

PREDIKSI KELULUSAN MAHASISWA DENGAN MENGGUNAKAN ALGORITMA DECISION TREE

Reki Kurnia Permana, Mamok Andri Senubekti, Novita Anggraini, Muhammad Syafiuddin Usman

Manajemen Informatika, Universitas Teknologi Digital Bandung

Jl. Cibogo indah III – Bodogol Rt.08/05 Kel. Mekarjaya Kec. Rancasari, Kota Bandung, Indonesia

reki22380022@digitechuniversity.ac.id

ABSTRAK

Penelitian ini dilatarbelakangi oleh kebutuhan akan pengelolaan data akademik yang semakin kompleks di Universitas Teknologi Digital yang seiring dengan meningkatnya volume data akademik. Permasalahan yang diangkat adalah belum adanya sistem prediksi kelulusan mahasiswa yang dapat membantu pihak universitas dalam mengantisipasi mahasiswa yang berisiko tidak lulus tepat waktu. Tujuan penelitian ini adalah untuk membangun model prediksi kelulusan mahasiswa berbasis algoritma Decision Tree yang dapat digunakan sebagai alat bantu pengambilan keputusan. Metode yang digunakan meliputi pengumpulan data akademik mahasiswa, pra-pemrosesan data menggunakan perangkat lunak RapidMiner, pembuatan model prediksi dengan algoritma Decision Tree, serta evaluasi model melalui confusion matrix untuk mengukur akurasi, presisi, dan recall. Hasil penelitian menunjukkan bahwa model Decision Tree yang dibangun memiliki akurasi prediksi sekitar 89–90%, dengan variabel utama yang memengaruhi kelulusan adalah IPS semester 2, IPS semester 3, dan status mahasiswa (bekerja atau tidak bekerja) dari 161 data yang diambil dari mahasiswa tahun 2020 dan 2021 di Universitas. Model ini diharapkan dapat mendukung pengambilan keputusan berbasis data di lingkungan pendidikan tinggi

Kata kunci : Data mining, Pendidikan, Prediksi, Decision Tree, Klasifikasi

1. PENDAHULUAN

Pertumbuhan cepat teknologi informasi dalam beberapa tahun terakhir telah membawa dampak yang sangat besar terhadap cara kita mengelola, menganalisis, dan memanfaatkan data. Data yang dihasilkan setiap hari kini semakin besar, kompleks, dan beragam, yang dikenal dengan istilah Big Data. Menurut [1] ilmu mengekstraksi informasi yang berguna dari kumpulan data besar atau database dikenal sebagai data mining, jika datanya berupa text maka disebut text mining [2], [3]. Ini adalah disiplin baru, terletak di persimpangan statistik, pembelajaran mesin, manajemen data dan database, pengenalan pola, kecerdasan buatan, dan area lainnya. Semua ini berhubungan dengan aspek-aspek tertentu dari analisis data, sehingga mereka memiliki banyak kesamaan tetapi masing-masing juga memiliki cita rasa yang berbeda, yang menekankan pada masalah dan jenis solusi tertentu. Fenomena ini tidak hanya terjadi dalam dunia bisnis dan industri, tetapi juga merambah ke berbagai sektor lain, termasuk pendidikan. Dalam konteks pendidikan, pengelolaan dan analisis data akademik yang dihasilkan oleh mahasiswa, dosen, serta kegiatan pembelajaran lainnya menjadi semakin penting, karena dapat memberikan wawasan berharga untuk meningkatkan kualitas pendidikan itu sendiri.

Menurut [4] tingkat kelulusan mahasiswa bisa diprediksi menggunakan suatu sistem. Namun, beberapa perguruan tinggi belum memiliki sistem untuk memprediksi keterlambatan kelulusan mahasiswa, sehingga perguruan tinggi tidak dapat melakukan pencegahan akan hal tersebut. Beberapa peneliti sudah membuat sebuah sistem untuk memprediksi tingkat kelulusan mahasiswa. Dalam

penelitian [5] yang berjudul “Implementasi Algoritma Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa” menghasilkan akurasi sebesar 80,55% dengan menggunakan data training sebesar 170 data. Sedangkan untuk penelitian [6] yang berjudul “Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika” menghasilkan akurasi sebesar 62,44%. variabel yang digunakan antara lain: jenis kelamin, asal daerah, IPK, dan TOEFL. Hasil akurasi yang diperoleh dari penelitian sebelumnya masih kecil dan belum termasuk dalam kategori baik.

Studi ini berfokus pada Universitas Teknologi Digital yang berlokasi di Kampus 1, Jl. Cibogo No.Indah 3, Mekarjaya, Kec. Rancasari, Kota Bandung, Jawa Barat, sebagai objek penelitian. Fokus utama penelitian ini adalah pada pengolahan data kelulusan mahasiswa menggunakan metode Decision Tree dan perangkat lunak RapidMiner. Dengan menggunakan data yang mencakup informasi akademik mahasiswa, seperti nama, NIM, jenis kelamin, tahun masuk, tahun lulus, IPK, dan berbagai data terkait lainnya, diharapkan dapat diidentifikasi faktor-faktor utama yang mempengaruhi kelulusan mahasiswa. Melalui penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam tentang elemen-elemen yang mempengaruhi keberhasilan akademik mahasiswa dan bagaimana faktor-faktor tersebut dapat dimanfaatkan untuk meningkatkan kualitas pendidikan di universitas tersebut serta ke akurasian data yang lebih baik dari penelitian sebelumnya.

2. TINJAUAN PUSTAKA

2.1. Prediksi Kelulusan

Kelulusan adalah salah satu kriteria penilaian dalam proses akreditasi institusi suatu perguruan tinggi, jika mahasiswa lulus tepat waktu akan membantu penilaian akreditasi suatu perguruan tinggi. Sebaliknya, keterlambatan kelulusan mahasiswa sering kali sulit dideteksi secara dini. Oleh karena itu, dibutuhkan teknik prediksi, salah satunya melalui data mining dengan metode klasifikasi [7].

Beberapa penelitian terdahulu telah dilakukan untuk memprediksi kelulusan mahasiswa dengan berbagai algoritma. Dalam penelitian yang dilakukan oleh [6] "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika" menggunakan variabel seperti jenis kelamin, asal daerah, IPK, dan TOEFL. Penelitian tersebut menghasilkan akurasi sebesar 62,44%, yang dinilai masih rendah dan belum masuk dalam kategori baik sehingga perlu dikaji metode lain untuk meningkatkan akurasi.

Selain itu, [5] menerapkan algoritma Support Vector Machine (SVM) dan memperoleh akurasi sebesar 80,55% menggunakan 170 data mahasiswa. Sementara [4] menggunakan Naïve Bayes dan menekankan perlunya tambahan variabel non-akademik untuk meningkatkan performa model.

2.2. Data

Data adalah deskripsi atau keterangan sebuah objek yang belum memiliki arti sepenuhnya yang dapat berbentuk angka (numerik), karakter (text), gambar, suara, ataupun lambang (simbol). Menurut [8] data dapat didefinisikan sebagai fakta-fakta yang dikumpulkan, diukur, atau dianalisis, yang memiliki potensi untuk memberikan wawasan bagi pengambilan keputusan. Hasil dari pengolahan data dalam suatu bentuk yang lebih berguna dan lebih mudah dipahami penerima dapat disebut sebagai informasi. Informasi mempresentasikan lebih lanjut kejadian-kejadian nyata yang akan diaplikasikan dalam pengambilan keputusan.

Dalam kehidupan sehari-hari, data mendasari hampir semua aspek, mulai dari keputusan sederhana seperti pemilihan rute terbaik dalam perjalanan hingga keputusan yang lebih kompleks seperti perencanaan keuangan atau strategi bisnis. Dengan munculnya Big Data saat ini proses pengumpulan, penyimpanan, dan analisis data menjadi lebih dinamis dan canggih dalam proses pengambilan keputusan. Dengan memanfaatkan teknologi dan alat analisis canggih, dapat memperoleh wawasan yang lebih dalam, mengatasi keterbatasan tradisional, dan membuat keputusan yang lebih tepat waktu dan terinformasi [9]. Sebagai contoh, aplikasi pemetaan yang mengandalkan data lalu lintas real-time dapat membantu pengguna menghindari kemacetan,

sementara platform e-commerce menggunakan data perilaku konsumen untuk merekomendasikan produk yang relevan. Lebih jauh lagi, menurut [10] dalam *The Handbook of Science and Technology Studies*, data juga berperan penting dalam pembangunan sosial dan ekonomi, memberikan wawasan yang diperlukan untuk merancang kebijakan publik yang lebih efektif dan merespons kebutuhan masyarakat. Dengan kata lain, data bukan hanya alat untuk efisiensi, tetapi juga menjadi dasar dalam upaya untuk memajukan kualitas hidup secara keseluruhan.

2.3. Data Mining

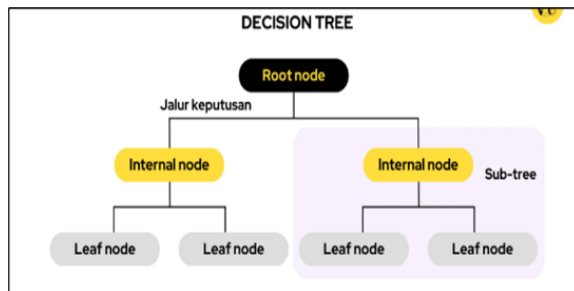
Data mining merupakan suatu proses untuk menggali informasi atau pola yang tersembunyi dalam sejumlah besar data. Data mining mencari informasi baru, berharga dan berguna dalam sekumpulan data dengan melibatkan komputer dan manusia serta bersifat iteratif baik melalui proses yang otomatis ataupun manual [11]. Teknik ini digunakan untuk menganalisis data yang sangat besar dan kompleks guna menemukan hubungan atau pola yang mungkin tidak terlihat secara langsung. Proses ini melibatkan penggunaan berbagai metode statistik, algoritma machine learning, dan kecerdasan buatan untuk mengeksplorasi data dan menarik kesimpulan yang berguna.

2.4. Decision Tree

Decision Tree atau Pohon keputusan adalah sebuah struktur data yang terdiri dari simpul (node) dan rusuk (edge) simpul pada sebuah pohon dibedakan menjadi tiga, yaitu simpul akar (root node), simpul percabangan (branch node) dan simpul daun atau leaf node [12]. Algoritma Decision Tree merupakan suatu metode pengklasifikasian yang menggunakan Gambaran pohon, menyatakan node yang menggambarkan tiap atribut, yang mana daun menggambarkan tiap kelas, juga setiap cabangnya menggambarkan nilai dari tiap kelas. Node akar menyatakan node yang berada paling atas dari pohon. Setiap node ini menggambarkan node pembagi, yang mana tiap node ini merupakan satu masukan dan memiliki sedikitnya dua keluaran [13].

3. METODE PENELITIAN

Penelitian ini menggunakan metode klasifikasi Decision Tree untuk memprediksi kelulusan mahasiswa. Algoritma ini dipilih karena kemampuannya menyajikan model dalam bentuk pohon keputusan yang mudah dipahami dan diinterpretasikan oleh pengguna non-teknis, seperti pihak akademik. Decision Tree mampu menangani data dengan kombinasi variabel numerik dan kategorikal, serta efektif dalam mengidentifikasi aturan keputusan berbasis atribut utama yang memengaruhi kelulusan mahasiswa pada Gambar 1.



Gambar 1. Decision tree

Dari Gambar 1 diatas dapat diuraikan beberapa tahapan penting adalah sebagai berikut :

- a. Pra-pemrosesan data:
Data dibersihkan dari duplikasi dan nilai hilang. Operator Replace Missing Value dan Impute Missing Values (dengan algoritma K-NN) pada RapidMiner digunakan untuk menangani missing value.
- b. Pembangunan model Decision Tree:
Model dibangun menggunakan algoritma Decision Tree dengan kriteria gain ratio sebagai pemilih atribut terbaik pada setiap node. Metode validasi silang (cross validation) digunakan untuk mengukur performa model.
- c. Evaluasi model:
Model diuji menggunakan confusion matrix untuk menghitung akurasi, presisi, recall, dan F1-score. Hasil model divisualisasikan dalam bentuk pohon keputusan untuk memudahkan interpretasi.
- d. Pengujian data baru:
Data uji dari mahasiswa angkatan terbaru diterapkan ke model untuk memprediksi status kelulusan berdasarkan pola pohon keputusan yang dihasilkan.

Algoritma decision tree dapat didefinisikan sebagai metode pembelajaran mesin yang menggunakan representasi visual menyerupai diagram alur dengan struktur hierarkis seperti pohon. Teknik ini berfungsi sebagai alat bantu dalam proses pengambilan keputusan serta penyelesaian permasalahan yang terkait dengan tugas prediksi dan pengelompokan data. Dalam struktur decision tree, komponen utama terdiri dari beberapa elemen penting. Node utama atau titik awal (root node) berfungsi sebagai representasi dari permasalahan utama yang hendak diselesaikan. Dari node utama ini, terbentuk percabangan yang menggambarkan berbagai alternatif jalur pengambilan keputusan, dimana setiap percabangan akan menuju ke node-node perantara (internal nodes).

Setiap algoritma decision tree dapat memiliki multiple internal nodes yang berfungsi sebagai opsi-opsi alternatif dalam proses decision making. Node-node perantara ini dapat berkembang lebih lanjut menjadi node-node terminal (leaf nodes) yang merepresentasikan hasil keputusan final dari proses analisis.

3.1. Sumber Data

Data yang didapatkan merupakan data kelulusan mahasiswa dari Universitas Teknologi Digital tahun 2020 dan 2021 sejumlah 161 data, dengan persetujuan dari bidang LPPM dan Akademik. Data yang didapat dari berbagai program studi, meskipun terbatas dan tidak terlalu lengkap dalam pemberian data yang diinginkan, namun sudah bisa ditelaah menjadi data yang informatif.

3.2. Data Uji

Data uji didapatkan dari hasil google form yang dibagikan kepada mahasiswa Universitas Teknologi Digital angkatan 2022 dan 2023 sejumlah 13 data yang akan diprediksi sesuai model data decision tree nanti yang akan dibuat

3.3. Software Rapidminer

RapidMiner adalah platform analitik data dan kecerdasan buatan (AI) yang memudahkan proses analisis data, mulai dari persiapan data hingga penerapan model prediktif. Pengolahan data yang dilakukan menggunakan software rapidminer, karena memberikan pemahaman dan kemudahan dalam penggunaan operator atau fitur yang ada untuk mengolah data. Perangkat lunak ini bersifat open source dan dibangun dengan menggunakan program Java di bawah lisensi GNU Public Licence dan RapidMiner dapat dijalankan pada sistem operasi apapun. Tidak diperlukan keahlian coding khusus untuk menggunakan RapidMiner karena semua model sudah disediakan.. Model yang ditawarkan juga cukup banyak dan lengkap, seperti Model Bayesian, Modelling, Tree Induction, Neural Network, Decision Tree dan lain-lain

3.4. Hasil

- a. Column chart adalah jenis grafik batang yang umum digunakan untuk membandingkan nilai antar kategori. Column chart menyajikan batang secara vertikal atau tegak lurus. Jenis grafik ini sangat efektif untuk membandingkan nilai dari beberapa kategori dan terutama berguna dalam menampilkan perubahan data dari waktu ke waktu (data deret waktu) atau ketika label kategorinya singkat. Tinggi kolom secara intuitif merepresentasikan besarnya nilai
- b. Confusion Matrix merupakan tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi, baik itu klasifikasi biner (dua kelas) maupun multikelas. Tabel ini membandingkan hasil prediksi model dengan nilai sebenarnya, sehingga memungkinkan kita untuk memahami bagaimana model membuat kesalahan dan seberapa baik kinerjanya.
- c. Pohon keputusan adalah alat visualisasi berbentuk struktur pohon yang digunakan dalam pengambilan keputusan dan klasifikasi. Setiap simpul internal dalam pohon merepresentasikan sebuah fitur (atau atribut), setiap cabang

merepresentasikan aturan keputusan, dan setiap simpul daun merepresentasikan hasil atau keputusan akhir.

3.5. Wawancara

Salah satu metode pengumpulan data yang melibatkan interaksi langsung antara peneliti dengan narasumber (informan) untuk mendapatkan informasi mendalam terkait data penelitian. Setelah mengajukan proposal perizinan kepada pihak LPPM, wawancara untuk penelitian dilakukan kepada pihak akademik, sesuai data yang dibutuhkan dan sesuai perizinan yang sudah diterbitkan.

4. HASIL DAN PEMBAHASAN

4.1. Data

Terdapat data yang akan diproses untuk diolah terlebih dahulu, yaitu data latih dari data kelulusan mahasiswa tahun 2020 dan 2021. Data yang didapatkan berasal dari akademik melalui wawancara dengan variabel nama, ips, ipk dan lain-lain. Data ini akan diproses untuk pembuatan model prediksinya. Berikut merupakan beberapa data kelulusan mahasiswa pada Tabel 1 berikut :

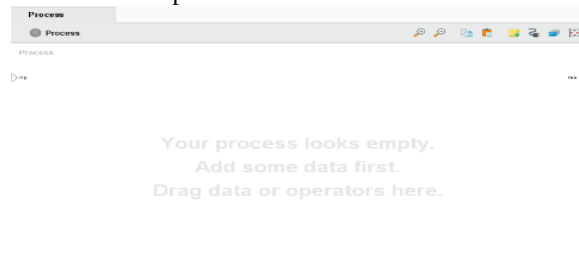
Tabel 1. Data kelulusan mahasiswa tahun 2020-2021

Nama	Program Studi	Tahun Masuk	IPK	Tahun Lulus	Status Kelulusan	Jenis Kelamin	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	Status Mahasiswa
Ulan Habibah Ummu Salamah	Manajemen S1	2020	3,51	2023	Tepat Waktu	Perempuan	3,26	3,28	3,19	3,46	3,29	2,9835	Bekerja
Umayatul Azizah	Manajemen S1	2020	3,63	2023	Tepat Waktu	Perempuan	3,51	3,34	3,43	3,30	3,58	3,0855	Bekerja
Abdah Maghfuroh	Ekonomi Syariah S1	2020	3,74	2023	Tepat Waktu	Laki-Laki	3,48	3,59	3,62	3,73	3,63	3,179	
Aisyah	Ekonomi Syariah S1	2020	3,68	2023	Tepat Waktu	Perempuan	3,37	3,44	3,60	3,57	3,54	3,128	Belum Bekerja
Amir Fatah	Ekonomi Syariah S1	2020	3,66	2023	Tepat Waktu	Laki-Laki	3,59	3,42	3,55	3,60	3,34	3,111	Belum Bekerja
Muhammad Rizal Khoirul Insan	Manajemen Informatika D3	2021	0,72	2023	Terlambat	Laki-Laki	0,60	0,64	0,69	0,65	0,60	0,828	Bekerja
Muhammad Rifky Pratama	Manajemen Informatika D3	2021	0,65	2023	Terlambat	Laki-Laki	0,46	0,56	0,61	0,58	0,57	0,7475	Bekerja

4.2. Preprocessing Data

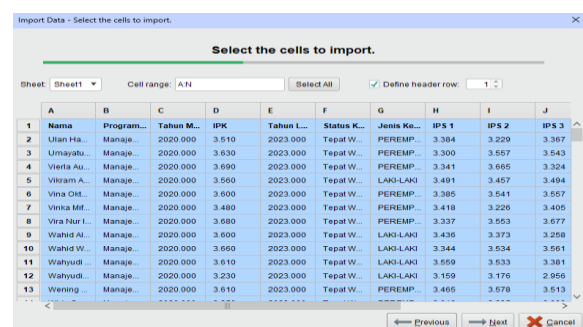
Pra-pemrosesan data merupakan salah satu langkah penting dalam proses data mining. Sebelum data diproses lebih lanjut, data mentah perlu melalui tahap pengolahan awal. Pada tahap ini, data yang tidak relevan atau tidak sesuai biasanya akan disaring dan dihapus. Selain itu, data juga akan dikonversi ke dalam format yang lebih mudah dipahami oleh sistem. Berikut merupakan langkah-langkah sebelum melakukan preprocessing data menggunakan software Rapid Miner yang dilakukan pada penelitian ini :

- Membuat new blank process atau halaman baru yang didalamnya terdapat suatu proses alur data bisa dilihat pada Gambar 2.



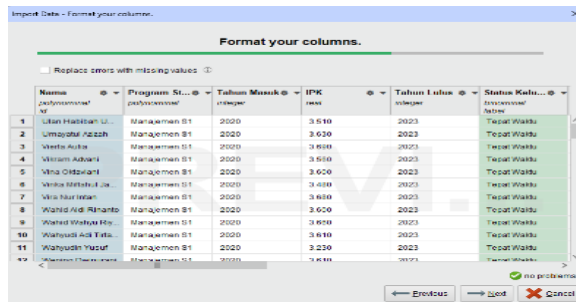
Gambar 2. New blank process

- Mengimport data yang sudah diberikan pihak akademik atau data yang akan diteliti dapat berupa excel ataupun csv. Usahakan ketika membuat suatu data pada bagian paling atas atau column 1 row 1 itu akan dihitung sebagai head atau column, jadi sesuaikan dengan kebutuhan. Gunakanlah operator Read Excel untuk membaca data Excel bisa dilihat pada Gambar 3.



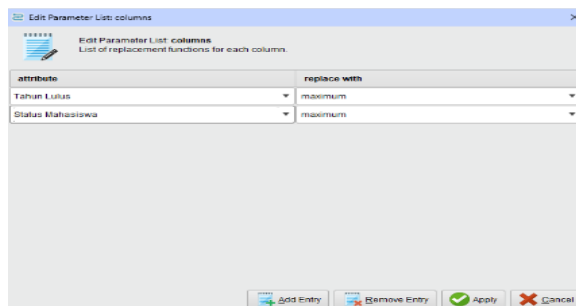
Gambar 3. Import data excel

c. Langkah berikutnya adalah format column, sesuaikan kolom-kolom dengan formatnya, seperti pada kolom yang menjadi keputusan untuk menentukan sebuah nilai diberi format label diubah rolenya, dan untuk angka seperti biasa integer atau real. Disini saya menggunakan Role pada Variabel data Status kelulusan sebagai label dan Nama sebagai Id yang bisa dilihat pada Gambar 4.



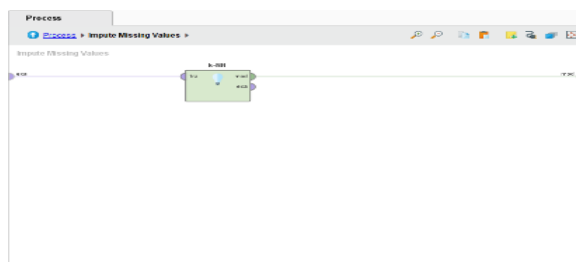
Gambar 4. Format kolom

d. Gunakan operator Replace Missing Value, gunanya untuk menentukan kolom yang terdapat missing value dapat diisi / di replace. Untuk replace with disini, saya menggunakan maximum dari data yang ada, karena posisi attribute data tidak diketahui sama sekali pada attribute tahun lulus dan status kelulusan bisa dilihat pada Gambar 5.



Gambar 1. Operator replace missing value

e. Gunakan operator Impute Missing Values, dimana operator ini untuk mengisi kembali atau menuliskan operator replace sebelumnya. Nah disini saya menggunakan algoritma K-NN (K-Nearest Neighbor) ini adalah algoritma machine learning yang digunakan untuk mengklasifikasikan data berdasarkan kedekatan dengan data lain yang sudah diklasifikasikan bisa dilihat pada Gambar 6.

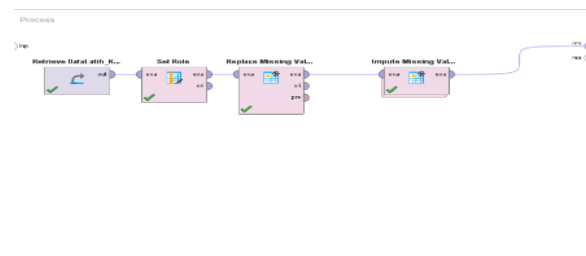


Gambar 2. Operator impute missing values

f. Run program dan lihat hasil dari statistik apakah masih ada missing value atau tidak. Berikut merupakan hasil dari data cleaning bisa dilihat pada Gambar 7 dan alur data nya pada Gambar 8.

Nama	Status Kelulusan	Program Studi	Tahun Masuk	IPK	Tahun Lulus	Jenis Kelamin
Ulan Haddisan U.	Tepat Waktu	Manajemen S1	2020	3.510	2023	T
Umayyah Adzah	Tepat Waktu	Manajemen S1	2020	3.620	2023	T
Winda Aulia	Tepat Waktu	Manajemen S1	2020	3.600	2023	T
Widani Adhara	Tepat Waktu	Manajemen S1	2020	3.500	2023	T
Vina Oktavia	Tepat Waktu	Manajemen S1	2020	3.600	2023	T
Vina Oktavia	Tepat Waktu	Manajemen S1	2020	3.400	2023	T
Vita Nur Hafid	Tepat Waktu	Manajemen S1	2020	3.400	2023	T
Wahid Ridwan	Tepat Waktu	Manajemen S1	2020	3.600	2023	T
Wahid Wahyu R.	Tepat Waktu	Manajemen S1	2020	3.800	2023	T
Wahid Adhara	Tepat Waktu	Manajemen S1	2020	3.810	2023	T
Wahid Nur	Tepat Waktu	Manajemen S1	2020	3.220	2023	T
Wahid Nur	Tepat Waktu	Manajemen S1	2020	3.400	2023	T

Gambar 3. Hasil data clean

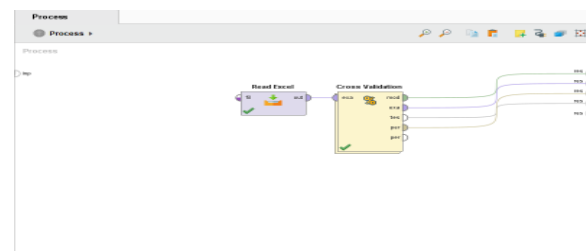


Gambar 4. Alur data

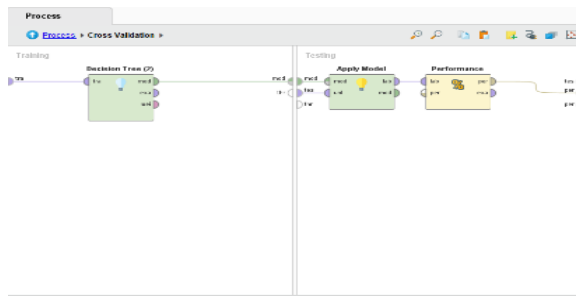
4.3 Model Decision Tree

Sebagaimana kita ketahui, Decision Tree merupakan metode pengambilan keputusan yang menyusun setiap opsi menjadi bentuk yang bercabang. Dengan begitu membuat metode ini disebut pohon keputusan. Setelah melakukan preprocessing data langkah selanjutnya yaitu, membuat sebuah prediksi untuk hasil yang ingin kita uji menggunakan model yang banyak seperti bayesian, tree, lazy, dan lain-lain. Tentunya ketika sudah menentukan model, hasil yang didapatkan merupakan ke akuratan prediksi data tersebut. Disini kita menggunakan model tree yaitu decision tree. Berikut merupakan langkah nya :

a. Gunakan operator read excel dan masukan data yang sudah kita lakukan proses data cleaning, data yang dimana sudah bersih dari missing values dan sudah lengkap. Biasakan untuk format column nya disesuaikan dengan sebelumnya. Setelah itu kita akan uji performa model ini dengan algoritma Decision tree yang bisa dilihat pada Gambar 9 ubah kriteria dengan gain ratio dan gunakan metode cross Validation yang bisa dilihat pada Gambar 10.



Gambar 5. Langkah tes model



Gambar 6. Langkah dalam cross validation

- b. Kita bisa melihat hasil dari model ini akurasi prediksi berapa persen dari confusion matrix dan model decision tree seperti apa pada Gambar 11.

Table View Plot View

accuracy: 90.00% +/- 6.72% (micro average: 90.06%)

	true Tepat Waktu	true Terlambat	class precision
pred. Tepat Waktu	112	12	90.32%
pred. Terlambat	4	33	89.19%
class recall	96.55%	73.33%	

Gambar 7. Confusion matrix

Kita harus ketahui dulu tabel confusion matrix ini, dimana tiap kolom memiliki istilah. Pada bagian paling atas horizontal True lulus dan True belum lulus itu disebut dengan Actual/Ground Truth, dan bagian vertical paling kiri itu disebut Predictive value. Selanjutnya ada istilah TP (True positif) yang diisi angka 112, TN (True Negative) angka 33, FP (False Positif) diangka 12, dan FN (False Negative) angka 4. Kegunaan dari tabel confusion matrix ini adalah untuk mengukur : accuracy, recall, precision, dan f1. berikut merupakan rumus confusion matrix accuracy :

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Jika dilihat dari rumus maka, data yang sudah kita uji dapat dihitung dengan rumus accuracy tersebut.

$$accuracy = \frac{112 + 33}{112 + 33 + 4 + 12} = \frac{145}{161} = 0,90$$

Untuk menghitung Precision berikut rumus nya :

$$Precision1 = \frac{TP}{TP+FP} \text{ \& } Precision2 = \frac{TN}{TN+FN}$$

Maka didapat :

$$Precision1 = \frac{112}{112+4} = \frac{112}{116} = 0,965$$

$$Precision2 = \frac{33}{33+12} = \frac{33}{45} = 0,73$$

Untuk menghitung Recall berikut rumusnya :

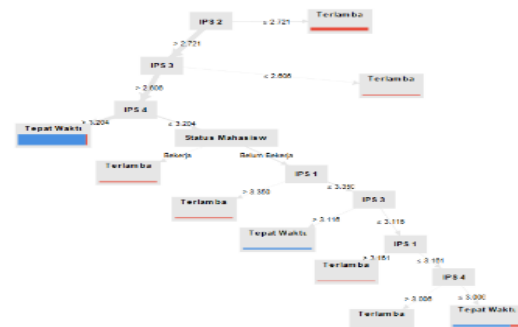
$$Recall1 = \frac{TP}{TP+FN} \text{ \& } Recall2 = \frac{TN}{TN+FP}$$

Maka didapat :

$$Recall1 = \frac{112}{112 + 12} = \frac{112}{124} = 0,9032$$

$$Recall2 = \frac{33}{33 + 4} = \frac{33}{37} = 0,8919$$

- c. Lalu, ini merupakan model decision tree yang didapatkan untuk acuan bagi data uji. Gambaran model ini merupakan hasil dari data tahun 2020 dan 2021 yang ditentukan oleh model decision tree. Algoritma decision tree dapat didefinisikan sebagai metode pembelajaran mesin yang menggunakan representasi visual menyerupai diagram alur dengan struktur hierarkis seperti pohon. Teknik ini berfungsi sebagai alat bantu dalam proses pengambilan keputusan serta penyelesaian permasalahan yang terkait dengan tugas prediksi dan pengelompokan data. Bisa dilihat pada Gambar 12.



Gambar 8. Model decision tree

IPS 2 > 2.721
 | IPS 3 > 2.606
 | | IPS 4 > 3.204: Tepat Waktu {Tepat Waktu=96, Terlambat=2}
 | | IPS 4 ≤ 3.204
 | | | Status Mahasiswa = Bekerja: Terlambat {Tepat Waktu=0, Terlambat=6}
 | | | Status Mahasiswa = Belum Bekerja
 | | | IPS 1 > 3.350: Terlambat {Tepat Waktu=0, Terlambat=2}
 | | | IPS 1 ≤ 3.350
 | | | | IPS 3 > 3.116: Tepat Waktu {Tepat Waktu=10, Terlambat=0}
 | | | | IPS 3 ≤ 3.116
 | | | | IPS 1 > 3.161: Terlambat {Tepat Waktu=0, Terlambat=2}
 | | | | IPS 1 ≤ 3.161
 | | | | | IPS 4 > 3.006: Terlambat {Tepat Waktu=0, Terlambat=2}
 | | | | | IPS 4 ≤ 3.006: Tepat Waktu {Tepat Waktu=10, Terlambat=2}
 | IPS 3 ≤ 2.606: Terlambat {Tepat Waktu=0, Terlambat=2}
 IPS 2 ≤ 2.721: Terlambat {Tepat Waktu=0, Terlambat=27}

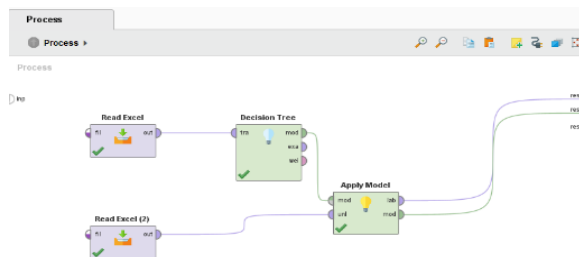
4.4 Interpretasi Hasil

Setelah melewati process pembuatan model decision tree, inteprestasi hasil yang akan didapat merupakan pengujian dari data uji yang sudah disiapkan dari data Tahun angkatan 2022 dan 2023. Berikut merupakan datanya pada Tabel 2 :

Tabel 2. Data uji tahun 2022 dan 2023

Nama	Program Studi	Tahun Masuk	Tahun Lulus	Status Kelulusan	Jenis Kelamin	IPS 1	IPS 2	IPS 3	IPS 4	Status Mahasiswa
Reki Kurnia Permana	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	3,8	3,71	3,85	4	Belum Bekerja
M. Zidan Ramadhan	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	3,83	3,77	3,84	3,96	Belum Bekerja
Muhammad Faizal Pratama	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	3,67	3,73	3,85	3,88	Belum Bekerja
Andri Maulana Yusuf	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	3	3,56	3,25	3,09	Belum Bekerja
Thomas Lazuardi Muhammad	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	3,7	3,83	3,95	3,85	Bekerja
Yaris Fajar Baihaqi	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	3,4	3,4	3,4	3,4	Belum Bekerja
Ilman Nur Rohman Yusuf	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	3	3	3	3	Belum Bekerja
Chandra Ramadhan	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	3,9	2,98	3,21	2,89	Bekerja
Rifki Fajar Nugraha	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	2,8	2,94	3,44	3,21	Bekerja
Seanno Shaggy Irawansyah	D3 Manajemen Informatika	2022	2025	?	Laki-Laki	3,42	3,03	2,89	2,97	Belum Bekerja
Novitasari Ramadhani Lubis	D3 Manajemen Informatika	2022	2025	?	Perempuan	3,74	3,69	3,84	3,91	Belum Bekerja
Adinda Salsabila	D3 Manajemen Informatika	2022	2025	?	Perempuan	3,65	3,68	3,34	3,69	Belum Bekerja
Chica Anggraeni	D3 Manajemen Informatika	2023	2026	?	Perempuan	3,78	3,66	3,97	3,7	Belum Bekerja

Buat process seperti pada Gambar 13 ini, gunakan operator decision tree dan gunakan kriteria gain ratio, gunakan operator apply model untuk memproses data uji.

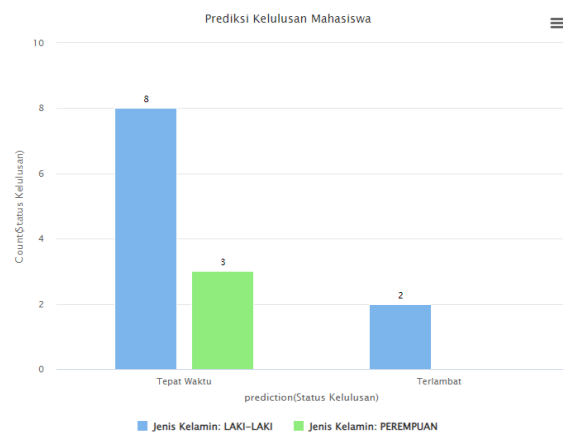


Gambar 9. Proses pengujian hasil

Kita bisa lihat hasilnya seperti pada Gambar 14 untuk hasil prediksi dan Gambar 15 visualisasi dengan bar column mengenai hasil prediksi dibawah ini :

na	Stat.	prediction(S...	conf.	conf.	Program St...	Tah...	Ta...	Jenis Kelamin	IPS 1	IPS 2	IPS 3
Reki Kurnia P...	?	Tepat Waktu	0.990	0.020	D3 Manajem...	2022	2025	LAKI-LAKI	3.800	3.710	3.850
M. Zidan Ra...	?	Tepat Waktu	0.990	0.020	D3 Manajem...	2022	2025	LAKI-LAKI	3.830	3.770	3.840
Muhammad F...	?	Tepat Waktu	0.980	0.020	D3 Manajem...	2022	2025	LAKI-LAKI	3.670	3.730	3.850
Andri Maula...	?	Tepat Waktu	1	0	D3 Manajem...	2022	2025	LAKI-LAKI	3	3.560	3.250
Thomas Laz...	?	Tepat Waktu	0.980	0.020	D3 Manajem...	2022	2025	LAKI-LAKI	3.700	3.830	3.950
Yaris Fajar...	?	Tepat Waktu	0.980	0.020	D3 Manajem...	2022	2025	LAKI-LAKI	3.400	3.400	3.400
Ilman Nur R...	?	Tepat Waktu	0.833	0.167	D3 Manajem...	2022	2025	LAKI-LAKI	3	3	3
Chandra Ram...	?	Terlambat	0	1	D3 Manajem...	2022	2025	LAKI-LAKI	3.900	2.980	3.210
Rifki Fajar...	?	Tepat Waktu	0.980	0.020	D3 Manajem...	2022	2025	LAKI-LAKI	2.800	2.940	3.440
Seanno Shag...	?	Terlambat	0	1	D3 Manajem...	2022	2025	LAKI-LAKI	3.420	3.030	2.890
Novitasari...	?	Tepat Waktu	0.980	0.020	D3 Manajem...	2022	2025	PEREMPUAN	3.740	3.690	3.840
Adinda Sals...	?	Tepat Waktu	0.980	0.020	D3 Manajem...	2022	2025	PEREMPUAN	3.650	3.680	3.340
Chica Anggr...	?	Tepat Waktu	0.980	0.020	D3 Manajem...	2023	2026	PEREMPUAN	3.780	3.660	3.970

Gambar 10. Hasil prediksi pengujian



Gambar 11. Bar column

5. KESIMPULAN DAN SARAN

Metode Decision Tree terbukti efektif dalam memprediksi kelulusan mahasiswa dengan menggunakan data akademik seperti IPS, IPK, tahun masuk, status mahasiswa, dan status bekerja. Struktur pohon keputusan yang dihasilkan memberikan gambaran yang jelas terhadap pola kelulusan mahasiswa. Dari model yang dibangun menggunakan perangkat lunak RapidMiner, diperoleh akurasi prediksi yang cukup tinggi, yaitu mencapai 89–90%, sebagaimana ditunjukkan oleh confusion matrix. Ini menunjukkan bahwa metode yang digunakan mampu mengklasifikasikan mahasiswa yang lulus tepat waktu dan yang terlambat dengan cukup akurat. Variabel

yang paling berpengaruh dalam model adalah IPS semester 2, IPS semester 3, dan status mahasiswa (bekerja atau tidak). Hal ini menandakan bahwa performa akademik pada awal perkuliahan serta status pekerjaan mahasiswa dapat menjadi indikator kuat terhadap prediksi kelulusan. Hasil dari prediksi data uji menunjukkan bahwa model ini juga mampu menggeneralisasi data baru dengan baik, dan dapat dijadikan alat bantu pengambilan keputusan bagi pihak kampus.

Perluasan Dataset untuk hasil yang lebih akurat dan general, sebaiknya dilakukan perluasan jumlah data historis mahasiswa dari berbagai program studi dan tahun ajaran. Semakin banyak data, semakin kuat model yang dapat dibangun. Penambahan Variabel Non-Akademik dalam penelitian selanjutnya, disarankan untuk menambahkan variabel lain seperti motivasi belajar, kehadiran, partisipasi organisasi, atau kondisi sosial-ekonomi yang mungkin juga berpengaruh terhadap kelulusan mahasiswa. Penggunaan Metode Pembandingan dalam pengembangan lebih lanjut, sebaiknya dilakukan perbandingan dengan algoritma lain seperti Naïve Bayes, Random Forest, atau SVM agar dapat diketahui metode mana yang paling optimal untuk konteks data dan institusi yang sama.

DAFTAR PUSTAKA

- [1] M. A. Senubekti and L. A. Puspita Dewi, "PRINSIP KLASIFIKASI DAN DATA MINING DENGAN ALGORITMA C4.5," *NUANSA*, vol. 16, no. 2, pp. 87–93, Jul. 2022, doi: 10.25134/nuansa.v16i2.5834.
- [2] N. Anggraini, E. S. N. Harahap, and T. B. Kurniawan, "Text Mining - Analisis Teks Terkait Isu Vaksinasi COVID-19 (Text Mining - Text Analysis Related to COVID-19 Vaccination Issues)," *IPTEKKOM*, vol. 23, no. 2, pp. 141–153, Dec. 2021, doi: 10.17933/iptekkom.23.2.2021.141-153.
- [3] N. Anggraini and H. Suroyo, "Comparison of Sentiment Analysis against Digital Payment 'T-cash and Go-pay' in Social Media Using Orange Data Mining," *journalisi*, vol. 1, no. 2, pp. 152–163, Sep. 2020, doi: 10.33557/journalisi.v1i2.21.
- [4] N. Khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, "PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAIVE BAYES," *Technologia*, vol. 13, no. 3, p. 207, Aug. 2022, doi: 10.31602/tji.v13i3.7312.
- [5] A. Pratama, R. C. Wihandika, and D. E. Ratnawati, "Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa," 2022.
- [6] R. P. S. Putri and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *khif*, vol. 4, no. 1, pp. 1–7, Jul. 2020, doi: 10.23917/khif.v4i1.5975.
- [7] Eko Rohmawan, "PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE DESICION TREE DAN ARTIFICIAL NEURAL NETWORK," *jurnalatrik*, vol. 20, no. 1, Jan. 2019, Accessed: Mar. 18, 2025. [Online]. Available: <https://journal.binadarma.ac.id/index.php/jurnalatrik/article/view/105>
- [8] K. C. Laudon and J. P. Laudon, *Management information systems: managing the digital firm*, Seventeenth edition, Global edition. Harlow, England London New York Boston: Pearson, 2022.
- [9] O. Feriyanto, A. N. Ilmi, V. Aulia, L. H. Jandriani, S. Safitri, and E. Irmayanti, "PERAN AKUNTANSI TERHADAP PENGAMBILAN KEPUTUSAN BISNIS MELALUI ANALISIS BIG DATA (STUDI LITERATUR)," *BIG DATA*, Sep. 2021.
- [10] U. Felt, R. Fouché, C. A. Miller, L. Smith-Doerr, and Society for Social Studies of Science, Eds., *The handbook of science and technology studies*, Fourth edition. Cambridge, Massachusetts London, England: The MIT Press, 2022. [Online]. Available: https://books.google.co.id/books/about/The_Handbook_of_Science_and_Technology_S.html?id=4GTUDQAAQBAJ&redir_esc=y
- [11] A. Salam and Moh. Sholik, "Implementasi Algoritma Apriori untuk Mencari Asosiasi Barang yang dijual di E-commerce OrderMas," *tc*, vol. 17, no. 2, pp. 158–170, May 2020, doi: 10.33633/tc.v17i2.1656.
- [12] F. A. Hermawati, *Data Mining*, 1st ed. Andi, 2020. [Online]. Available: <https://elibrary.nusamandiri.ac.id/readbook/200002/data-mining.html>
- [13] F. F. Harryanto and S. Hansun, "Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," *Jurnal Teknik Informatika dan Sistem Informasi (JATISI)*, vol. 3, pp. 95–103, Mar. 2020, doi: <https://doi.org/10.35957/jatisi.v3i2.71>.