pISSN: 2442-3386 eISSN: 2442-4293
Vol 9 No 2 Jul – Des 2023, 135 - 144
Melek IT
INFORMATION TECHNOLOGY JOURNAL

PREDIKSI PENERIMAAN SNMPTN MENGGUNAKAN ALGORITMA C4.5 DAN NAIVE BAYES

Indah Purnamasari*[1], Indah Suryani[2]
[1,2]Universitas Nusa Mandiri,
indah.ihi@nusamandiri.ac.id[*1], indah.ihy@nusamandiri.ac.id[2]

Abstrak

Pendidikan mempunyai peranan penting bagi kehidupan manusia. Sekolah sebagai tempat mengenyam pendidikan mengajarkan berbagai disiplin ilmu dengan tujuan agar para peserta didik dapat menjadi manusia yang unggul. Secara umum masyarakat di Indonesia ingin putra putrinya dapat menempuh pendidikan terbaik di Perguruan Tinggi Negeri (PTN). Banyak jalur yang bisa dilalui menuju PTN, salah satunya adalah jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) yang merupakan jalur masuk PTN tanpa tes atau seleksi nilai raport. Akan tetapi kuota penerimaan PTN melalui jalur SNMPTN hanya minimal 20% dan tidak dapat ditentukan passing grade setiap PTN sehingga kemungkinan untuk bisa masuk PTN melalui jalur SNMPTN tidak diketahui. Oleh karena itu untuk dapat memprediksi siswa siswi yang dapat diterima di PTN melalui jalur SNMPTN maka dilakukan prediksi penerimaan SNMPTN menggunakan data mining klasifikasi dengan dua algoritma yaitu C4.5 dan Naive Bayes. Tujuannya adalah dengan siswa mengetahui kemungkinan diterima pada PTN melalui jalur SNMPTN maka diharapkan semakin banyak siswa yang dapat diterima PTN melalui jalur SNMPTN tersebut. Penelitian ini menunjukkan prediksi penerimaan PTN melalui jalur SNMPTN menggunakan data mining klasifikasi dengan algoritma C4.5 memberikan hasil akurasi yang jauh lebih tinggi yaitu sebesar 85,09 % dan nilai AUC 0,873 sedangkan algoritma Naive Bayes memberikan hasil akurasi 63.01% dan nilai AUC 0,665.
Kata Kunci: SNMPTN, Prediksi, Naive Bayes, C4.5

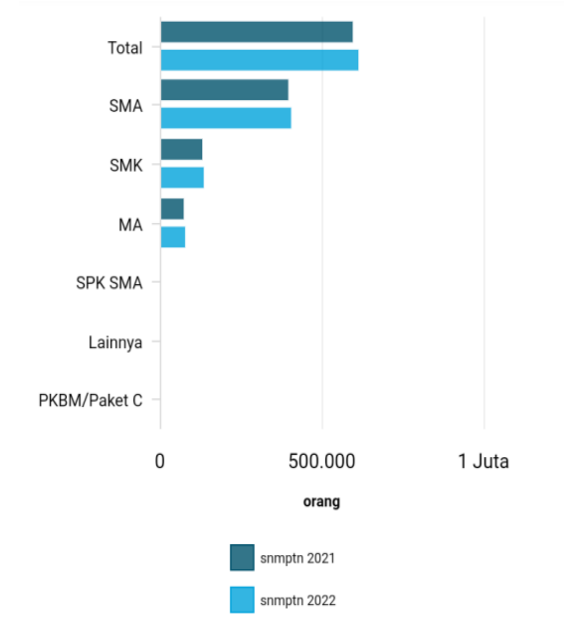
Abstract

School is a formal education unit where teachers teach various disciplines which aim to make students become superior human beings. Education is important for humans. In general, people in Indonesia want their sons and daughters to be able to get an education at State Universities (PTN). There are many paths that can be taken to PTN, one of which is the National Selection for State University Entrance (SNMPTN) which is the PTN entrance route without a test or selection of report cards. However, the quota for acceptance of PTN through the SNMPTN pathway is at least 20% and the passing grade of each PTN cannot be determined so that the possibility of being accepted into PTN through the SNMPTN route is unknown. Therefore, to be able to predict students who can be accepted into PTN through the SNMPTN path, predictions of SNMPTN acceptance are made using classification data mining with the C4.5 and Naive Bayes algorithm. The goal is that with students knowing the possibility of being accepted into PTN through the SNMPTN pathway, it is hoped that more students can be accepted by PTN through the SNMPTN pathway. This study shows that the prediction of PTN acceptance through the SNMPTN path using data mining classification with the C4.5 algorithm produces a much higher accuracy of 85,09% and the AUC value 0,873, while the Naive Bayes algorithm produces an accuracy of 63,01% and the AUC value..
Keywords: *SNMPTN, Prediction, Naive Bayes, C4.5*

I. PENDAHULUAN

Saat ini semakin tinggi kesadaran masyarakat Indonesia akan pentingnya pendidikan. Pendidikan memiliki peran penting untuk memotong rantai kemiskinan dan kebodohan, di Indonesia tahapan pendidikan bermula dari Sekolah Dasar (SD) selama 6 tahun, Sekolah Menengah Pertama (SMP) selama 3 tahun, Sekolah Menengah Atas (SMA) atau Sekolah Menengah Kejuruan (SMK) selama 3 taun kemudian Perguruan Tinggi [1]. Berbagai upaya pemerintah dalam mendukung pendidikan diantaranya dengan meningkatkan anggaran pendidikan serta program-program pendidikan sehingga pendidikan semakin merata, mudah diperoleh di seluruh daerah dan semakin maju. Berdasarkan dari databoks dengan sumber Lembaga Tes Masuk Perguruan Tinggi (LTMPT) tanggal 29 Maret 2022 menunjukkan semakin tinggi jumlah masyarakat Indonesia yang mendaftar pendidikan ke Perguruan Tinggi [2] seperti gambar 1.

Upaya untuk meningkatan kualitas calon mahasiswa baru di PTN telah dilakukan sejak 1976 dengan berbagai istilah. Menteri Riset, Teknologi dan Pendidikan Tinggi Republik Indonesia pada tanggal 4 Januari 2019 meluncurkan Lembaga Tes Masuk Perguruan Tinggi (LTMPT).



Gambar 1. Jumlah Pendaftar SNMPTN meningkat 2022

Banyak jalur yang bisa dilalui menuju PTN, salah satunya adalah jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) yaitu salah satu jalur masuk PTN tanpa tes atau seleksi berdasarkan nilai raport. Persentase kuota pada tahun 2019 ditetapkan oleh Lembaga Tes Masuk Perguruan Tinggi (LTMPT) sebesar minimal 20% dari besaran kuota daya tampung setiap prodi di PTN [3] dan tidak dapat ditentukan *passing grade* setiap PTN sehingga kemungkinan untuk bisa masuk PTN melalui SNMPTN tidak diketahui.

Proses penerimaan calon mahasiswa baru yang dilakukan oleh suatu perguruan tinggi, pada dasarnya harus objektif (tidak diskriminatif) dan transparan [4]. Selain dari itu antusias para siswa sangatlah besar untuk dapat diterima pada Perguruan Tinggi Negeri favoritnya, akan tetapi para siswa perlu memahami terlebih dahulu untuk memilih prodi dan PTN yang sesuai hasil pencapaiannya [5]. Potensi siswa siswa dari setiap sekolah yang diterima melalui jalur SNMPTN dapat dikenali polanya. Tetapi bila analisa data penerimaan melalui jalur SNMPTN dilakukan dengan cara manual tentunya akan memerlukan waktu yang lama [6].

Data mining dengan model prediksi mampu membantu dalam memprediksi. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma Decission Tree C.45, Artificial Neural Networks (ANN), KNearest Neighbor (KNN), algoritma Naive Bayes, Neural Network serta algoritma lainnya[7]

Algoritma Decision tree memiliki kelebihan yaitu dapat menghasilkan pohon keputusan yang mudah diinterprestasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat menangani atribut bertipe diskret dan numerik[8]

Faktor-faktor yang mempengaruhi waktu kelulusan bagi mahasiswa seperti, jenis kelamin, status pernikahan, umur, IPK, dan status pekerjaan. Faktor-faktor tersebut merupakan variabel yang akan diolah dengan data mining algoritma KNN (K-Nearest Neighbor), Naive Baye, C4.5. Data mahasiswa digunakan sebagai data preprocessing yang terdiri dari jenis kelamin, status pernikahan, status pekerjaan, umur dan IPK. Berdasarkan hasil confusion matrix, menunjukkan bahwa Naive Bayes mempunyai accuracy 100.00% dan AUC 1.000 lebih tinggi dibandingkan dengan C4.5 dan KNN. Sehingga algoritma Naive Bayes mempunyai kinerja yang lebih baik dibandingkan dengan KNN dan C4.5[9].

Data mahasiswa terus bertambah setiap tahun sehingga menumpuk seperti data yang terabaikan karena jarang digunakan. Data tentang mahasiswa yang lulus dapat memberikan informasi yang berguna jika dimanfaatkan dengan maksimal.. Penelitian ini mengunakan metode pohon keputusan yang dibangun dengan algoritma C4.5 disertai dengan algoritma error-based pruning untuk proses pemotongan pohon keputusan. Variabel-variabel yang digunakan yaitu jenis kelamin, IPK, asal daerah dan TOEFL. Algoritma C4.5 menghasilkan prediksi kelulusan dengan nilai rata-rata precision 63.93%, recall 60.73%, dan akurasi 60.52%. Dengan menggunakan metode error-based pruning, didapatkan hasil yang lebih baik dengan menggunakan nilai confidence 0,4 menghasilkan precision 70.70%, recall 50.65%, dan akurasi 61.57%. Sedangkan dengan menggunakan nilai confidence 0,25 menghasilkan precision 73.77%, recall 48.84%, dan akurasi 62.44%[10]

Banyak faktor mempengaruhi kelulusan seorang mahasiswa seperti kondisi ekonomi keluarga, nilai mahasiswa atau karena faktor lain yang berhubungan dengan tempat mahasiswa belajar. Kelulusan merupakan salah satu penilaian proses akreditasi suatu perguruan tinggi. Oleh karena itu apabila mahasiswa banyak yang lulus tepat waktu akan mempengarui nilai akreditasinya. Permasalahan tersebut di atas harus segera diatasi dengan suatu metode. Data mining salah satu metode yang paling tepat untuk mengatasi maslah tersebut di atas. Suatu keilmuan yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar. Permasalahan kelulusan tepat waktu merupakan hal yang prioritas pada suatu perguruan tinggi. Oleh karena itu peneliti mengusulkan untuk melakukan mengembangkan penelitian tentang Predkisi kelulusan tepat waktu yang semula hanya menggunakan metode naive bayes, peneliti menambahkan feature selection information gain sebagai feature untuk menyeleksi atribut yang berbobot.[11]

Salah satu kewajiban perguruan tinggi adalah menghasilkan lulusan yang kompeten. Selain itu, sidah

pISSN: 2442-3386 eISSN: 2442-4293
Vol 9 No 2 Jul – Des 2023, 137 - 144
Melek IT
INFORMATION TECHNOLOGY JOURNAL

menjadi impian para mahasiswa adalah lulus tepat waktu. Mahasiswa tidak perlu membayar biaya kuliah lagi dan bisa bekerja lebih cepat. Tetapi pada kenyataanya, mahasiswa belum tentu dapat menuntaskan masa studi tepat waktu. Banyak faktor yang menjadi pengaruh kelulusan mahasiswa terlambat, seperti status perkawinan mahasiswa, status mahasiswa (bekerja/tidak bekerja), tingkat pemahaman mahasiswa terhadap materi kuliah yang dapat dilihat dari IPK mahasiswa. Oleh karena itu, perlu adanya sistem untuk memprediksi tingkat kelulusan mahasiswa berdasarkan variabel-variabel yang ada. Dengan sistem yang dibuat diharapkan perguruan tinggi bisa membuat kebijakan sehingga mahasiswa dapat lulus tepat waktu. Penelitian ini menggunakan 379 data, dengan metode Naive bayes, dengan rincian data training 303 data dan data testing 76 data. Atribut yang digunakan nama, status mahasiswa, status perkawinan, IPS, IPK, dan status kelulusan. Dengan tahapan identifikasi masalah, pengumpulan data, data cleaning, data transformation (dibagi menjadi data training dan data tesing), klasifikasi dengan KNN, validasi, evaluasi dan hasil. Hasil penelitian yang diperoleh yaitu akurasi = 88,16%, precision = 93,62% dan recall = 88%, termasuk dalam kategori good classification.[12]

Dengan memanfaatkan perkembangan ilmu pengetahuan dalam data mining klasifikasi algoritma Naive Bayes dan algoritma C4.5 serta memanfaatkan teknologi yang dapat membantu dalam menganalisis data yang cukup besar dengan menggunakan *software data mining* yaitu rapidminer, maka dapat memberikan sebuah pola berdasarkan potensi nilai siswa yang dapat lolos seleksi SNMPTN.

Melalui penelitian ini diharapkan bisa membantu siswa untuk memilih perguruan tinggi negeri favoritnya dengan tepat sesuai hasil pencapain masing-masing siswa.

1.1 Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN)

Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) merupakan salah satu jalur untuk masuk PTN berdasarkan nilai rapotnya secara akademis dan prestasi dalam bidang lainnya seperti perlombaan tingkat nasional, olahraga, penghapal quran dll.

SNMPTN merupakan jalur undangan bagi siswa/i sederajat SMA sebagai calon mahasiswa/i baru yang akan melalui tahap seleksi berdasarkan kompetensi nilai rapor SMA semester 1-5. Siswa yang dapat masuk PTN melalui jalur SNMPTN dianggap sebagai siswa unggul berprestasi secara akademik di sekolahnya [13].

1.2 Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN)

Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) merupakan salah satu jalur untuk masuk PTN berdasarkan nilai rapotnya secara akademis dan prestasi dalam bidang lainnya seperti perlombaan tingkat nasional, olahraga, penghapal quran dll.

SNMPTN merupakan jalur undangan bagi siswa/i sederajat SMA sebagai calon mahasiswa/i baru yang akan melalui tahap seleksi berdasarkan kompetensi nilai rapor SMA semester 1-5. Siswa yang dapat masuk PTN melalui jalur SNMPTN dianggap sebagai siswa unggul berprestasi secara akademik di sekolahnya [13].

1.3 Data Mining

Data Mining merupakan bagian inti dari tahapan proses KDD yang meliputi prediksi algoritma dengan mengeksplor data, membangun model data mining dan menemukan pola klasifikasi yang belum diketahui [14]

Data mining merupakan istilah untuk menemukan berbagai pengetahuan yang ada di dalam suatu database dan masih tersembunyi. *Data mining* merupakan suatu proses penggalian data dengan menggunakan matematika, teknik statistik, pembelajaran mesin (machine learning) dan kecerdasan buatan (*artificial intelligent*) dalam mengidentifikasi, mengekstrasi informasi dan pengetahuan yang bermanfaat terkait suatu database.[15]

Data mining didefinisikan sebagai proses untuk menemukan antara korelasi bermakna baru, trend dan pola oleh sejumlah data besar yang disimpan dalam suatu repositori dengan menggunakan teknik statistik dan teknologi pola serta matematika (Larose, 2005, p. 2). *Data mining* adalah proses analitik dirancang untuk mengeksplorasi sejumlah besar data mencari ilmu tersembunyi yang konsisten dan berharga, langkah pertama terdiri dalam eksplorasi dan data persiapan awal [17]. *Data mining* adalah proses untuk menemukan pola yang ada dalam data [18].

1.4 Algoritma Naive Bayes

Naive Bayes adalah salah satu algoritma klasifikasi sederhana berupa perhitungan probabilistik dengan menghitung probabilitas dari frekuensi dan kombinasi nilai pada kumpulan database yang ada. Teorema Bayes yang digunakan pada algoritma ini mengasumsikan bahwa semua atribut adalah independen karena nilai variabel dari kelas Ini kondisional dengan asumsi bahwa pada kenyataannya jarang berlaku kemerdekaan, maka dari itu karakterisasinya akan diasumsikan Naif namun algoritma Naive Bayes cenderung dapat belajar dengan cepat dan berkinerja baik untuk berbagai masalah *data mining* klasifikasi [19]. Teorema Bayes sebagai berikut [17] :

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Keterangan :

X : Data dengan kelas yang belum diketahui

H : Hipotesis data X suatu kelas spesifik

$(H|X)$: Probabilitas hipotesis H dari kondisi X (*posteriori probability*)

(H) : Probabilitas hipotesis H (*prior probability*)

$(X|H)$: Probabilitas X dari kondisi hipotesis H

(X) : Probabilitas X

Mengingat proses klasifikasi Teorema Bayes memerlukan petunjuk untuk menentukan kelas maka disesuaikan sebagai berikut :

$$P(H|X) = P(X|H).(H) \quad (2)$$

1.5 Algoritma C4.5

Algoritma C4.5 berasal dari algoritma sederhana yang akan membentuk suatu pohon keputusan. Dengan menghitung cara untuk menangani atribut yang bernilai numerik setelah itu, atasi nilai yang hilang. Masalah yang terpenting dalam pemangkasan pohon keputusan adalah bahwa meskipun pohon keputusan yang dibuat oleh suatu algoritma berkinerja baik pada set pelatihan, biasanya algoritma dilengkapi dengan data pelatihan dan tidak digeneralisasi dengan baik pada tes diatur secara bebas. Kemudian metode singkat bagaimana mengubah pohon keputusan menjadi aturan dalam klasifikasi dan perlu memeriksa opsi yang telah disediakan oleh algoritma C4.5 tersebut. Akhirnya, apa yang diterapkan pada sistem CART yang terkenal digunakan untuk mempelajari pohon keputusan dengan klasifikasi dan regresi dalam strategi pemangkasan pohon keputusan [18].

Tahapan algoritma C4.5 sebagai berikut: [20]

- Pilih atribut sebagai simpul akar.
- Buat cabang untuk tiap – tiap nilai.
- Bagi kasus dalam cabang.
- Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.
- Pemilihan atribut sebagai simpul, baik simpul akar(root) atau simpul internal berdasarkan pada nilai Gain tertinggi dari atribut-atribut yang ada

Beberapa label diperbolehkan pada daun pohon khusus, dan dengan rumus perhitungan entropi yang dimodifikasi sebagai berikut [17]

$$\text{Entropy}(D) = -\sum_{j=1}^{q}(p(\lambda j)logp(\lambda j) + q(\lambda j)logq(\lambda j)) \quad (3)$$

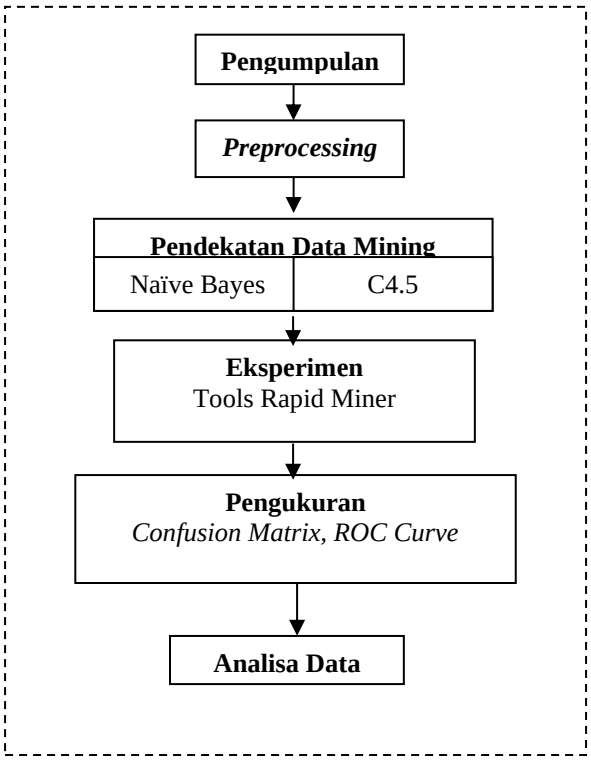
Keterangan :

$p(\lambda j)$ = frekuensi λj

$q(\lambda j) = 1 - p(\lambda j)$. (4)

II. METODE

Penelitian Prediksi Penerimaan SNMPTN ini menggunakan Algoritma C4.5 Dan Naive Bayes dengan melakukan tahapan penelitian seperti pada gambar 2 sebagai berikut :



Gambar 2. Tahapan Penelitian

2.1 Tahap Pengumpulan Data

Tahap ini melakukan pengumpulan data awal yang berupa data primer untuk mendapatkan informasi awal tentang data nilai para siswa/i dan mendeteksi informasi menarik mengenai penegtahuan penting yang masih tersembunyi.

2.2 Tahap *Preprocessing* Data

Untuk memperoleh data berkualitas tinggi, terdapat beberapa teknik yang perlu dilakukan yaitu :

- *Data Cleaning*
  Merupakan proses untuk membersihkan data yang inkonsistensi

- *Data Integration*
  Merupakan proses untuk menggabungkan data data terkait

pISSN: 2442-3386 eISSN: 2442-4293
Vol 9 No 2 Jul – Des 2023, 139 - 144
Melek IT
INFORMATION TECHNOLOGY JOURNAL

- *Data Reduction*
  Merupakan proses pada data yang tidak lengkap untuk dihilangkan

- *Data Transformation*
  Merupakan proses untuk mengubah berntuk data menjadi data yang dibutuhkan dalam proses klasifikasi *data mining.*

2.3 Eksperimen Data Mining

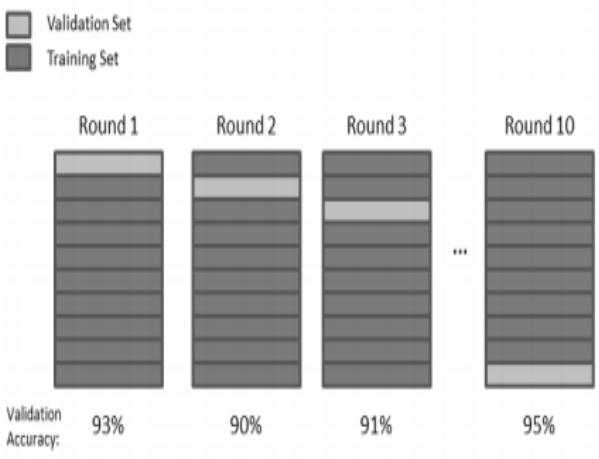
Eksperimen yang dilakukan mengunakan teknologi baik alat bantu berupa *hardware* maupun *software* Rapid miner untuk memproses data. Klasifikasi penerimaan para siswa/i melalui jalur SNMPTN menggunakan atribut-atribut dari data terkait dengan algoritma Naive Bayes dan C4,5dilakukan pada tahap ini.

2.4 Tahap Pengukuran

Tahap ini melakukan pengukuran untuk menguji model penelitian dengan menggunakan *cross validation, performance, confusion matriks, ROC curve.*

- K - Fold *Cross Validation*

  K-fold Cross Validation merupakan salah satu teknik validasi [21] dengan cara membagi data ke dalam K bagian secara acak kemudian masing-masing dari bagian tersebut akan dilakukan proses klasifikasi seperti pada gambar 3 di berikut ini :



Gambar 3. *10-fold cross validation*

- *Confusion Matrix*

  Confusion Matrix merupakan tabel terdiri dari kelas prediksi dan kelas klasifikasi yang memberikan informasi tentang klasifikasi system dengan cara mengevaluasi menggunakan data dalam matriks seperti ditunjukkan pada tabel 1 berikut ini :

Tabel 1. *Confusion Matrix*

CLASSIFICATION	*PREDICTED CLASS*	
	Class=Yes	*Class=No*
Class=Yes	*A (True Positive - tp)*	*B (False Negative - fn)*
Class=No	*C (False Positive - fp)*	*D (True Negative - tn)*

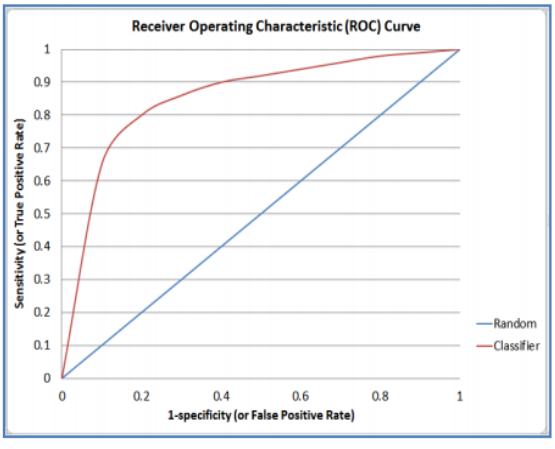
Nilai akurasi dapat dihitung dengan persamaan berikut ini :

$$Akurasi = \frac{tp+tn}{tp+tn+fp+fn} \quad (5)$$

TP (True Positive) = jumlah dokumen dari kelas 1 yang benar diklasifikasikan sebagai kelas 1
TN (True Negative) = jumlah dokumen dari kelas 0 yang benar diklasifikasikan sebagai kelas 0
FP (False Positive) = jumlah dokumen dari kelas 0 yang salah diklasifikasikan sebagai kelas 1
FN (False Negative) = jumlah dokumen dari kelas 1 yang salah diklasifikasikan sebagai kelas 0.

- *ROC Curve*

  ROC curve (Receiver Operating Characteristic) merupakan pengukuran untuk mengevaluasi akurasi klasifikasi secara visual berbentuk kurva. Gambar 4 berikut tampilan dari ROC Curve :



Gambar 4. ROC *curve*

Akurasi 0.90 – 1.00 = *Excellent classification*

Akurasi 0.80 – 0.90 = *Good classification*

Akurasi 0.70 – 0.80 = *Fair classification*

Akurasi 0.60 – 0.70 = *Poor classification*

Akurasi 0.50 – 0.60 = *Failure*.

III. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Data yang diperoleh merupakan data primer dari siswa kelas 12 tahun 2022 yang terdiri atas 268 data dengan 33 atribut termasuk *class label atribute* (atribut output) yaitu atribut Hasil. Dari 268 data terdapat 34 siswa LULUS dan 234 siswa TIDAK. 33 atribut tersebut adalah :

1. NISN (Nomor Induk Siswa Nasional)
2. Nama Siswa
3. Nilai Bahasa Indonesia semester 1-5
4. Nilai Bahasa Inggris semester 1-5
5. Nilai Matematika semester 1-5
6. Nilai Kimia semester 1-5
7. Nilai Biologi semester 1-5
8. Nilai Fisika semester 1-5
9. Hasil Kelulusan

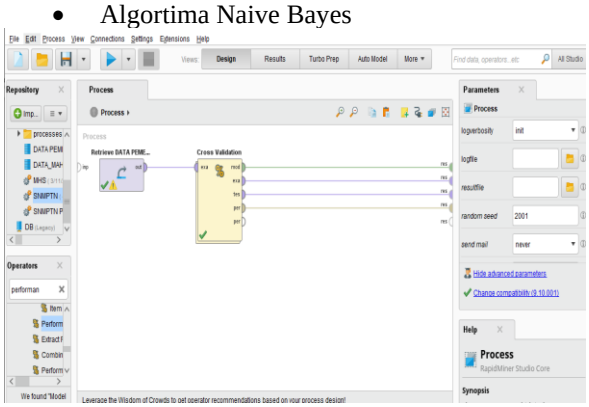
3.2. Data Preprocessing

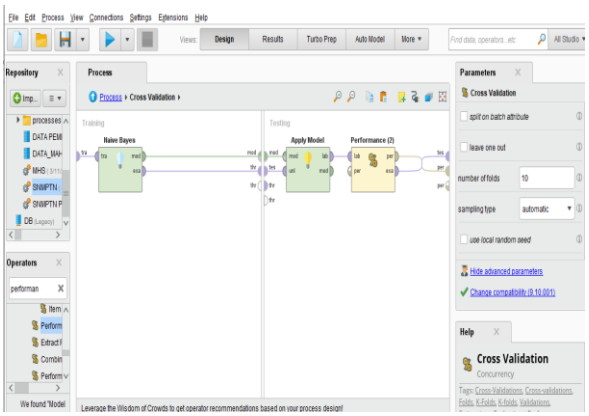
Pada tahap *preprocessing*, data dalam penelitian ini memiliki 33 atribut yang termasuk atribut label "Hasil". Akan tetapi dalam data tersebut terdapat 2 atribut yang memiliki nilai unik (dimana setiap data tidak ada yang sama nilainya) untuk masing-masing siswa sehingga pada penelitian ini mereduksi 2 atribut tersebut dan hanya memakai 31 atribut.

3.3. Eksperimen *Data Mining*

Eksperimen *data mining* klasifikasi prediksi penerimaan SNMPTN dilakukan menggunakan alat bantu *software Rapidminer* 9.1. Hasil eksperimen yang dilakukan seperti yang ditunjukkan gambar 5 berikut ini :

- Algortima Naive Bayes

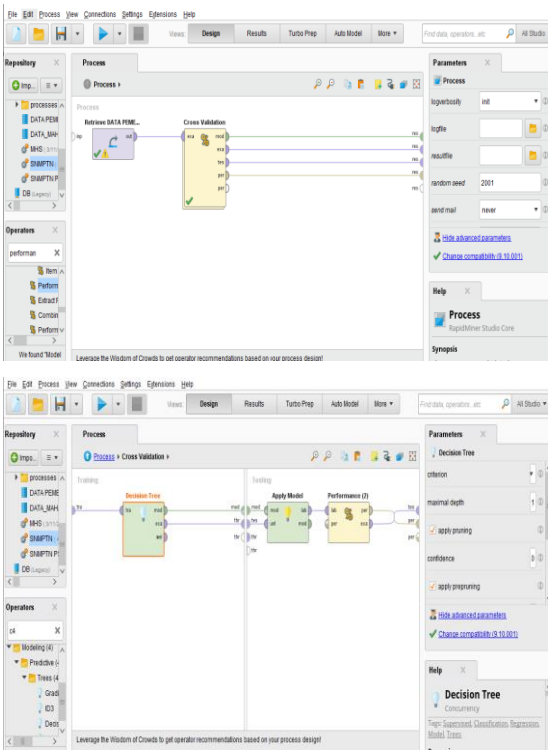




Gambar 5. Model algoritma Naive Bayes

Gambar 4 menunjukkan model *data mining* klasifikasi dengan menggunakan algoritma Naive Bayes dimana untuk menampilkan tingkat keakurasiannya menggunakan parameter *accuracy* dan *Area Under Curve* (AUC) pada operator performance.

- Algoritma C4.5



Gambar 6. Model algoritma C4.5

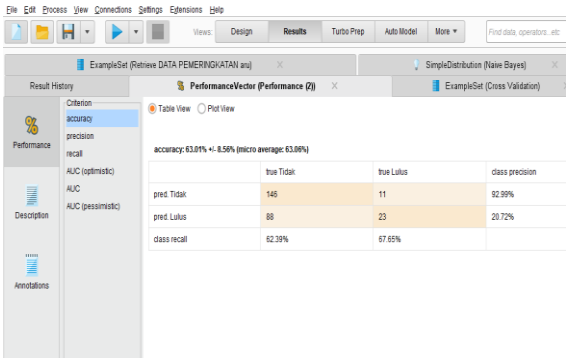
Gambar 6 menunjukkan model data mining klasifikasi dengan menggunakan algoritma C4.5 dimana untuk menampilkan tingkat keakurasiannya menggunakan parameter *accuracy* dan *Area Under Curve* (AUC) pada operator performance.

pISSN: 2442-3386 eISSN: 2442-4293
Vol 9 No 2 Jul – Des 2023, 141 - 144
Melek IT
INFORMATION TECHNOLOGY JOURNAL

3.4. Pengukuran

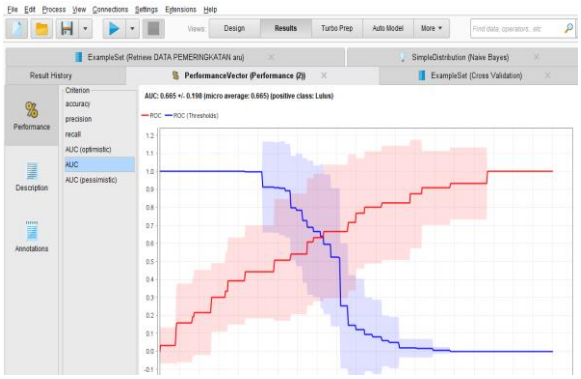
Hasil pengukuran akurasi dan AUC *data mining* klasifikasi prediksi penerimaan SNMPTN dilakukan menggunakan alat bantu *software Rapidminer* 9.1 berikut ini :

- Algoritma Naive Bayes



Gambar 7. Hasil Pengukuran algoritma Naive Bayes

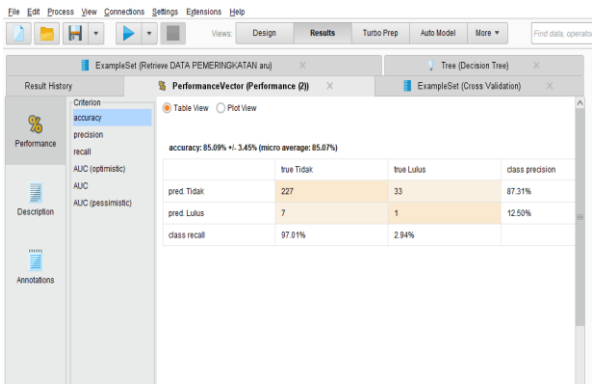
Gambar 7 menunjukkan hasil pengukuran *data mining* klasifikasi algoritma Naive Bayes dimana nilai akurasi nya adalah 63,01 %



Gambar 8. ROC *Curve* algoritma Naive Bayes

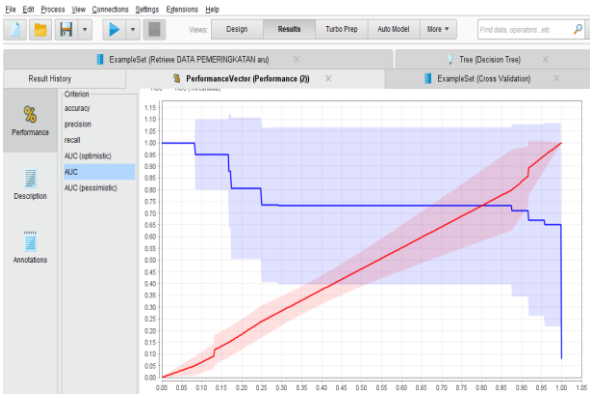
Gambar 8 menunjukkan kurva ROC *data mining* klasifikasi algoritma Naive Bayes dimana nilai AUC adalah 0,665.

- Algoritma C4.5



Gambar 9. Hasil Pengukuran algoritma C4.5

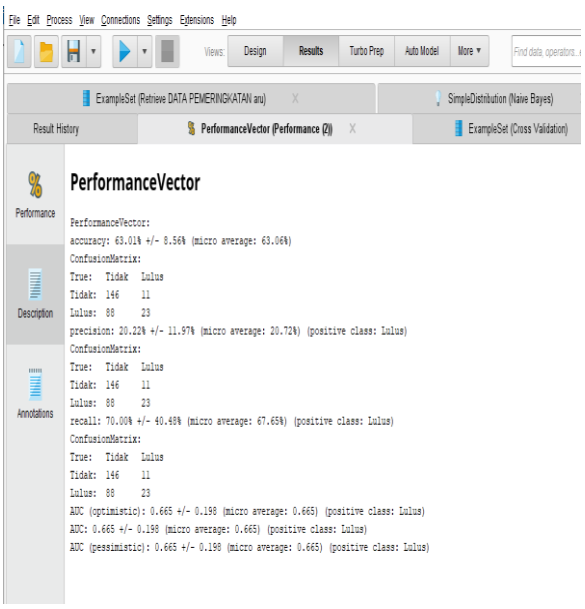
Gambar 9 menunjukkan hasil pengukuran *data mining* klasifikasi dengan algoritma C4.5 dimana nilai akurasi nya adalah 85,09 %



Gambar 10. ROC *Curve* algoritma C4.5

Gambar 10 menunjukkan kurva ROC *data mining* klasifikasi dengan algoritma C4.5 dimana nilai AUC adalah 0,873.

3.5. Pembahasan

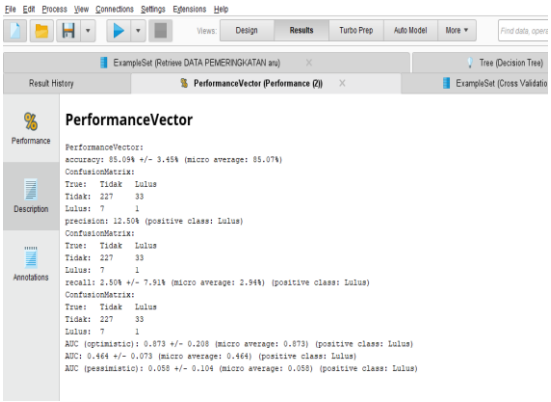


Gambar 11. Performansi algoritma Naive Bayes

Gambar 11 menunjukkan nilai performansi *data mining* klasifikasi algoritma Naive Bayes dimana nilai akurasi adalah 63,01% +/- 8.56% (*micro average*: 63.06%). Perhitungan nilai akurasi berdasarkan persamaan 5, yaitu :

$$\text{Akurasi} = \frac{146 + 23}{146 + 23 + 11 + 88} = 0{,}6306$$

Sedangkan hasil AUC dari algoritma Naive Bayes adalah 0.665 sehingga dapat disimpulkan bahwa model dengan algoritma *Naive Bayes* ini merupakan model *data mining* klasifikasi yang Kurang Bagus untuk digunakan dalam penelitian ini dimana akurasi 0.60 – 0.70 = *Poor classification.*



Gambar 12. Performansi algoritma C4.5

Gambar 12 menunjukkan nilai performansi *data mining* klasifikasi dengan algoritma C4.5 dimana nilai akurasi adalah 85.09% +/- 3.45% (micro average: 85.07%). Berdasarkan persamaan 5, perhitungan nilai akurasi yaitu :

$$\text{Akurasi} = \frac{227 + 1}{227 + 1 + 7 + 33} = 0{,}8507$$

Sedangkan hasil AUC dari algoritma C4.5 adalah 0.873 sehingga dapat disimpulkan bahwa model *C4.5* ini merupakan model *data mining* klasifikasi yang Bagus untuk digunakan dalam penelitian ini dimana akurasi 0.80 – 0.90 = *Good classification*

Berdasarkan analisa hasil evaluasi akurasi dan nilai AUC data mining untuk klasifikasi dengan algoritma Naive Bayes dan C4.5 maka dapat diketahui sebagai berikut :

Tabel 2. Perbandingan Performansi

Pengukuran	C4.5	Naive Bayes
Accuracy	85,09%	63,01%
AUC	0.873	0,665

Tabel 2 perbandingan performansi diatas menunjukkan bahwa dalam penelitian ini prediksi menggunakan d*ata mining* klasifikasi dengan algoritma C4.5 memiliki hasil akurasi yang lebih baik dari algoritma *Naive Bayes.*

Prediksi penerimaan SNMPTN menggunakan data mining klasifikasi akan lebih akurat bila menggunakan algoritma C4.5 sehingga bisa membantu siswa untuk memilih perguruan tinggi negeri favoritnya dengan tepat sesuai hasil pencapain masing-masing siswa.

Dan berdasarkan hasil AUC menunjukkan bahwa algoritma C4.5 merupakan model klasifikasi dengan kriteria “Good” untuk digunakan pada prediksi penerimaan SNMPTN.

IV. PENUTUP

4.1. Kesimpulan

Penelitian ini menggunakan data mining klasifikasi dengan algoritma Naive Bayes dan C4.5 untuk memprediksi penerimaan SNMPTN berdasarkan 268 data siswa 31 atribut nilai siswa semester 1 sampai 5 dan atribut label yaitu Hasil.

Dengan memanfaatkan *software rapid miner* 9.1 untuk eksperimen dan pengukuran menunjukkan bahwa hasil performansi algoritma C4.5 lebih baik dengan akurasi 85,09% dan AUC 0,873 sedangkan performansi algoritma Naive Bayes menghasilkan akurasi 63,01% dan AUC 0,665.

Prediksi penerimaan SNMPTN menggunakan data mining klasifikasi akan lebih akurat bila menggunakan algoritma C4.5 sehingga bisa membantu siswa untuk memilih perguruan tinggi negeri favoritnya dengan tepat sesuai hasil pencapain masing-masing siswa

4.2. Saran

Penelitian selanjutnya dapat menggunakan algoritma seperti *Logistic Regression, Artificial Neural Network* (ANN), *Support Vector Machines* (SVM) dan dapat menggunakan seleksi fitur agar hanya fitur yang memiliki pengaruh yang digunakan.

DAFTAR PUSTAKA

[1] E. Darmawani, “JUANG: Jurnal Wahana Konseling HIV,” *Metod. Ekspositori Dalam Pelaks. Bimbing. dan Konsleing Klasikal*, vol. 1, no. 2, pp. 30–44, 2018.

[2] Lembaga, M. Perguruan, T. Ltmpt, and N. Masuk, “Jumlah Pendaftar SNMPTN 2022 Meningkat , Ini Rinciannya,” p. 2022, 2022.

[3] H. Hasanah, N. A. Sudibyo, and E. Kurniawan, “Prediksi Jurusan pada Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) Menggunakan Metode Klasifikasi Naïve Bayes,” *DoubleClick J. Comput. Inf. Technol.*, vol. 4, no. 1, p. 55, 2020, doi: 10.25273/doubleclick.v4i1.6623.

[4] L. K. Simanjuntak, T. Y. M. Sihite, M. Mesran, N. Kurniasih, and Y. Yuhandri, “Sistem Pendukung Keputusan SNMPTN Jalur Undangan Dengan Metode Electre,” *Jurasik (Jurnal Ris. Sist. Inf. dan Tek. Inform.*, vol. 3,

pISSN: 2442-3386 eISSN: 2442-4293
Vol 9 No 2 Jul – Des 2023, 143 - 144
Melek IT
INFORMATION TECHNOLOGY JOURNAL

no. 3, p. 14, 2018, doi: 10.30645/jurasik.v3i0.63.

[5] J. K. Nazarius, "Analisis Pemilihan Siswa Untuk Jalur SNMPTN dengan Metode Weighted Product (WP) Dan Weighted Sum Model (WSM)," vol. 5, pp. 135–142, 2021.

[6] W. R. Izzati, M. Komarudin, H. D. Septama, and Y. Mulyani, "Analisis Potensi Asal Sekolah pada Jalur Penerimaan Mahasiswa Baru di Universitas Lampung menggunakan Algoritma K-Means," *Electrician*, vol. 13, no. 1, p. 7, 2019, doi: 10.23960/elc.v13n1.2087.

[7] C.-T. Hsieh and C.-S. Hu, "Fingerprint Recognition by Multi-objective Optimization PSO Hybrid with SVM," *J. Appl. Res. Technol.*, vol. 12, no. 6, pp. 1014–1024, 2014, doi: 10.1016/S1665-6423(14)71662-1.

[8] A. Purwanto, A. Primajaya, and A. Voutama, "Penerapan Algoritma C4.5 Dalam Prediksi Potensi Tingkat Kasus Pneumonia Di Kabupaten Karawang," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 4, p. 390, 2020, doi: 10.26418/justin.v8i4.41959.

[9] E. F. Wati and B. Rudianto, "Penerapan Algoritma KNN, Naive Bayes Dan C4.5 Dalam Memprediksi Kelulusan Mahasiswa," *Format J. Ilm. Tek. Inform.*, vol. 11, no. 2, p. 168, 2023, doi: 10.22441/format.2022.v11.i2.009.

[10] R. P. S. Putri and I. Waspada, "Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 1–7, 2018, doi: 10.23917/khif.v4i1.5975.

[11] S. Lestari and A. Suryadi, "Model Klasifikasi Kinerja Dan Seleksidosen Berprestasi Dengan," *Proseding Semin. Bisnis Teknol.*, pp. 15–16, 2014.

[12] N. Khasanah and A. Salim, "Rachman Komarudin 4) , Yana Iqbal Maulana 5) 1) Teknik Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri 2,3) Sistem Informasi, Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika 4) Sistem Informasi, Fakultas Teknologi Informasi, Universitas Nusa Mandiri 5) Teknik Informatika," *Fak. Teknol. Inf.*, vol. 13, no. 3, pp. 207–214, 2022.

[13] N. et al Maulina, "PERBANDINGAN TINGKAT ANSIETAS MAHASISWA KEDOKTERAN YANG DITERIMA MELALUI JALUR SNMPTN, SBMPTN DAN MANDIRI DAN KECENDERUNGAN CABIN FEVER DALAM MELAKSANAKAN UJIAN BLOK PADA MASA PANDEMI," *Syntax Lit. J. Ilm. Indones.*, vol. 6, no. 2, p. 6, 2021.

[14] A. Krisna Ferdinan Leo Simanjuntak, Annita Carolina Br Barus, "Implementasi Metode Decision Tree Dan Algoritma C4.5 Untuk Klasifikasi Kepribadian Masyarakat," *JOISIE J. Inf. Syst. Informatics Eng.*, vol. 5, no. 1, pp. 51–59, 2021.

[15] S. Syahputra, S. Ramadani, A. Manaor, and H. Pardede, "MENENTUKAN STRATEGI PROMOSI MENGGUNAKAN ALGORITMA CLUSTERING K-MEANS didapatkan informasi anggota cluster 1 terdiri dari 164 siswa yang berasal dari kecamatan Kuala sebanyak 75 siswa , dengan asal sekolah terbanyak dari SMP Negeri 1 Salapian sebanyak 21 s," vol. 4, no. 1, pp. 7–14, 2020.

[16] D. T. Larose, *Discovering Knowledge in Data an introduction to data mining*. 2005.

[17] F. Gorunescu, *Data Mining Concepts, Models and Techniques*. Berlin: Springer, 2011.

[18] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical Machine Learning Tools and Techniques*. 2016.

[19] D. Haryono, Y. Zulianda, Wirta, and Lusiana, "Sistem Pendeteksian Serangan Jaringan Local Area Network (Lan) Menggunakan Algoritma Naive Bayes," *JOISIE (Journal …*, vol. 5, no. 1, pp. 1–8, 2021.

[20] K. Umam, D. Puspitasari, and A. Nurhadi, "Penerapan Algoritma C4.5 Untuk Prediksi Loyalitas Nasabah PT Erdika Elit Jakarta," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 65, 2020, doi: 10.30865/mib.v4i1.1652.

[21] I. H. Witten, E. Frank, and M. a. Hall, *Data Mining: Practical Machine Learning Tools and Techniques, Third Edition*, vol. 54, no. 2. 2011.

[Halaman ini sengaja dikosongkan]