Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

A Study Concentration Selection With a C4.5 Algorithm, KNN, and Naive Beyes

Muhammad Busyro*[1], Tri Astuti[2], Deuis Nur Astrida[3], Primandani Arsi[4], Pungkas Subarkah[5]

[1]Informatics, Faculty of Computer Science, Universitas Amikom Purwokerto, Indonesia

Email: [1]busyro0203@gmail.com

Received : Nov 27, 2024; Revised : May 24, 2025; Accepted : May 20, 2025; Published : Aug 18, 2025

Abstract

The course of concentration is a crucial aspect for students at the university amikom purwokerto.This decision doesn't just affect their academic journey., but also determine their readiness in the face of the working world.Various factors that affect the concentration selection, the challenges that students face, as well as solutions to help them choose concentrations that fit their interests and career goals.There are still many students who have been confused in deciding which courses best fit their interests and career goals..This confusion is often caused by a lack of adequate information and proper guidance. This study attempts to analyze the lecture amikom purwokerto concentration of students in the universities of the use of the method c4.5 algorithm 3, k-neareset naighbors and naïve beyes. Academic student data used as the basis analysis to determine the dominance in the lecture concentration.Of the result of the research uses phon 60,24 % decision is, there are using k-neareset naighbors 75.36 % and use naïve beyes 100,00 % there are, the prediction could be the basis for deciding the lecture the concentration by mainstream student.The result is expected to help the university in recommended it to students study concentration related to the election.

Keywords: *Course Selection, Decision Tree, K-Neareset Naighbors, Naïve Beyes.*

1. PENDAHULUAN

Pemilihan mata kuliah konsentrasi merupakan tantangan signifikan bagi mahasiswa Universitas Amikom Purwokerto. Kebingungan dan kurangnya informasi seringkali berdampak negatif pada prestasi akademik dan kesiapan kerja. Penelitian sebelumnya menunjukkan ketidaksesuaian minat, bakat, dan kemampuan dengan pilihan konsentrasi sebagai faktor utama kesulitan pengambilan keputusan [1]. Hal ini berimplikasi pada penundaan penyelesaian tugas akhir dan penurunan motivasi belajar. Oleh karena itu, diperlukan solusi yang efektif untuk membantu mahasiswa memilih konsentrasi yang tepat.[2]

Alih-alih mengandalkan intuisi atau informasi yang terbatas, penelitian ini mengusulkan pendekatan berbasis algoritma klasifikasi untuk memprediksi kesesuaian mahasiswa dengan pilihan konsentrasi. Pendekatan ini memanfaatkan data mahasiswa yang tersedia untuk mengidentifikasi pola dan hubungan antara karakteristik mahasiswa dengan pilihan konsentrasi yang tepat (Anggraeni & Christy, 2019).

Penelitian ini akan menganalisis dan membandingkan tiga algoritma klasifikasi, yaitu C4.5, K-Nearest Neighbors (KNN), dan Naive Bayes, untuk menentukan model yang paling akurat dan efektif dalam konteks data mahasiswa Universitas Amikom Purwokerto. Meskipun penelitian sebelumnya telah mengeksplorasi pemilihan mata kuliah dengan menggunakan algoritma individu seperti C4.5, K-

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

Nearest Neighbors, dan Naive Bayes secara terpisah, masih terbatas penelitian yang membandingkan performa beberapa model secara bersamaan untuk mengidentifikasi metode yang paling akurat dan efektif dalam memberikan panduan bagi mahasiswa di konteks Universitas Amikom Purwokerto[4]. Penelitian ini mengisi celah tersebut dengan membandingkan tiga algoritma klasifikasi—C4.5, K-Nearest Neighbors, dan Naive Bayes—untuk menentukan model mana yang paling tepat dalam memprediksi pilihan konsentrasi berdasarkan data mahasiswa.

Penelitian ini bertujuan untuk membandingkan tiga algoritma (Decision Tree, KNN, dan Naïve Bayes) untuk menentukan model yang paling efektif untuk panduan pemilihan mata kuliah.

2. METODOLOGI PENELITIAN

Metode dari *algoritma C.45* merupakan sebuh metode algoritma yang digunakan untuk membuat pohon keputusan.Menggunakan metode ini, kita dapat mengubah fakta yang sangat signifikan menjadi pohon keputusan yang mewakili situasi yang sebenarnya. Dalam penelitian ini penggunaan *algoritma C.45* untuk menentukan keputusan pemilihan mata kuliah kinsentrasi yaitu Pemrograman dan Multimedia. Di samping itu, C4.5 yang digunakan untuk membuat sebuah putusan dalam bentuk putusan pohon dengan kriteria menulis yang berfungsi sebagai dasar untuk pembangunan putusan (Sinambela, 2016).

Penelitian ini menggunakan algoritma C4.5 karena kemampuannya dalam menghasilkan pohon keputusan yang mudah diinterpretasi, sehingga memudahkan pemahaman terhadap faktor-faktor yang mempengaruhi [sebutkan variabel target]. Untuk memastikan akurasi model, dilakukan validasi menggunakan metode k-fold cross-validation dengan nilai k = 10, yang membagi data menjadi 10 bagian, melatih model pada 9 bagian, dan menguji pada 1 bagian, proses ini diulang 10 kali untuk mendapatkan estimasi performa yang lebih robust.

Interpretasi pohon keputusan yang dihasilkan mudah dibaca dan dipahami, sehingga memudahkan interpretasi hasil dan identifikasi faktor-faktor yang berpengaruh.Menangani data kategorikal dan nusmerik: C4.5 dapat menangani data dengan tipe yang berbeda. Robust terhadap noise, C4.5 relatif tahan terhadap data yang noisy atau mengandung kesalahan. Untuk beberapa model validasinya seperti menggunakan k-fold cross-validation, data dibagi menjadi k bagian (biasanya 10). Model dilatih pada k-1 bagian dan diuji pada 1 bagian. Proses ini diulang k kali, dengan setiap bagian digunakan sebagai data uji sekali, Stratified k-fold cross-validation: Sama seperti k-fold, tetapi memastikan bahwa setiap fold memiliki proporsi kelas yang sama dengan data aslinya. Penting untuk data yang tidak seimbang (imbalanced dataset). Hold-out validation, Data dibagi menjadi dua bagian: data latih dan data uji. Model dilatih pada data latih dan diuji pada data uji sekali.

Knowledge Discovery in Databases (KDD) adalah proses otomatis atau semi-otomatis untuk menemukan pola-pola yang bermakna, tren, dan informasi baru dari sejumlah besar data. KDD mencakup serangkaian langkah-langkah, pemilihan data dari data yang dibutuhkan, integrasi data, pemilihan data, transformasi data, penambangan data (data mining), interpretasi, dan evaluasi pola. Tujuan utama dari KDD adalah untuk mengekstrak pengetahuan yang berguna dan dapat diinterpretasikan dari data mentah yang besar, yang dapat digunakan untuk pengambilan keputusan yang lebih baik di berbagai bidang seperti bisnis, kesehatan, ilmu pengetahuan, dan teknologi (Hijriana & Muttaqin, n.d.).

1. Pemilihan Data *(Data Selection)*
   Pemilihan data agar relevan tujuannya untuk tugas analisis dari data yang telah diintegrasikan.
2. Preprocesing Data *(Pemrosesan Awal Data)*
   Tahap ini melibatkan penghapusan data yang tidak lengkap, tidak akurat, atau data yang memiliki anomali untuk meningkatkan kualitas data yang akan digunakan.
3. Transformasi Data *(Data Transformation)*

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

Mengubah data menjadi format yang sesuai untuk penambangan data. Ini bisa melibatkan normalisasi, agregasi, atau konstruksi atribut baru.

4. Penambangan Data *(Data* Mining*)*
   Proses inti dari KDD yang melibatkan penerapan metode atau algoritma untuk mengekstraksi pola-pola yang bermakna dari data. Teknik-teknik yang digunakan bisa mencakup klasifikasi, klastering, asosiasi, dan deteksi anomali dengan 3 model alagoritama seperti berikut:
   a. Algoritma C.45
      Keunggulan utama C4.5 adalah kemampuannya menghasilkan pohon keputusan yang mudah diinterpretasi dan divisualisasikan. Aturan-aturan klasifikasi dapat diekstrak dari pohon keputusan dalam bentuk "IF-THEN," sehingga mudah dipahami oleh manusia.
   b. K-Neareset Naighbors
      Konsepnya sangat intuitif dan mudah dipahami. KNN merupakan algoritma *lazy learning*, yang berarti tidak ada model yang dibangun secara eksplisit selama fase pelatihan. Klasifikasi dilakukan saat ada data baru yang akan diprediksi dengan mencari k tetangga terdekat.
   c. Naïve Beyes
      Naïve Bayes sangat cepat dalam proses pelatihan dan klasifikasi, terutama untuk data berukuran besar. Algoritma ini bekerja dengan baik pada data dengan jumlah fitur yang banyak. Naïve Bayes sering digunakan dalam aplikasi klasifikasi teks, seperti *spam filtering*.
5. Evaluasi Pola *(Pattern Evaluation)*
   Mengevaluasi pola-pola yang ditemukan untuk menentukan apakah mereka benar-benar bermakna dan berguna. Ini bisa melibatkan pengujian validitas pola dan interpretasi hasil.
6. Metrik Evaluasi klasifikasi
   a. Akurasi *(Accuracy)*
      Akurasi adalah Mengukur proporsi prediksi yang benar dari seluruh total prediksi dengan rumus sebagai berikut:

$$Akurasi = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

      Semakin tinggi akurasi, semakin baik performa model secara keseluruhan. Namun, akurasi bisa menyesatkan jika dataset tidak seimbang (imbalanced dataset), yaitu jumlah data untuk setiap kelas sangat berbeda.
   b. Presisi *(Precision)*
      Mengukur proporsi prediksi positif yang benar dari seluruh prediksi positif. Presisi menjawab pertanyaan "Dari semua yang diprediksi positif, berapa yang benar-benar positif?" Rumus Presisi adalah sebagai berikut:

$$Presisi = \frac{TP}{(TP+FP)} \quad (2)$$

      Interpretasi: Presisi penting jika biaya dari false positive tinggi. Contohnya, dalam deteksi spam, presisi yang tinggi berarti email yang ditandai sebagai spam benar-benar spam, sehingga meminimalisir email penting yang masuk ke folder spam.
   c. *Recall* (Sensitivitas/Sensitivity atau True Positive Rate/TPR)
      Mengukur proporsi data aktual positif yang diprediksi dengan benar. *Recall* menjawab pertanyaan "Dari semua yang benar-benar positif, berapa yang berhasil diprediksi dengan benar?" Rumus Recall Adalah sebagai berikut:

$$Recall = \frac{TP}{(TP+FN)} \quad (3)$$

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

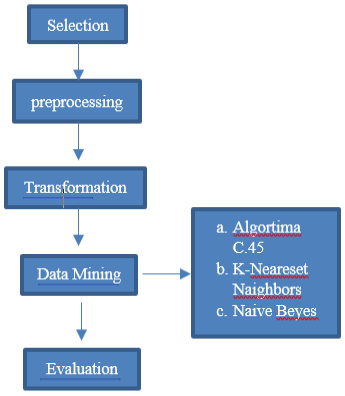
Interpretasi: *Recall* penting jika biaya dari *false negative* tinggi. Contohnya, dalam deteksi penyakit, *recall* yang tinggi berarti model berhasil mendeteksi sebagian besar kasus penyakit, sehingga meminimalisir kasus yang terlewat dan tidak diobati.

d. F1-*score*

Rata-rata harmonik antara presisi dan recall. F1-score mempertimbangkan baik presisi maupun recall, sehingga memberikan metrik yang lebih seimbang, terutama jika terjadi trade-off antara keduanya. Rumus F1-score adalah sebagai berikut:

$$F1 - score = 2 \times \frac{(Presisi \times Recall)}{(Presisi + Recall)} \quad (5)$$

Interpretasi: F1-score mencapai nilai tertinggi (1) jika presisi dan recall keduanya tinggi, dan nilai terendah (0) jika salah satunya bernilai 0.



Gambar 1. Flowchart Tahapan Studi

2.1. Tahapan Penelitian

1) Studi Pendahuluan

Pada Tahap ini penulis meakukan pengamatan langsung dan studi awal di Universitas Amikom Purwokerto untuk memahami situasi dan permasalahan yang terkait dengan pengambilan mata kuliah konsentrasi oleh mahasiswa. Observasi ini bertujuan untuk mengumpulkan data nilai mahasiswa dari smester 1 hingga semester 4 untuk di analisis dengan data mining.

2) Kajian Metode

Pada tahap ini untuk analisis penambangan sebuah data akan menggunakan metode algoritma C4.5 untuk membuat pohon keputusuan untuk mencari apakah dengan metode ini relevan tidak untuk jadi acuan sebagai pengambilan mata kuliah konsentrasi mahasiswa.

3) Pengolahan Data

Tujuan analisis dalam penelitian ini adalah untuk mengidentifikasi pola dominasi dalam pengambilan mata kuliah konsentrasi oleh mahasiswa dan menghasilkan rekomendasi yang dapat membantu mahasiswa dalam memilih mata kuliah konsentrasi yang sesuai.

4) Pengujian Sistem Menggunakan Rapid Miner Versi 9.10.

Tahap ini penulis mengumpulkan data nilai akademik mahasiswa dari Universitas Amikom Purwokerto yang akan di analisis. Setelah data terkumpul, dilakukan pengujian *dengan C.45 algoritme* untuk membangun sebuah pohon keputusan. Analisis ini akan memberikan hasil berbagai pola yang bisa mendaji patokan penentuan atau pemilihan mata kuliah konsentrasi lebih dominan.

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

3. HASIL DAN PEMBAHASAN

3.1 Pengolahan Data

a. Selection Data (Data Selection)

Proses analisis data dimulai dengan pengumpulan sekumpulan data mahasiswa yang relevan. Selanjutnya, algoritma C4.5 akan digunakan untuk menganalisis data tersebut sebelum tahap seleksi informasi dilakukan. Tujuan dari analisis ini adalah untuk mengidentifikasi pola dan tren dalam data yang dapat digunakan untuk menentukan pengambilan mata kuliah konsentrasi (Harapan & Rismayanti, 2018).

Analisa data mendalam dilakukan terhadap keseluruhan data mahasiswa program studi Sistem Informasi di Universitas Amikom Purwokerto mulai dari semester satu hingga semester lima. Fokus utama penelitian ini adalah pada data pengambilan mata kuliah konsentrasi mahasiswa yang terhimpun pada tahun ajaran 2023/2024. Sampel penelitian terdiri dari 560 data mahasiswa

1) Data matakuliah

Kurikulum program studi Sistem Informasi di Universitas Amikom Purwokerto dirancang cukup padat dengan total 56 matakuliah yang harus diselesaikan mahasiswa dalam 146 SKS. Sebagai gambaran, setiap semester mahasiswa diharuskan mengelola beban studi maksimal 9 matakuliah atau setara dengan 24 SKS. Beban studi yang terbilang intensif ini bertujuan untuk membekali lulusan dengan kompetensi yang relevan dengan kebutuhan industri.

Tabel 1. Matakuliah Semester 1

No	Kode MK	Nama Matakuliah	SKS
1	NSIFW001	Pendidikan Pancasila dan Kewarganegaraan	3
2	NSIFW002	Pendidikan Agama	3
3	USIFW001	Sikap Mental Amikom	1
4	PSIFW001	Kalkulus Dasar	2
5	PSIFW003	Sistem Basis Data	4
6	PSIFW004	Arsitektur dan Organisasi Komputer	3
7	USIFW002	Technopreneurship	2
8	NSIFW003	Pendidikan Anti Korupsi	2

Tabel 2. Matakuliah Semester 2

No.	Kode MK	Nama Matakuliah	SKS
1	NSIFW008	Bahasa Indonesia	3
2	FSIFW001	Pengantar Ilmu komputer	3
3	USIFW004	Pengantar Multimedia	3
4	PSIFW005	Pengantar Sistem Cerdas	2
5	PSIFW014	Logika Digital dan Sistem Digital	3
6	FSIFW002	Aljabar Linier dan Matrik	3
7	FSIFW003	Matematika Diskret	3

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

Tabel 3. MataKuliah Semester 3

No.	Kode MK	Nama Matakuliah	SKS
1	PSIFW006	Pemrograman Berorientasi Objek	3
2	PSIFW008	Algoritma dan Struktur Data	3
3	USIFW005	Bahasa Inggris	2
4	PSIFW009	Bahasa Pemrograman Python	4
5	FSIFW004	Sistem Operasi	3
6	PSIFW010	Jaringan Komputer	3
7	PSIFW002	Kalkulus Lanjut	2

Tabel 4. MataKuliah Semester 4

No.	Kode MK	Nama Matakuliah	SKS
1	PSIFW012	Statistik Probabilitas	3
2	PSIFW013	Pemodelan 2 Dimensi	3
3	PSIFW031	Cloud Computing	3
4	PSIFW027	Pemrograman Mobile	4
5	PSIFW022	Cloud Computing	2
6	FSIFW005	Bahasa Inggris Lanjut	2
7	PSIFW021	Teori Graf dan Otomata	3

2) Data Penelitian

Dalam penelitian ini menggunakan data mahsaiswa yang berjumlah 560 dengan data nilai semester 1 sampai dengan semester 4 berikut adalah tabel data mahasiswanya.

Tabel 5. Data penilitian

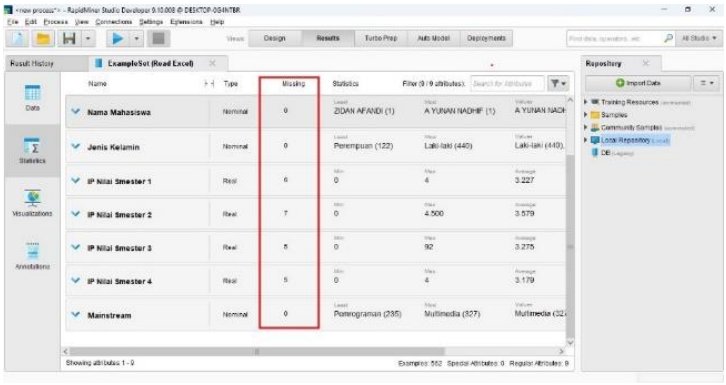
Jenis Kelamin	Semester 1	Semester 2	Semester 3	Semester 4	Konsentrasi
Laki-laki	3.56	3.78	3.85	3.78	Multimedia
Laki-laki	3.56	3.64	3.5	3.85	Pemrograman
Laki-laki	4	3.92	3.92	3.78	Multimedia
Perempuan	3.5	4	4	3.92	Multimedia
Laki-laki	3.06	3.42	3.07	3.07	Multimedia
Laki-laki	3.31	3.78	3.64	3.85	Multimedia
Perempuan	3.6	3.78	3.92	3.78	Pemrograman
Laki-laki	3.25	3.71	3.71	3.78	Multimedia
Laki-laki	3.3	3.42	3.42	3.57	Multimedia
Laki-laki	3.31	3.85	4	3.92	Multimedia
Laki-laki	3.6	3.85	4	4	Multimedia
Laki-laki	3.56	3.85	3.92	3	Pemrograman
Laki-laki	3.12	3.57	3.35	3.71	Multimedia

b. Preprocesing Data *(Pemrosesan Awal Data)*

Pada tahap ini pengumpulan seluruh data nilai mahasiswa yang terkumpul semua terdapat 560 siswa data awal dan akan menyeleski beberapa data yang hilang atau *missing* sehingga tidak menggangu saat proses data mining.

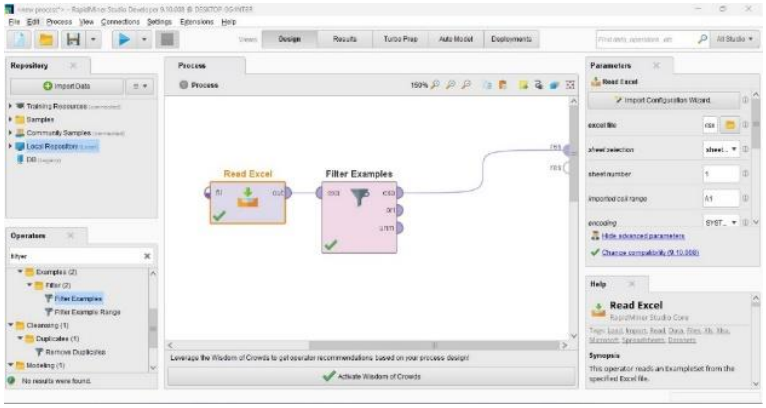
Pada gambar 2 merupakan tahapan preprocessing data dari 560 siswa terdapat bebrapa data missing pada variable IP Nila Smester 1 hingga 4 yang harus di perbaiki atau diseleksi sehingga data keseluruhan siap untuk digunakan analisis data mining.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157



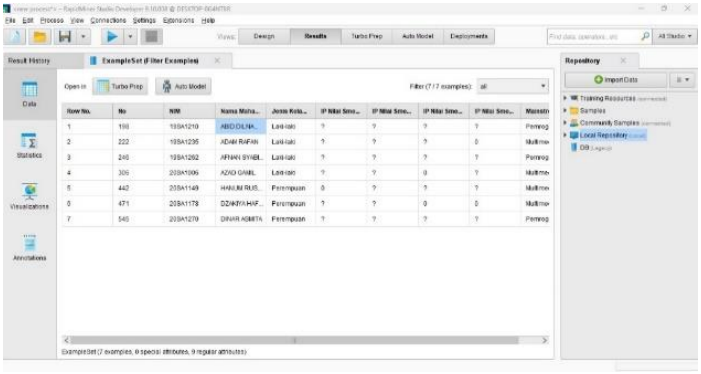
Gambar 2. *Preprocesing* data *missing*

Pada gambar 2 merupakan tahapan preprocessing data dari 560 siswa terdapat bebrapa data missing pada variable IP Nila Smester 1 hingga 4 yang harus di perbaiki atau diseleksi sehingga data keseluruhan siap untuk digunakan analisis data mining.



Gambar 3. *Filter examples*

Pada gambar 3 merupakan proses pemilihan data mahasiswa yang missing atau tidak jelas dengan menggunakan operator *Read Excel* kemudian di sambungkan dengan operator *Filter Examples* dengan disambugkan pada jalur *exa* dan dari operator *Filter Examples* dilanjutkan dari jalur exa disambnungkan ke jalur *res* lalu tekan tombol run untuk menjalanakan.

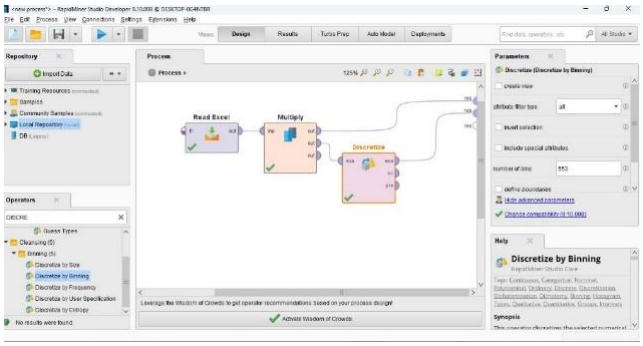


Gambar 4. Jumlah data mahasiswa yang hilang

Pada gambar 4 merupakan hasil proses *Filter Exampels* yang terdapat 7 data mahasiswa yang missing atau tidak jelas dari mulai smester 1 hingga 4, sehingga data ini tidak akan dipakai

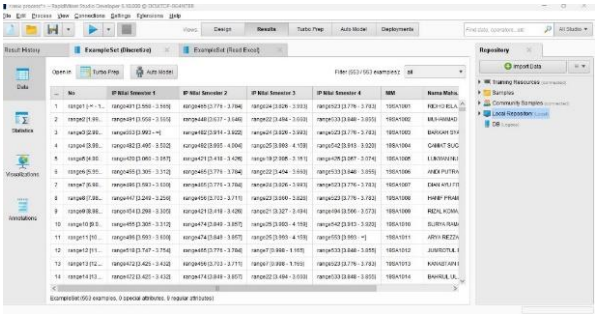
Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

untuk dimasukan ke proses data mining, jadi dengan adanya pengurangan 7 data mahasiswa tersebut dari data awal berjumlah 560 mahasiswa yang siap untuk di analisis menggunakan data mining menjadi 553 data mahasiswa.



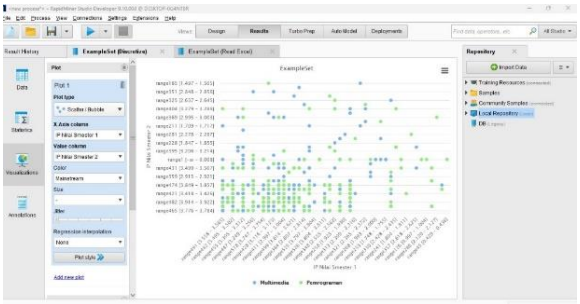
Gambar 5. Desain Proses diskretisasi data

Pada gambar 5 merupakan proses diskretisasi data menggunakan rapid miner dengan melakukan import data dengan menggunakan operator *Read Excel* kemudian disambungkan pada operator *multiply* dan dilanjutkan dari *multiply* disambungkan pada operator *discretize* dengan jalur out dari *multiply* disambungkan ke jalur *exa* pada operator *discretize* dan dialnjutkan ke jalur *res* untuk di jalankan.



Gambar 6. Merupakan hasil proses diskretisasi data

Pada Gambar 6 Merupakan hasil proses diskretisasi data data dengan jumlah range pada setiap data mahasiswa menggunakan *rapid miner.*



Gambar 7. Visualisasi deskritisasi data

Pada gambar 7 merupakan hasil visualisai data rangedari setiap mahasiswa dengan warna biru konsentrasi multimedia dan warna hijau konsentrasi pemrograman. Dari semua data siswa menghasilkan klasifikasi data nilai range dengan kategori rendah adalah [1.497 – 1.855], Sedang

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

[2.278 – 2.921] dan Tinggi [3.206 – 3.922] sehingga data mahasiswa bisa di kategorisasi dengan 3 range tersebut.

c. Transformasi *(Data Transformasi)*

Sebelum memulai proses penambangan data. Transformasi data memakan waktu lebih lama untuk memproses. Proses transformasi dilakukan untuk menemukan basis data yang akan digunakan. Proses transformasi dilakukan dengan membuat kategori untuk setiap atribut. Sebelum melakukan transformasi data, proses pembersihan data harus diselesaikan terlebih dahulu.Klasifikasi atribut nilai dibagi menjadi tiga bagian, yaitu

Tabel 5. Klasifikasi Nilai

No	IPK Semester	Kelas
1	IPK >= 3.5	Tinggi
2	IPK >= 2.75	Sedang
3	IPK < 2.75	Kecil

Berdasarkan hasil transformasi pada klasifikasi nilai table 5 dapat disimpulkan seperti terlihat pada tabel 6 berikut ini :

Tabel 6. Transformasi data.

IP Nilai Smester 1	IP Nilai Smester 2	IP Nilai Smester 3	IP Nilai Smester 4	Mainstream
Tinggi	Sedang	Kecil	Tinggi	Multimedia
Tinggi	Sedang	Tinggi	Tinggi	Pemrograman
Tinggi	Tinggi	Tinggi	Tinggi	Multimedia
Tinggi	Kecil	Tinggi	Tinggi	Multimedia
Sedang	Sedang	Tinggi	Sedang	Multimedia
Sedang	Sedang	Tinggi	Sedang	Multimedia
Tinggi	Tinggi	Sedang	Tinggi	Pemrograman
Sedang	Sedang	Tinggi	Sedang	Multimedia
Sedang	Tinggi	Sedang	Sedang	Multimedia
Sedang	Tinggi	Tinggi	Sedang	Multimedia
Tinggi	Tinggi	Tinggi	Tinggi	Multimedia
Tinggi	Tinggi	Sedang	Tinggi	Multimedia
Tinggi	Sedang	Tinggi	Sedang	Multimedia
Tinggi	Tinggi	Sedang	Tinggi	Multimedia
Tinggi	Tinggi	Tinggi	Sedang	Pemrograman
Sedang	Sedang	Sedang	Tinggi	Multimedia
Tinggi	Sedang	Tinggi	Tinggi	Multimedia
Tinggi	Sedang	Tinggi	Sedang	Multimedia
Sedang	Kecil	Sedang	Sedang	Multimedia

d. Penambangan Data *(Data Mining)*

1. Algoritma C.45

Analisis akan dilakukan berdasarkan data berubah dalam rangka untuk menghasilkan berupa keputusan pohon C4.5 menggunakan algoritma. Pada umumnya, algoritma yang digunakan untuk menciptakan C4.5 berikut pohons:

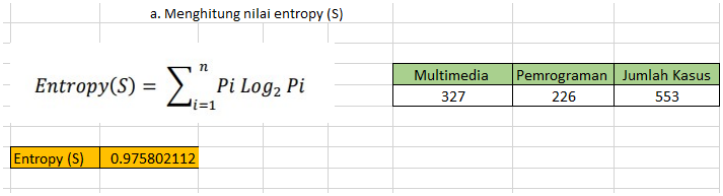
a) berlatih dengan ekstraksi data.Pelatihan berasal dari data data sejarah yang telah terjadi dan telah yang telah diuraikan dalam kelas-kelas yang bersangkutan.
b) membuat akar dari pohon.Akar akan dianalisis oleh menghitung keuntungan dari masing masing individu atribut.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

Mendapatkan persentase tertinggi akan menjadi dasar akar.Sebelum penghitungan memperoleh manfaat dari atribut, hendaknya terlebih dahulu dihitung entropi.

1) Merancang Decision tree menggunakan *algoritma C4.5.*

Langkah selanjutnya adalah menentukan entropi dan keuntungan untuk menentukan akar atau node berdasarkan tabel data penelitan. Berikut adalah hasil dari entropi dan keuntungan perhitungan pada node 1. Bisa dilihat pada tabel 8.

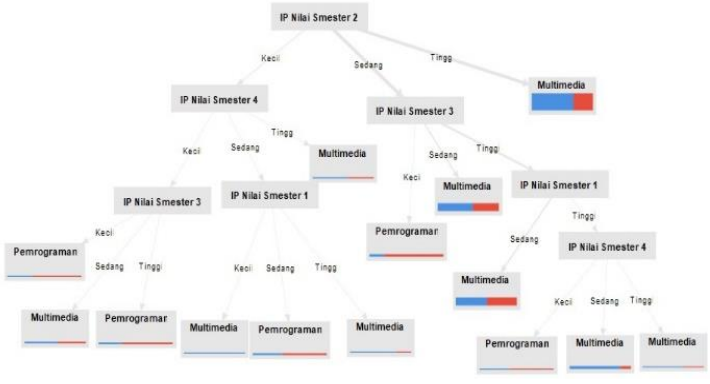


Gambar 8. Hasil Hitungan nilai entropy

Tabel 8. *Node 1*

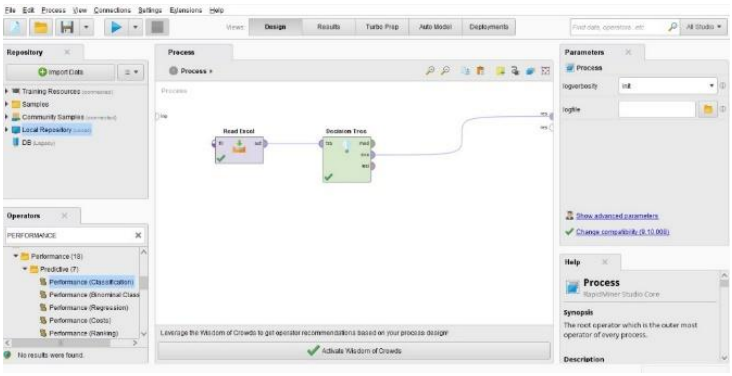
Jumlah Kasus	Pemrograman	Multimedia	Entropy	Gain
154	118	36	0.784519139	
254	198	56	0.761029228	0.148828713
145	63	82	0.987578748	
176	95	81	0.995430863	
164	32	132	0.712064055	0.154562936
213	166	47	0.761365857	
180	122	58	0.906781691	
232	161	71	0.888542793	0.093508128
141	103	38	0.840751354	
195	139	56	0.865049879	
183	86	97	0.997392108	0.069994926
175	49	126	0.855450811	
434	176	258	0.974093636	0.055251354
119	95	24	0.725276731	

Menurut tabel di atas, dapat dilihat bahwa atribut tertinggi adalah IP semester 2 yaitu 0.154562936. Melakukan perhitungan serupa untuk menentukan peningkatan dan entropi dan gain berikutnya. Atribut dari ip semester 1 sebagai akar utama atau node 1. Berdasarkan hasil dari hitungan nilai entropi dan mendapatkan di titik 1, pohon keputusan atau keputusan manual dapat ditampilkan seperti yang diperlihatkan dalam ilustrasi di bawah ini:



Gambar 9. Merupakan hasil dari pohon keputusan menggunakan rapid minder

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157



Gambar 10. Proses pembuatan pohon keputusan

Pada gambar 10 merupakan tahapan untuk merancang pohon keputusan Dengan menggunakan aplikasi *Rapidminer* dengan menggunakan operatao *Read excel* dengan memasukan data mahasiswa yang berupa *Excel* dan dilanjutkan disambungkan ke operator *decision tree* untuk dijalankan dengan menekan tombol *running.*

Gambar 11. Hasil description pohon keputusan

Pada gambar 11 Merupakan hasil *results* dari *decision tree* yang berhasil dan tidak mengalami error, dengan hal tersebutbahwa model keputusan berjalan lancar.

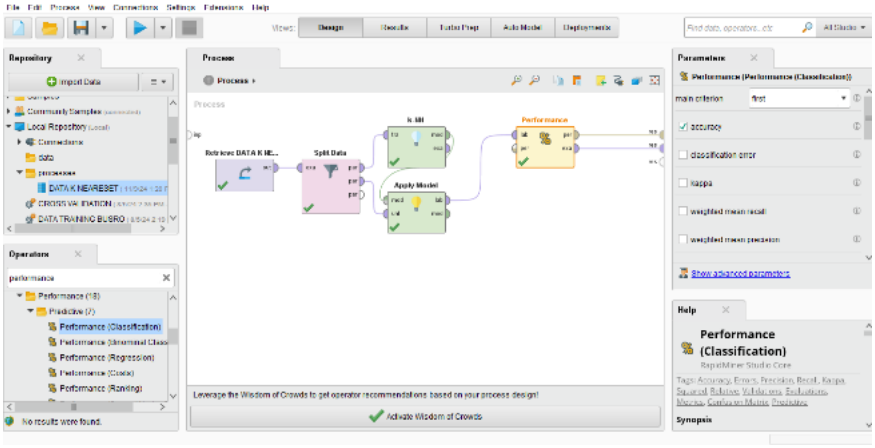
2. K-Nearest Neighbors

Metode yang disebut KNN (K-Nearest Neighbor) mengklasifikasikan objek dengan memanfaatkan data dari tetangga terdekat. Jarak Euclidean digunakan untuk menentukan seberapa dekat atau jauhnya suatu tetangga. Metode ini terdiri dari dua fase pelatihan (training) dan klasifikasi (testing) seperti pada aplikasi rapidminer seperti berikut:

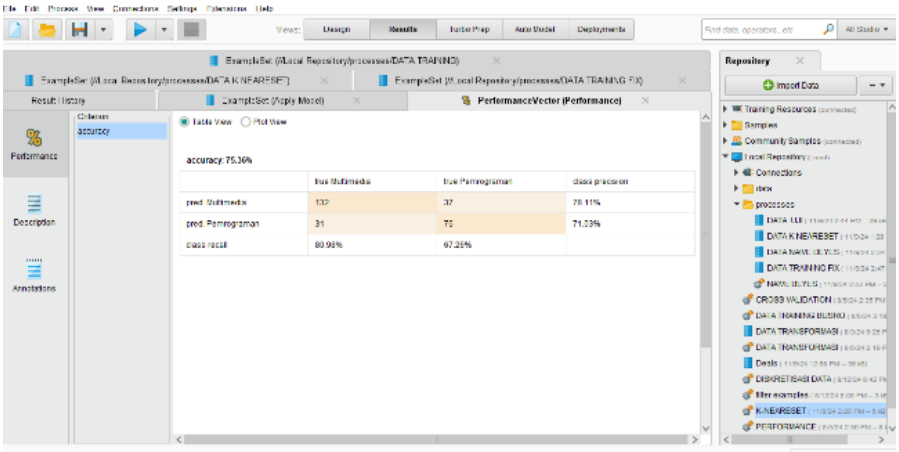
Pada 12 merupakan tahapan untuk melakukan eksplorasi data awal untuk memahami distribusi data, mencari missing values, dan mengecek tipe data. Pada tahap selanjutmya yaitu melakukan split data Bagi dataset menjadi dua bagian: data latih (training data) dan data uji (testing data). Data latih digunakan untuk melatih model, sedangkan data uji digunakan untuk mengevaluasi kinerja model. RapidMiner menyediakan operator Partition untuk membagi data. Pemilihan Atribut untuk menentukan atribut mana yang relevan untuk proses klasifikasi atau regresi atribut yang tidak relevan dapat menurunkan kinerja model dan menggunakan operator Select Attributes untuk memilih atribut yang diinginkan. Penerapan Algoritma K-NN. Masukan

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

operator KNN di palet operator RapidMiner. Jumlah tetangga terdekat yang akan dipertimbangkan. Nilai K yang optimal biasanya ditentukan melalui eksperimen kemudian pilih model klasifikasi.



Gambar 12. Proses penyambungan atribut pada rapidminer



Gambar 13. Hasil Pengujian sistem model pohon keputusan menggunakan rapid miner.

Pada gambar 13 merupakan hasil pengujian dan evaluasi model K-Neareset Naighbors yang sudah dirancang dengan mendapatkan hasil *accuracy* 75,36% kemudian untuk *classprecision* multimedia 78,11% dan pemrograman 71,03%, dan untuk *class recall* pada multimedia 80,98% dan pemrograman 67,26%. Berdasarkan hasil pengujian tersebut bahwa pohon keputusan cukup layak digunakan sebagai metode yang efisien untuk pemilihan mata kuliah konsentrasi karena *accuracy* dan *avvarage* di atas 50% jadi sangat bisa membantu mahasiswa Universitas Amikom Purwokerto yang masih kesulitan dan kebingungan untuk menentukan mata kuliah konsentrasi.

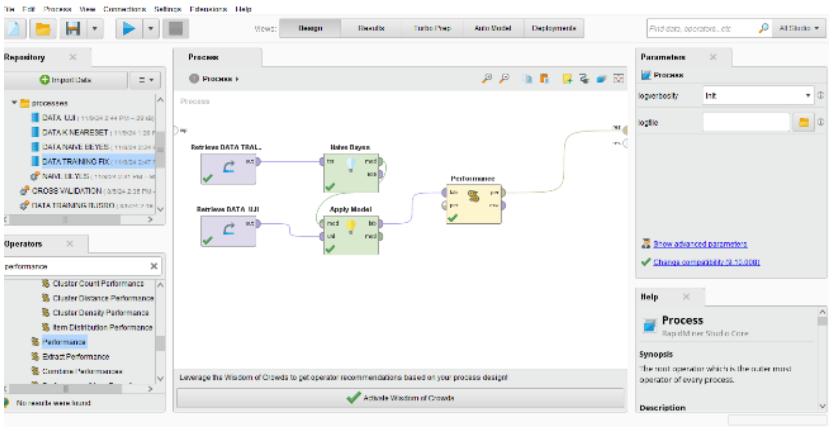
3. Naive Bayes

Dalam penelitian ini, digunakan algoritma Naive Bayes untuk mengklasifikasikan data mahasiswa untuk memilih mata kuliah konsentrasi dengan tahapan sebagai berikut:

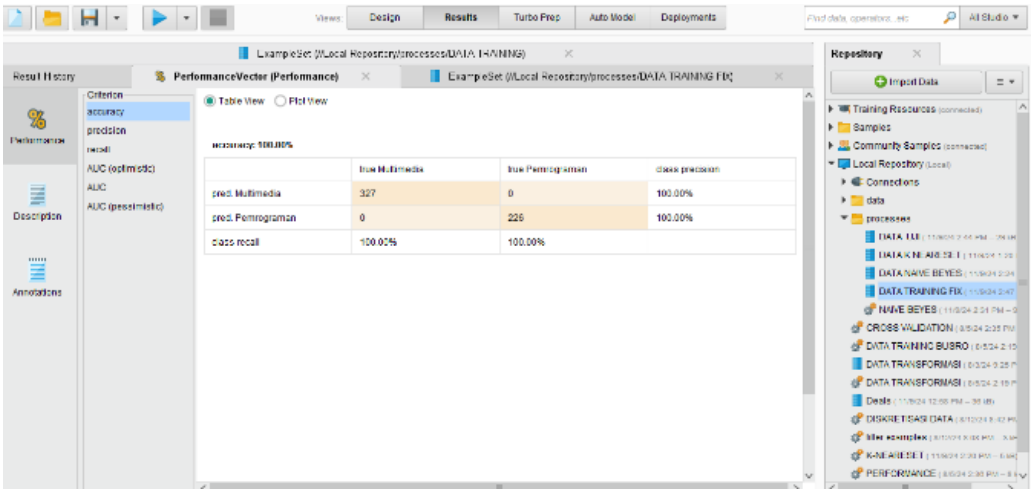
Pada gambar 14 merupakan tahapan untuk melakukan klasifikasi data dengan mengimport dataset impor dataset ke dalam RapidMiner, kemdudian melakukan preprocessing data, seperti tokenisasi, stop word removal, dan stemming, mengekstrak fitur dari teks menggunakan teknik TF-IDF. Bagi dataset menjadi data latih dan data uji.

Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

Memasangkan model Naive Bayes menggunakan operator Naive Bayes di RapidMiner. Evaluasi kinerja model menggunakan matriks konfusi dan berbagai metrik evaluasi lainnya. Analisis hasil klasifikasi dan interpretasikan kinerja model.



Gambar 14. Proses penyambungan atribut pada rapidminer



Gambar 15. Hasil Pengujian sistem model pohon keputusan menggunakan rapid miner.

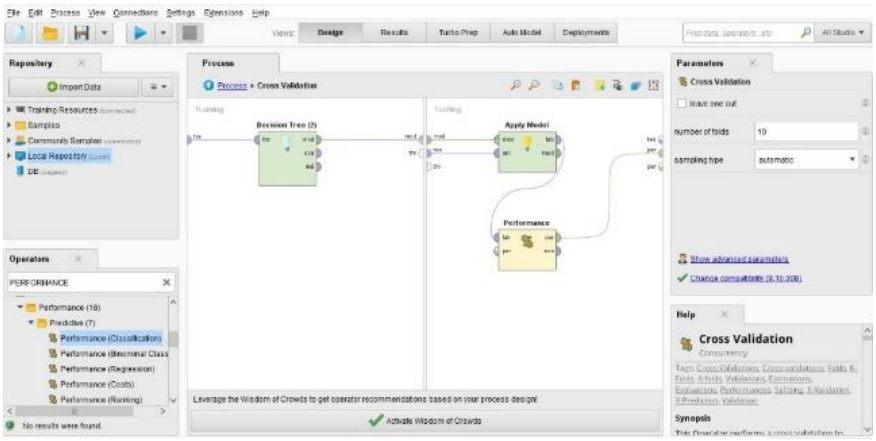
Pada gambar diatas Merupakan hasil pengujian dan evaluasi model K-Neareset Naighbors yang sudah dirancang dengan mendapatkan hasil *accuracy* 100.00% kemudian untuk *classprecision* multimedia 100.00% dan pemrograman 100.00%, dan untuk *class recall* pada multimedia 100.00% dan pemrograman 100.00%. Berdasarkan hasil pengujian tersebut bahwa pohon keputusan sangat layak digunakan sebagai metode yang efisien untuk pemilihan mata kuliah konsentrasi karena *accuracy* dan *avvarage* sempurna 100.00% % jadi sangat bisa membantu mahasiswa Universitas Amikom Purwokerto yang masih kesulitan dan kebingungan untuk menentukan mata kuliah konsentrasi.

a. Evaluasi Pola *(Pattern Evaluation)*

Pada tahap ini dilakukan evaluasi pola-pola yang ditemukan untuk menentukan apakah mereka benar-benar bermakna dan berguna. Ini bisa melibatkan pengujian validitas pola dan interpretasi hasil seperti pada gambar berikut ini:

Pada gambar 16 Merupakan proses evaluasi apakah model keputusan yang dirancang sudah layak untuk di jadikan acuan sebagai metode pengambilan matakuliah konsentrasi.

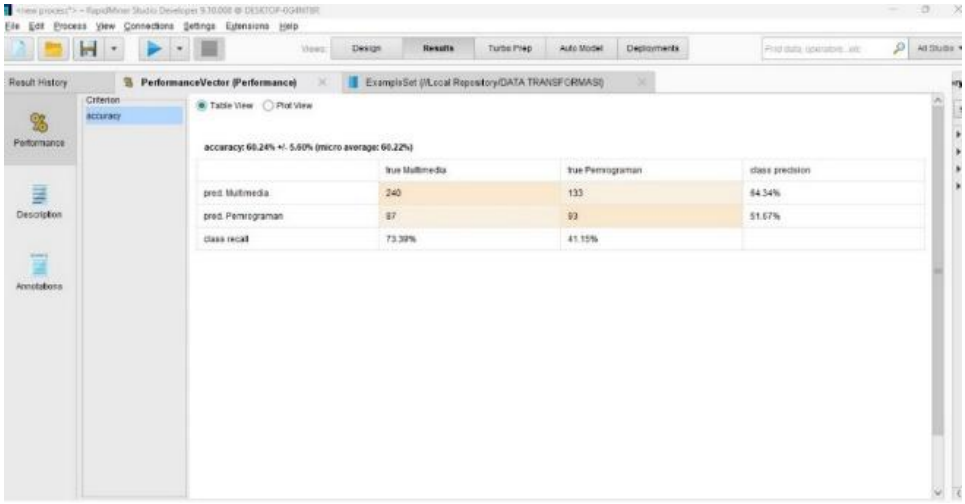
Jurnal Teknik Informatika (JUTIF)
P-ISSN: 2723-3863
E-ISSN: 2723-3871
Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157



Gambar 16. Proses evaluasi pohon keputusan

4. DISKUSI

4.1. Pengujian Sistem Manggunakan Aplikasi Rapid Miner 9.10



Gambar 17. Hasil Pengujian sistem model pohon keputusan menggunakan rapid miner.

Pada gambar 17 Merupakan hasil pengujian dan evaluasi model phonon keputusan yang sudah dirancang dengan mendapatkan hasil *accuracy* 60,24% dan micro *avvarege* 60,22%, kemudian untuk *classprecision* multimedia 64,34% dan pemrograman 51,67%, dan untuk *class recall* pada multimedia 73,39% dan pemrograman 41,15%. Berdasarkan hasil pengujian tersebut bahwa pohon keputusan cukup layak digunakan sebagai metode yang efisien untuk pemilihan mata kuliah konsentrasi karena *accuracy* dan *avvarage* di atas 50% jadi sangat bisa membantu mahasiswa Universitas Amikom Purwokerto yang masih kesulitan dan kebingungan untuk menentukan mata kuliah konsentrasi.

4.2. Hasil Pengujian 3 Model Algoritma Manggunakan Aplikasi Rapid Miner 9.10

Tabel 9 hasil data mini 3 model algoritma menggunakan rapidminer

Metode	Accuracy
Algoritma C.45	60,24%
KNN	75,36%
Naive Beyes	100.00%

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

Pada tabel diatas terdapt hasil dari 3 data mining yang sangat berbeda terutama pada metode Naive Beyes terdapat hasil sempurna yaitu 100.00% kemudain pada metode K-Neareset naighbors terdapat hasil angka 75,36% dan metode menggunakan Algoritma C.45 paling kecil terdapat hasil 60,24% sehingga metode paling cocok digunakan yaitu menggunakan metode algoritma Naive beyes.

5. KESIMPULAN

Berdasarkan hasil penelitian, *algoritma C4.5* dengan model pohon keputusan dapat menjadi salah satu alternatif metode dalam memprediksi pilihan matakuliah konsentrasi mahasiswa. Namun, perlu dilakukan penelitian lebih lanjut untuk meningkatkan akurasi model dan mempertimbangkan faktor-faktor lain yang mempengaruhi hasil prediksi dengan 3 model Algoritma menggunakan rapid miner seperti menggunakan *Algoritma C4.5* dengan model pohon keputusan, K-Neareset Naighbors dan Naïve beyes yang menunjukkan terdapat satu model yang potensi dalam memprediksi pilihan matakuliah konsentrasi mahasiswa yaitu model Niave Beyes dengan akurasi 100,00% mengindikasikan bahwa model mampu mengklasifikasikan data dengan benar lebih dari setengah dari waktu. kemudain pada metode K-Neareset naighbors terdapat hasil angka 75,36% dan metode menggunakan Algoritma C.45 paling kecil terdapat hasil 60,24% sehingga metode paling cocok digunakan yaitu menggunakan metode algoritma Naive beyes. Dengan demikian, dapat disimpulkan bahwa hasil penelitian ini memberikan titik awal yang baik untuk pengembangan sistem pendukung keputusan dalam pemilihan matakuliah konsentrasi bisa menggunakan model algoritma Naïve beyes.

Meskipun memiliki potensi, akurasi 100,00% masih butuh analisis elbih luas lagi dengan menggunakan algortima yang lainnya lagi sehingga akurasinya bisa dikatakan valid kalo sudah diuji dengan berbagai metode atau model algoritma.

Hal ini menunjukkan bahwa model sudah cukup baik dalam menangkap pola yang kompleks dalam data. Akurasi model juga dipengaruhi oleh kualitas data yang digunakan, pemilihan atribut, dan parameter tuning algoritma. ntuk meningkatkan akurasi model, beberapa hal yang dapat dilakukan seperti menggunakan teknik *feature engineering* Mengubah atau menggabungkan atribut yang ada untuk menciptakan fitur baru yang lebih informatif. Mencoba algoritma lain Membandingkan kinerja *algoritma C4.5,* K-Neareset naighbors dan naïve beyes dengan algoritma lain seperti atau *Support Vector Machine, randm forest* dan Menyesuaikan parameter algoritma dengan mencoba berbagai kombinasi parameter untuk menemukan nilai optimal.

6. SARAN

Untuk penelitian selanjutnya diharapkam mencoba menggunakan algoritma yang lainnya seperti SVM, Random Fores, dan lainnya dengan potensi menemukan algoritma yang lebih praktis dan maksimal disbanding penelitian ini. Penelitian di masa depan harus mengeksplorasi metode ansambel atau algoritma pembelajaran mendalam untuk lebih meningkatkan akurasi.

REFERENCES

[1] L. Swastina, “Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa,” vol. 2, no. 1, 2013.

[2] F. Nasari, S. Informasi, P. Keputusan, and D. Selection, “Penerapan algoritma c4.5 dalam pemilihan bidang peminatan program studi sistem informasi di stmik potensi utama medan,” pp. 30–34, 2014.

[3] D. Anggraeni, A. Z. Syah, and S. Informasi, “Tips Dan Trik Membangun Relationship Dan Query Dalam,” vol. 1, no. 2, 2018.

[4] A. I. Tarigan, “Model Optimasi Pemetaan Mata Kuliah Berprasyarat Untuk Rencana Studi Mahasiswa (Studi Kasus Program Studi Matematika Fmipa Ut).”

[5] . Studi, T. Informatika, U. Harapan, and M. Sumatera, “Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan),” vol. 5341, no. April, pp. 16–24, 2018.

Vol. 6, No. 4, August 2025, Page. 1683-1698
https://jutif.if.unsoed.ac.id
DOI: https://doi.org/10.52436/1.jutif.6.4.4157

[6] M. S. Sinambela, E. Rosely, F. I. Terapan, and U. Telkom, "THE DECISION SUPPORT SYSTEM FOR SUBJECT SPECIALIZATION STUDENTS IN," vol. 2, no. 3, pp. 858–866, 2016.

[7] D. Anggraeni, and T. Christy, Analisa Kinerja Algoritma C4.5 Dalam Menentukan Pola Dominasi Mainstream Mahasiswa. Vol, 6 No. 4, pp. 333-334, 019.

[8] Shrooq Algarni et al., "Systematic Review of Recommendation Systems for Course Selection," Machinel Learning & knowledge extraction., Vol 5, no. 3, pp. 560–596. doi: 10.3390/make5020033.

[9] Christine Lahoud et al., "A comparative analysis of diferent recommender systems for university major and career domain guidance." Education and Information Technologies., Vol 7, no.6, pp. :8733–8759. Doi: 10.1007/s10639-022-11541-3.

[10] Normah et al., "Comparison of Classification C4.5 Algorithms and Na¨ıve Bayes Classifier in Determining Merchant Acceptance on Sponsorship Program." Journal of Physics: Conference Series., Vol 3, no. 4, pp. doi:10.1088/1742-6596/1641/1/012006.

[11] Vraj Sheth et al., "A Comparative Analysis of Machine Learning Algorithms for nce on Innovative Data Comm." Procedia Computer Science., Vol 4, no. 4, pp. doi: 215 (2022) 422–431.