E-ISSN 2797-4278
P-ISSN 2797-4294

Prediksi Kelulusan Mahasiswa Berdasarkan Data Kunjung dan Peminjaman Buku menggunakan Rapid Miner dengan Metode C.45 dan Random Forest

[1]Romdan Muhamad Ubaidilah, [2]Zulfi Anugerahwati, [3]Imaniar Ikko Mulya Rizky, [4]Sri Lestari
[1,2,3,4]Magister Teknik Informatika, Ilmu Komputer, Institut Informatika dan Bisnis Darmajaya, Indonesia.
[1]romdan.2221210029@mail.darmajaya.ac.id

ABSTRAK

Waktu kelulusan dan ketidakseimbangan antara jumlah pendaftar dan lulusan menjadi perhatian utama perguruan tinggi. Masalah ini menyebabkan penumpukan data mahasiswa yang belum lulus, serta peningkatan kunjungan dan peminjaman buku setiap harinya. Penelitian ini dilakukan untuk mendapatkan pengetahuan menggunakan teknik klasifikasi dengan memproses data guna menemukan pola pada data kelulusan dan transaksi perpustakaan. Pengolahan data menggunakan metode decision tree dengan algoritma C4.5 dan Random Forest. Atribut non-kelas yang digunakan meliputi IP semester 1-4, jumlah kunjungan, pinjam buku dan status kelulusan. Prosesnya meliputi memuat data, membersihkan data, pemilihan fitur, pemodelan C4.5 and Random Forest, pengujian akurasi menggunakan confusion matrix. Percobaannya menunjukkan bahwa kedua metode tersebut memberikan tingkat akurasi prediksi yang cukup tinggi dalam memprediksi kelulusan mahasiswa. Dalam rasio training 90% dan testing 10%, metode C4.5 memiliki akurasi sebesar 94,90%, sementara Random Forest mencapai akurasi sebesar 95,13%. Manfaat dari penelitian ini dapat digunakan oleh IAIN Metro untuk memprediksi masa studi mahasiswa.

Keyword: ***data kunjung, peminjaman buku, rapid miner, c.45, random forest.***

1 PENDAHULUAN

Institut Agama Islam Negeri Metro selanjutnya disebut IAIN Metro, merupakan perguruan tinggi keagamaan yang berkedudukan di Kota Metro dan berdiri pada tanggal 1 Agustus 2016 berdasarkan Peraturan Presiden Nomor 71 Tahun 2016 sebagai perubahan bentuk dari Sekolah Tinggi Agama Islam Negeri Jurai Siwo Metro (Presiden Republik Indonesia, n.d.) Jumlah penerimaan mahasiswa baru IAIN Metro meningkat setiap tahun, namun mahasiswa yang lulus jumlahnya tidak sebanding dengan kuota penerimaan. Semakin meningkatnya jumlah mahasiswa baru, secara otomatis jumlah data yang dihasilkan dan disimpan dalam database bertambah banyak. Jumlah data transaksi peminjaman buku dan data kunjungan perpustakaan IAIN Metro semakin bertambah setiap hari. Berdasarkan data unit perpustakaan IAIN Metro rata-rata kunjungan mahasiswa setiap hari adalah sebanyak 74 kunjungan dengan total transaksi kunjungan mahasiswa dari angkatan tahun 2017-2019 sebanyak 91.302 transaksi. Sedangkan jumlah data transaksi peminjaman buku sebanyak 4.831 mahasiswa dengan data peminjaman sebanyak 58.376 transaksi. Data perpustakaan dapat digunakan untuk memprediksi kelulusan mahasiswa IAIN Metro, sehingga dapat dilakukan evaluasi efektifitas strategi pendidikan yang dimiliki dan upaya – upaya yang diperlukan untuk meningkatkan keberhasilan akademik mahasiswa sehingga dapat lulus tepat waktu. Prediksi kelulusan penting dilakukan untuk mengetahui besarnya pengaruh perpustakaan terhadap proses penyelesaian studi mahasiswa, karena data transaksi perpustakaan menyimpan informasi tingkat keterlibatan dan keaktifan mahasiswa dalam bidang akademik. Data ini bermanfaat bagi IAIN Metro untuk memprediksi kelulusan mahasiswa dengan menggunakan teknik data mining.

Penelitian yang berjudul Prediksi Kelulusan Mahasiswa Berdasarkan Data Berkunjung dan Pinjam Buku di Perpustakaan Menggunakan Metode C4.5, Penelitian ini menggunakan teknik klasifikasi untuk menemukan pola dalam kelulusan mahasiswa, data kunjungan dan peminjaman buku perpustakaan. Data diproses untuk memprediksi kategori yang belum diketahui. Teknik klasifikasi yang digunakan adalah decision tree dengan algoritma C4.5. Penelitian ini menggunakan atribut berupa usia, toefl, IPK, kunjungan, peminjaman buku, dan atribut kelas adalah status kelulusan. Tahapannya meliputi pemrosesan data, pengujian akurasi, pembuatan aturan, prediksi, dan pengujian akurasi menggunakan confusion matrix. Penelitian ini menghasilkan aturan untuk memprediksi kelulusan dengan tepat waktu menggunakan pohon keputusan. Dalam pengujian dengan 325 data, ditemukan 53 aturan dengan tingkat akurasi 94%, presisi 78%, dan recall 53%. Manfaat penelitian ini adalah sebagai sumber pengetahuan yang berguna bagi program studi dalam memprediksi masa studi dan kelulusan mahasiswa (Nurhanifa,1, Tedy Setiadib, 2020).

Pada penelitian yang berjudul Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta), penelitian ini bertujuan untuk menganalisis mahasiswa yang lulus tepat waktu maupun tidak tepat waktu, dengan data mining menggunakan metode Random Forest dan Gradient Boosting untuk mengetahui tingkat akurasi dan menentukan mana model klasifikasi yang terbaik pada ketepatan lulus mahasiswa. Analisis menggunakan data mahasiswa S1 FIK UPN Veteran Jakarta angkatan 2015 - 2017. Hasil uji coba sampel pada 590 data, algoritma random forest 10 k-fold mendapatkan akurasi 82,64% dan pada gradient boosting 3 k-fold mendapatkan akurasi 79,66%. Hasil penelitian ini digunakan sebagai dasar untuk mengambil keputusan dalam menentukan kebijakan oleh fakultas (Labib et al., 2023).

E-ISSN 2797-4278
P-ISSN 2797-4294

Dalam sebuah penelitian yang berjudul "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes", ditemukan bahwa kinerja model Naive Bayes lebih baik dibandingkan dengan model K-Nearest Neighbor dan Decision Tree. Hasil penelitian tersebut menunjukkan bahwa dari 35 data uji yang digunakan, model Naive Bayes memiliki tingkat akurasi sebesar 89% dan presisi sebesar 88%. Sementara itu, model K-Nearest Neighbor memiliki tingkat akurasi sebesar 77% dan presisi sebesar 76%, sedangkan model Decision Tree memiliki tingkat akurasi sebesar 74% dan presisi sebesar 84%.

Sebuah studi klasifikasi menggunakan tiga model dengan kasus studi pada Program Studi Teknik Informatika di Universitas Islam Madura merekomendasikan penggunaan model Naive Bayes dan K-NN karena performanya mendekati titik acuan yang ditetapkan (Hozairi et al., 2021). Jurnal tersebut berjudul *Predicting student graduation using library circulation data: a comparative study of machine learning algorithms*, hasil dari penelitian ini untuk menyajikan studi perbandingan berbagai teknik pemrosesan data yang baru-baru ini digunakan, algoritma klasifikasi, dampaknya pada dataset, serta hasil atribut prediksi secara jelas dan ringkas. Makalah ini juga mengidentifikasi atribut terbaik yang akan membantu dalam memprediksi kinerja siswa dengan cara yang efisien (Mounika & Persis, 2019).

.Selanjutnya penelitian berjudul *Predicting Students' Academic Performance Decision Tree and Neural Network*, hasil penelitian ini menyebutkan bahwa berbagai dataset siswa memberikan hasil yang berbeda dengan atribut yang berbeda pula. Hal ini menjadi alasan mengapa hasilnya bervariasi dengan berbagai ukuran evaluasi seperti akurasi, presisi, dan rata-rata geometrik. pendekatan dan algoritma pemrosesan data memberikan hasil yang bervariasi tergantung pada dataset dan atribut variabel yang digunakan untuk prediksi. jika kita menggunakan algoritma ADTree, JRip, Ridor, logistic regression dan neural network sesuai dengan kebutuhan kita, algoritma - algoritma ini memberikan hasil yang sangat akurat untuk prediksi masa depan dan membantu dalam perbaikan sistem pendidikan (Junshuai, 2019).

Proses menganalisa data penelitian ini digunakan alat bantu berupa Rapidminer. Rapidminer merupakan sebuah perangkat lunak open source yang dapat diakses oleh siapa saja dan bersifat terbuka. Berbagai teknik seperti teknik deskriptif dan prediksi digunakan pada Rapidminer dan dalam pengoperasiannya Rapidminer menggunakan bahasa pemrograman Java(Setio et al., 2020).

Berdasarkan tinjauan penelitian sebelumnya, peneliti melakukan studi mengenai prediksi kelulusan mahasiswa berdasar pada data kunjungan dan peminjaman buku menggunakan metode C4.5 dan Random Forest. Hasil penelitian ini dapat memberikan kontribusi bagi Institut Agama Islam Negeri Metro dalam memahami pengaruh data kunjungan dan peminjaman buku terhadap kelulusan mahasiswa.

2 LITERATUR REVIEW

2.1 Kajian Terdahulu

Berikut ini adalah rangkuman penelitian terkait prediksi kelulusan mahasiswa menggunakan data kunjungan dan data peminjaman buku pada perpustakaan dengan menggunakan metode C4.5 dan Random Forest:

No	Penelitian	Tujuan Penelitian	Metodologi	Temuan
1	(Putri & Waspada, 2018)	Memprediksi kelulusan mahasiswa berdasarkan kunjungan ke perpustakaan dengan menggunakan metode C4.5	Pengumpulan data, preprocessing, pembuatan model dengan algoritma C4.5, pengujian model	Algoritma C4.5, mampu memprediksi kelulusan dengan nilai rata-rata precision 63.93 %, recall 60.73 %, dan akurasi 60.52 %
2	(Data et al., 2022)	Memprediksi kelulusan mahasiswa tepat waktu menggunakan metode Random Forest	Pengumpulan data, preprocessing, pembuatan model dengan algoritma Random Forest, pengujian model	Metode Random Forest mampu memprediksi Kelulusan mahasiswa dengan akurasi 90,74 %
3	(Nurhanifa ,1, Tedy Setiadib, 2020)	Prediksi Kelulusan Mahasiswa Berdasarkan Data Berkunjung dan Pinjam Buku di	Pengumpulan data, preprocessing, pembuatan model dengan Metode C4.5	Penelitian ini menghasilkan rule untuk memprediksi

E-ISSN 2797-4278
P-ISSN 2797-4294

	Perpustakaan Menggunakan Metode C4.5	tepat waktu kelulusan dengan pohon keputusan. Hasil pengujian menggunakan 325 data menghasilkan 53 rule dengan tingkat akurasi 94%, presisi 78% dan recall 53%.

Dari tabel tersebut dapat dilihat bahwa ketiga penelitian menghasilkan model prediksi kelulusan mahasiswa yang memiliki akurasi yang cukup baik. Namun, masing-masing penelitian menggunakan metode yang berbeda-beda, yaitu C4.5 dan Random Forest. Oleh karena itu, penelitian ini akan menggabungkan ketiga metode tersebut dalam satu model untuk meningkatkan akurasi prediksi kelulusan mahasiswa.

2.3 Data Mining

Data mining adalah proses yang dilakukan untuk menemukan pola dan tren baru dengan tingkat signifikansi melalui analisis data yang berjumlah besar atau volumin dalam suatu repositori. Teknik pengenalan data yang efektif melibatkan penerapan metode statistik dan matematika, Salah satu metode yang digunakan adalah algoritma C45, yang merupakan metode pengklasifikasian berdasarkan prinsip probabilitas dan statistik. (Yuli Mardi, 2019).

2.4 Tahapan Data Mining

Pada tahun 1996, para analis yang mewakili DaimlerChrysler, SPSS, dan NCR mengembangkan Proses Standar Lintas-Industri untuk Penambangan Data (CRISP-DM). CRISP-DM menyediakan suatu proses standar yang tidak dimiliki oleh satu entitas tertentu dan tersedia secara bebas untuk mengintegrasikan penambangan data ke dalam strategi pemecahan masalah umum dari bisnis atau unit penelitian (Hozairi et al., 2021). Menurut CRISP-DM, proyek penambangan data memiliki siklus hidup yang terdiri dari enam step sebagai berikut:

1. Step Pemahaman Bisnis: Di fase ini, langkah pertama adalah dengan jelas menetapkan tujuan dan persyaratan proyek dalam konteks unit bisnis atau penelitian secara keseluruhan. Setelah itu, tujuan dan batasan tersebut diartikan menjadi definisi masalah penambangan data, serta merancang strategi awal untuk mencapai tujuan yang telah ditetapkan.
2. Fase Pemahaman Data: Pada fase ini, data dikumpulkan, kualitasnya dievaluasi, dan dilakukan analisis eksploratif terhadap data tersebut. Tujuannya adalah untuk mendapatkan pemahaman yang lebih baik tentang data dan menemukan wawasan awal yang dapat digunakan pada tahapan selanjutnya.
3. Fase Persiapan Data: Pada fase ini, data mentah awal disiapkan dan diubah menjadi satu set data akhir yang akan digunakan pada semua fase berikutnya. Selain itu, juga dilakukan seleksi kasus dan variabel yang akan dianalisis, serta melakukan transformasi pada variabel tertentu agar siap digunakan oleh alat pemodelan.
4. Fase Pemodelan: Di fase ini, kita memilih teknik pemodelan yang sesuai lalu menerapkannya. Selain itu juga dilakukan kalibrasi pengaturan model guna mengoptimalkan hasil dari model tersebut.
5. Fase Evaluasi: fase ini melibatkan evaluasi satu atau lebih model hasil dari fase pemodelan untuk mengukur kualitas dan efektivitasnya sebelum digunakan secara operasional.. Tujuannya adalah untuk menentukan apakah model-model tersebut berhasil mencapai tujuan yang telah ditetapkan sejak awal.

2.5 Predikasi

Prediksi adalah suatu tindakan memperkirakan atau memproyeksikan suatu kejadian, hasil, atau peristiwa yang akan terjadi di masa depan (Lusiana & Yuliarty, 2020). Pengertian Prediksi sama dengan ramalan atau perkiraan. Menurut Kamus Besar Bahasa Indonesia, prediksi adalah tindakan memperkirakan sesuatu yang akan terjadi di masa depan berdasarkan fakta atau bukti yang ada (Kafil, 2019). Peramalan merupakan prosedur untuk menghasilkan informasi faktual tentang situasi sosial di masa depan berdasarkan informasi yang sudah ada mengenai masalah kebijakan. Terdapat tiga bentuk utama dari peramalan yaitu proyeksi, prediksi, dan perkiraan.

Proyeksi merupakan jenis peramalan yang didasarkan pada ekstrapolasi tren masa lalu dan saat ini ke masa depan. Proyeksi menggunakan pertanyaan khusus dengan argumen yang diperoleh dari metode tertentu serta kasus-kasus sebelumnya.

Prediksi adalah bentuk peramalan yang bergantung pada asumsi teoritis yang spesifik. Asumsi tersebut dapat berupa prinsip-prinsip teoritis seperti hukum penurunan nilai uang, proposisi-proposisi teoretis seperti proposisi bahwa kotak-kotaknya masyarakat sipil disebabkan oleh kesenjangan antara harapan dan kemampuan, atau Contoh

E-ISSN 2797-4278
P-ISSN 2797-4294

analogi lainnya adalah perbandingan antara pertumbuhan organisasi pemerintahan dan pertumbuhan organisme biologis.

Perkiraan (conjecture) adalah jenis peramalan yang berdasar pada penilaian informatif atau keahlian dari pakar mengenai situasi masyarakat di masa depan. Tujuan dari peramalan kebijakan adalah untuk memperoleh informasi tentang perkembangan di masa depan yang akan mempengaruhi implementasi kebijakan dan dampaknya.

2.6 Algoritma C4.5

Algoritma C4.5 adalah suatu algoritma yang digunakan untuk membangun pohon keputusan. Dalam algoritma ini, pohon keputusan dibentuk berdasarkan parameter-parameter tertentu yang dapat membantu dalam pengambilan keputusan. Beberapa elemen penting dalam pembentukan pohon keputusan tersebut adalah Root (akar), Node (simpul), dan Relationship (hubungan antar simpul).

Dalam penentuan akar pohon, atribut dengan nilai gain tertinggi akan dipilih sebagai atribut yang menjadi akar dari pohon tersebut. Penghitungan gain menggunakan persamaan berikut (Prasetyo et al., 2020):

$$Gain(S, A) = Entropy(S) - \sum |Si| / |S| * Entropy(Si)\ n\ i=1 \quad (1)$$

Keterangan:
S : Himpunan Kasus
A : Atribut
n : Jumlah partisi atribut A
|Si| : Jumlah kasus pada partisi ke-i
|S| : Jumlah kasus dalam S

Sementara itu, untuk mencari nilai entropi digunakan persamaan berikut:

$$Entropy(S) = \sum - pi * log2\ pi\ n\ i=1 \quad (2)$$

Keterangan:
S : Himpunan Kasus
n : Jumlah partisi S
pi : proporsi dari Si terhadap S

Gain ratio adalah sebuah modifikasi dari information gain yang digunakan dalam algoritma C4.5 untuk mengurangi bias atribut. Dalam pemilihan akar pohon keputusan menggunakan gain ratio, kita perlu mencari nilai split info terlebih dahulu.

Split info merupakan ukuran dari homogenitas data pada suatu simpul berdasarkan atribut yang sedang dievaluasi. Gain ratio dan split info dinyatakan dalam persamaan berikut:

$$Split\ Info\ (S, A) = - \sum Si\ S\ log2\ Si\ S\ n\ i=1 \quad (3)$$

Keterangan:
S : Himpunan Kasus
A : Atribut
Si : Jumlah sampel untuk atribut i

$$Gain\ Ratio\ (a) = gain(a)\ split(a) \quad (4)$$

Keterangan:
a : Atribut
Gain (a) : Nilai gain pada atribut a
Split (a) : Nilai split pada atribut a

2.7 Algoritma Random Forest

Random forest merupakan teknik yang dikembangkan dari Classification dan Regression Tree (CART). Teknik ini mengaplikasikan random feature selection dan teknik bagging (bootstrap aggregating) dalam pengembangannya. Random forest adalah sebuah classifier yang terdiri dari kumpulan pohon klasifikasi {h(x, Sb), b = 1, ..., B}, dengan setiap pohon memiliki vektor acak Sb yang tidak saling berkaitan dan didistribusikan secara identik pada setiap pohon. Setiap pohon memberikan suara atau vote untuk kelas dengan frekuensi tertinggi pada input x.

Ensemble classifier h1(x), h2(x), ..., hb(x) diberikan dengan menggunakan data training yang diambil secara acak berdasarkan distribusi vektor acak X, Y. Hal ini didefinisikan dalam Persamaan 1 (Naufalrifqi, 2022).$mg(X, Y) = avbI(hb\ (X) = Y) -$ max $j \neq Y\ avb(hb\ (X) = j)$ (1)

Keterangan:
$I(\cdot)$ merupakan fungsi indikator dan avb merupakan hasil rata – rata dari fungsi indikator,
$hb\ (X) = Y$ adalah hasil dari prediksi Y
$hb\ (X) = j$ adalah hasil dari prediksi j.

Margin yang ditentukan digunakan sebagai ukuran seberapa besar nilai rata – rata pada vote X, Y untuk kelas yang tepat agar dapat melebihi rata – rata vote kelas lainnya. Karena, semakin besar jarak margin maka semakin akurat nilainya. Setelah itu, generalization error diberikan Persamaan 2 (Naufalrifqi, 2022).

$$395\ PE * = PX,Y(mg(X, Y) < 0 \quad (2)$$

Keterangan
$PE *$ merupakan geleralization error
PX,Y sebagai indikasi untuk probabilitas yang melebihi ruang X, Y.

RF, $hb\ (X) = h(X, Y\ b\ )$.

Untuk pohon dengan jumlah yang banyak, terdapat Tree Structurer dan Strong Law of Large Numbers. Semakin tingginya jumlah pohon, maka hampir dalam semua barisan S 1 , … akan menyebabkan nilai $PE *$ menjadi konvergen ke Persamaan 3 (Naufalrifqi, 2022) .

E-ISSN 2797-4278
P-ISSN 2797-4294

$PX,Y\ (PS(h(X, S\ b\ ) = Y) - \max\ j \neq Y\ PS\ (h(X, S\ b\ ) = j) < 0)$ (3)

2.9 Rapid Miner

RapidMiner adalah sebuah platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan dengan nama yang sama. Platform ini menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran mendalam, penambangan teks, dan analisis prediktif (Sudarsono et al., 2021). Dalam konteks penggunaannya, RapidMiner digunakan untuk memproses raw data atau data mentah dari data kunjung dan peminjaman buku. Tujuannya adalah untuk menghasilkan dataset baru yang lebih bersih dan siap digunakan.

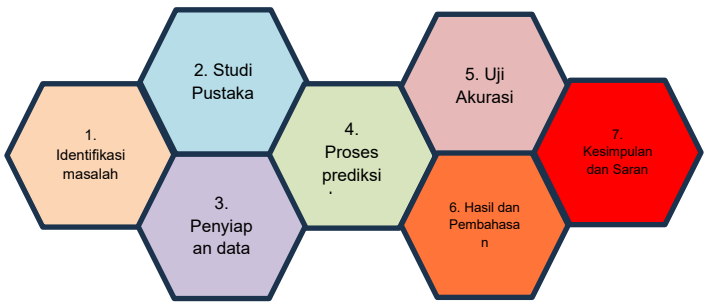
Karena jumlah data dalam database kunjungan dan peminjaman buku sangat besar, mengevaluasi kinerja secara menyeluruh menjadi sulit dilakukan. Oleh karena itu, teknologi seperti RapidMiner diperlukan sebagai alat bantu dalam proses tersebut (Mardalius, 2018).

Dengan menggunakan RapidMiner, penulis dapat memperoleh insight berharga dari data kunjung dan peminjaman buku serta melakukan analisis yang lebih efisien. Hal ini akan membantu dalam pengambilan keputusan yang lebih baik di bidang usaha.

3 METODOLOGI

3.1 Tahapan Penelitian

Tujuan dari penelitian ini adalah melakukan analisa perbandingan metode C.45 dan random forest dalam memprediksi kelulusan mahasiswa IAIN Metro Tahun Akademik 2017 – 2019. Dalam menganalisa dataset penelitian, peneliti menggunakan alat bantu yaitu aplikasi Rapid Miner. Adapun tahapan proses pada penelitian ini dapat dilihat pada Gambar. 3.1



Sumber: (Amalia, 2018)
Gambar 3.1 Tahapan Penelitian

Tahapan pertama dari penelitian ini adalah peneliti melakukan identifikasi masalah dan perumusan masalah. Selanjutnya melakukan studi pustaka untuk mempelajari dan menganalisis penelitian sebelumnya mengenai kelulusan mahasiswa termasuk algoritma yang digunakan untuk prediksi. Tahapan selanjutnya melakukan penyiapan dataset sebagai sumber data yang akan dianalisis. Dataset dibagi menjadi 2 bagian yakni data train dan data test. Setelah dataset siap, selanjutnya melakukan proses prediksi menggunakan aplikasi rapid miner. Proses uji akurasi metode C.45 dan Random Forest dilakukan setelah proses prediksi selesai. Tahapan keenam adalah hasil dan pembahasan penelitian yang telah dilakukan, berisi akurasi. Terakhir kesimpulan dan saran.

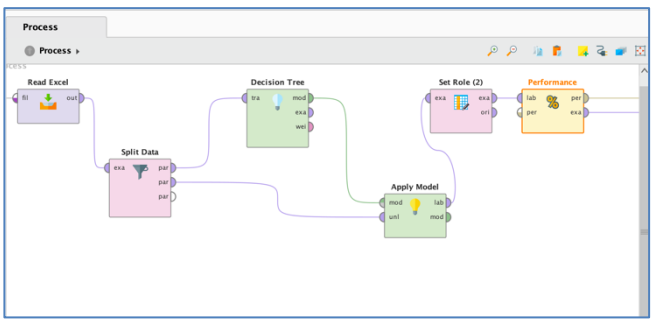
3.2 Data

Dalam penelitian ini yang menjadi objek penelitian adalah data kelulusan dan data transaksi perpustakaan IAIN Metro. Berdasarkan data unit perpustakaan IAIN Metro rata-rata kunjungan mahasiswa setiap hari adalah sebanyak 74 kunjungan dengan total transaksi kunjungan dari angkatan tahun 2017-2019 sebanyak 91.302 transaksi. Sedangkan jumlah data transaksi peminjaman buku sebanyak 4.831 mahasiswa dan jumlah data peminjaman sebanyak 58.376 transaksi.

4 HASIL DAN PEMBAHASAN

Hasil percobaan menunjukkan bahwa kedua metode, yaitu C.45 dan Random Forest, memberikan tingkat akurasi prediksi yang cukup tinggi dalam memprediksi kelulusan mahasiswa. Pada rasio data training 90% dan data testing 10%, C.45 menghasilkan akurasi sebesar 94,90%, sedangkan Random Forest mencapai akurasi sebesar 95,13%. Peningkatan sedikit terjadi saat rasio diubah menjadi 80% training - 20% testing dengan akurasi C.45 sebesar 94,04% dan Random Forest sebesar 94,53%. Selanjutnya, peningkatan perbandingan train-test ratio menjadi 70% -30% menunjukkan sedikit penurunan kinerja pada kedua metode tersebut dengan akurasi C.45 sebesar 93,65% dan Random Forest sebesar 93.73%. Meskipun ada penurunan kecil dalam performa prediksi pada rasio ini, tetapi masih dapat diterima secara praktis. Analisis hasil juga melibatkan pembahasan faktor-faktor penting yang berkontribusi dalam prediksi kelulusan mahasiswa menggunakan kedua metode ini. Faktor seperti kunjungan ke perpustakaan dan peminjaman buku dapat memberikan informasi yang berharga dalam memprediksi kelulusan mahasiswa. Pada tahap ini, dilakukan pemuatan data awal mengenai IAIN Metro dari angkatan tahun 2017 hingga 2019, serta data transaksi perpustakaan dari tahun 2017 hingga 2019 menggunakan file "DATA SIAP - Normalisasi.xls" dengan jumlah data sebanyak 4.984 mahasiswa. Tahap pembersihan data atau cleaning dilakukan untuk menghapus data yang tidak lengkap atau kosong, sebanyak 681 data mahasiswa dihapus karena tidak memiliki data kunjungan dan peminjaman buku. Sehingga diperoleh jumlah data sebanyak 4.303 dengan data kunjungan sebanyak 91.302 transaksi dan 81.637 transaksi peminjaman buku. Seleksi Atribut Pada tahap seleksi, hanya atribut-atribut yang dianggap berpengaruh dalam penelitian yang dipertahankan dalam dataset. Atribut yang digunakan meliputi IPS (Indeks Prestasi Sementara) semester 1-4, kunjungan, pinjam buku, dan status.

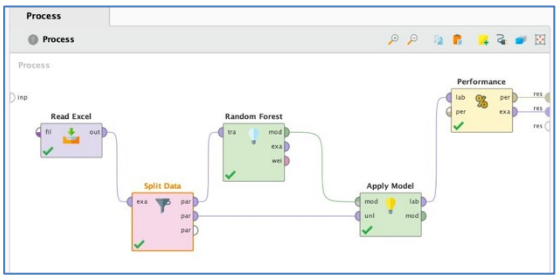
Pemodelan Algoritma C4.5



E-ISSN 2797-4278
P-ISSN 2797-4294

Dalam proses pemodelan C4.5, langkah pertama adalah memasukkan data sampel yang telah disiapkan untuk perhitungan. Data tersebut mencakup atribut-atribut, jumlah total data, jumlah data yang telah diklasifikasikan berdasarkan target yang ditentukan (misalnya lulus tepat waktu atau tidak), serta kolom nilai Entropi dan Gain.Selanjutnya, algoritma C4.5 diterapkan dengan menghitung nilai Entropi dan Gain pada setiap atribut untuk membentuk struktur pohon keputusan (Tree). Pohon keputusan ini merupakan aturan klasifikasi yang akan digunakan dalam proses pengujian (testing). Setelah proses perhitungan selesai dilakukan, akan dihasilkan sebuah aturan atau rule sebagai hasil akhir dari pemodelan.

Pemodelan Algoritma Random Forest



Gambar 2. Pemodelan Random Forest

Dalam proses pemodelan algoritma Random Forest, langkah pertama adalah memasukkan data sampel yang telah disiapkan untuk perhitungan. Data tersebut mencakup atribut-atribut, jumlah total data, dan hasil klasifikasi berdasarkan target yang ditentukan (misalnya lulus tepat waktu atau tidak). Selanjutnya, algoritma Random Forest diterapkan dengan membangun sejumlah pohon keputusan secara acak dari subset data yang diambil secara bootstrap. Setiap pohon keputusan dalam ensemble model ini menggunakan metode seperti Decision Tree atau C4.5. Kemudian, pada setiap pohon individual dalam Random Forest, dilakukan perhitungan nilai Entropi dan Gain pada tiap atribut untuk membentuk struktur pohon keputusan. Proses ini dilakukan secara independen pada setiap pohon. Setelah semua pohon selesai dibangun, prediksi baru dapat dilakukan melalui proses voting: Untuk masalah klasifikasi: Prediksi dari masing-masing tree dihitung dan mayoritas suara menjadi hasil akhir prediksi. Untuk masalah regresi: Prediksi rata-rata dari semua prediksi individu tree digunakan sebagai nilai prediksi akhir. Dengan demikian, hasil akhir dari pemodelan algoritma Random Forest adalah kombinasi aturan-aturan klasifikasi yang dihasilkan oleh setiap decision tree dalam ensemble model tersebut. Testing atau pengujian akurasi.

Testing atau pengujian akurasi

Pada fase pengujian ini, langkah yang dilakukan adalah menginputkan data uji atau data prediksi. Atribut yang digunakan dalam proses pengujian harus konsisten dengan atribut yang telah digunakan dalam proses pelatihan sebelumnya. Setiap atribut data akan dibandingkan dengan aturan yang terbentuk dari perhitungan pada data pelatihan sebelumnya.

Selanjutnya, data tersebut akan diklasifikasikan berdasarkan target yang ingin diketahui, yaitu apakah data tersebut mencerminkan mahasiswa dan transaksi perpustakaan serta apakah mereka lulus tepat waktu atau tidak.Data training memiliki jumlah sampel sebanyak 4.303, sementara data testing memiliki jumlah sampel sebanyak 403.

Dengan menggunakan metode ini, tujuannya adalah menguji performa model dan melihat bagaimana model dapat mengklasifikasikan dataset baru secara akurat.

accuracy: 94.90%

	true LULUS	true AKTIF	class precision
pred. LULUS	129	21	86.00%
pred. AKTIF	1	280	99.64%
class recall	99.23%	93.02%	

Gambar 3. Akurasi dengan model C45

accuracy: 95.13%

	true LULUS	true AKTIF	class precision
pred. LULUS	129	20	86.58%
pred. AKTIF	1	281	99.65%
class recall	99.23%	93.36%	

Gambar 4. Akurasi dengan model C45

Tabel dibawah ini menampilkan perbandingan akurasi metode C.45 dan Random Forest

No.	Data Latih	Algoritma	Akurasi	Software
1.	403	C.45	94.90%	Rapid
		Random Forest	95.13%	Miner
2.	860	C.45	94.04%	
		Random Forest	93,65%	
3.	1.291	C.45	94.53%	
		Random Forest	93.73%	

Tabel 1. Perbandingan Akurasi

5 CONCLUSION

Berdasarkan analisis hasil percobaan dengan Rapid Miner menggunakan metode C.45 dan Random Forest, dapat disimpulkan bahwa kedua metode tersebut efektif dalam memprediksi kelulusan mahasiswa berdasarkan data kunjung dan peminjaman buku. Meskipun terdapat sedikit perbedaan kinerja antara dua metode tersebut, baik C.45 maupun Random Forest menunjukkan tingkat akurasi yang cukup tinggi pada rasio data training dengan data testing yang digunakan dalam eksperimen. Oleh karena itu, kedua metode ini memiliki potensi untuk diaplikasikan pada skala yang lebih besar dalam konteks prediksi kelulusan mahasiswa. Sebagai saran untuk pengembangan selanjutnya, dapat melibatkan variabel lain atau melakukan pemodelan kombinasi dari beberapa metode prediksi lainnya guna meningkatkan performa model secara keseluruhan.

REFERENCES

Amalia, H. (2018). Perbandingan Metode Data Mining Svm Dan Nn Untuk Klasifikasi Penyakit Ginjal Kronis. Maret, 14(1), 1. www.bsi.ac.id

E-ISSN 2797-4278
P-ISSN 2797-4294

Data, I., Untuk, M., Kinerja, M., Karyawan, K., Metode, M., & Linier, R. (2022). *Implementasi Data Mining Untuk Memprediksi Kinerja*. *02*(1), 127–135.

Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, *6*(2), 133. https://doi.org/10.21107/nero.v6i2.237

Junshuai, F. (2019). Predicting Students' Academic Performance with Decision and Neural Network. *Αγαη*, *8*(5), 55.

Kafil, M. (2019). Penerapan Metode K-Nearest Neighbors. *Jurnal Mahasiswa Teknik Informatika (JATI)*, *3*(2), 59–66.

Labib, M., Zaidiah, A., & Yulistiawan, B. S. (2023). *Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta)*. 155–166.

Lusiana, A., & Yuliarty, P. (2020). *PENERAPAN METODE PERAMALAN (FORECASTING) PADA PERMINTAAN ATAP di PT X*.

Mardalius, M. (2018). Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means. *Jurteksi*, *4*(2), 123–132. https://doi.org/10.33330/jurteksi.v4i2.36

Mounika, B., & Persis, V. (2019). A Comparative Study of Machine Learning Algorithms for Student Academic Performance. *International Journal of Computer Sciences and Engineering*, *7*(4), 721–725. https://doi.org/10.26438/ijcse/v7i4.721725

Naufalrifqi, S. (2022). *Optimasi Random Forest Untuk Diagnosis Penyakit Ginjal Kronik Dengan Menggunakan Particle Swarm Optimization*. 393–400.

Nurhanifa,1, Tedy Setiadib, 2. (2020). *Prediksi Kelulusan Mahasiswa Berdasarkan Data Berkunjung dan Pinjam Buku di Perpustakaan Menggunakan Metode C4.5*. *8*(2), 24–33.

Prasetyo, E., Prasetiyo, B., & Korespondensi, P. (2020). *Increased Classification Accuracy C4.5 Algorithm Using Bagging Techniques in Diagnosing Heart Disease*. *7*(5), 1035–1040. https://doi.org/10.25126/jtiik.202072379

Presiden Republik Indonesia. (n.d.). *Peraturan Presiden Nomor 71 Tahun 2016.pdf*.

Putri, R. P. S., & Waspada, I. (2018). Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika. *Khazanah Informatika : Jurnal Ilmu Komputer Dan Informatika*, *4*(1), 1–7. https://doi.org/10.23917/khif.v4i1.5975

Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE - Journal of Business and Audit Information Systems*, *4*(1), 13–21. https://doi.org/10.30813/jbase.v4i1.2729

Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . Jurnal Edik Informatika. *Jurnal Edik Informatika*, *2*(2), 213–219.