Lunak Generate File Akun Uang Kuliah Tunggal (UKT) Universitas Palangka Raya. *Jurnal Saintekom*, 8(2), 166–178.
- Prasetya, F. D., Nugroho, H. W., & Triloka, J. (2022). Analisa Data Mining Untuk Prediksi Penyakit Hepatitis C Menggunakan Algoritma Decision Tree C. 45 Dengan Particle Swarm Optimization. *Prosiding Seminar Nasional Darmajaya*, 1, 198–209.
- Putri, S. U., Irawan, E., & Rizky, F. (2021). Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4. 5. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer Dan Manajemen)*, 2(1), 39–46.

- Rofik, A., & Prehanto, D. R. (2020). Sistem Informasi Pendukung Keputusan Penerimaan dan Monitoring Beasiswa Menggunakan Algoritma C4. 5 (Studi Kasus Universitas Negeri Surabaya). *Journal of Emerging Information System and Business Intelligence (JEISBI)*, 1(1), 16–23.
- Rokhman, M. T. N. (2017). Improving managerial performance through participation role of budget preparation: a theoretical and empirical overview. *Journal of Economics and Finance*, 8(1), 39–43.
- Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., & Yu, P. S. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14, 1–37.