

PREDIKSI KELULUSAN MAHASISWA FAKULTAS TEKNIK UNIVERSITAS BINA DARMA MENGGUNAKAN ALGORITMA NAÏVE BAYES

Robi Sepriansyah*¹⁾, Susan Dian Purnamasari²⁾, Kiky Rizky Nova Wardani³⁾, Nasrul Halim⁴⁾

1. Universitas Bina Darma, Palembang, Indonesia
2. Universitas Bina Darma, Palembang, Indonesia
3. Universitas Bina Darma, Palembang, Indonesia
4. Universitas Bina Darma, Palembang, Indonesia

Article Info

Kata Kunci: Klasifikasi, Naïve bayes, Prediksi, Rapid miner, Confusion Matrix

Keywords: Classification, Naïve bayes, Prediction, Rapid miner, Confusion Matrix

Article history:

Received 24 November 2022

Revised 8 December 2022

Accepted 15 December 2022

Available online 1 March 2023

DOI :

<https://doi.org/10.29100/jipi.v8i1.3459>

* Corresponding author.

Corresponding Author

E-mail address:

robi.sepriansyah83@gmail.com

ABSTRAK

Penelitian ini digunakan untuk memprediksi kelulusan mahasiswa fakultas teknik pada universitas bina darma sebagai salah satu penunjang dalam pengambilan keputusan dengan memanfaatkan data mahasiswa yang terdiri dari sks dan Indeks Prestasi Kumulatif. Tahapan penelitian ini menggunakan Knowledge Discovery In Database (KDD) dengan algoritma Naïve Bayes. Berdasarkan penelitian yang telah dilakukan dengan data Training (latih) sebanyak 151 record dan data Testing (uji) sebanyak 120 record. maka didapatkan hasil akurasi penelitian dengan menggunakan perhitungan confusion matrix multiclass bahwa nilai akurasi dari hasil prediksi menggunakan algoritma naïve bayes yaitu 78 mahasiswa yang akan lulus di semester 8 dengan akurasi sebesar 98%, lulus di semester 9 sebanyak 24 mahasiswa dengan akurasi 96%, lulus di semester 10 sebanyak 3 mahasiswa dengan akurasi 100%, dan lulus disemester 12 sebanyak 15 mahasiswa dengan akurasi 98%, Kemudian dari nilai akurasi secara keseluruhan untuk prediksi kelulusan mahasiswa menggunakan algoritma naïve bayes sebesar 95,33%. Berdasarkan hasil dari penelitian menggunakan algoritma naïve bayes maka akurasi tersebut sudah cukup untuk menentukan prediksi kelulusan mahasiswa.

ABSTRACT

This research is used to predict the graduation of students of the faculty of engineering at bina darma universities as one of the supports in decision making by utilizing student data consisting of credits and cumulative achievement indexes. This stage of the study used Knowledge Discovery In Database (KDD) with the Naïve Bayes algorithm. Based on research that has been carried out with Training data (training) as many as 151 records and Testing data (tests) as many as 120 records. then the results of research accuracy were obtained using the multiclass confusion matrix calculation that the accuracy value of the prediction results using the naïve bayes algorithm was 78 students who would graduate in semester 8 with an accuracy of 98%, graduated in semester 9 as many as 24 students with 96% accuracy, graduated in semester 10 as many as 3 students with 100% accuracy, and passed disemester 12 as many as 15 students with 98% accuracy, Then from the overall accuracy score for student graduation predictions using the naïve bayes algorithm of 95.33%. Based on the results of the study using the naïve bayes algorithm, this accuracy is enough to determine student graduation predictions.

PENDAHULUAN

Kelulusan mahasiswa merupakan salah satu bidang studi yang tercakup dalam standar penjaminan mutu internal (SPMI) disetiap perguruan tinggi[1]. Untuk mencapai sebuah kelulusan yang telah ditetapkan di perguruan tinggi untuk dapat menempuh masa studi maksimal delapan semester atau 4 tahun dan total Sks yang telah ditempuh sebanyak 144 sks[2]. Di universitas bina darma proses akreditasinya itu berkaitan dengan kelulusan mahasiswa tersebut, karena penting untuk penilaian akreditasi. Oleh karena itu, untuk mengurangi jumlah mahasiswa yang lulus tepat waktu diperlukan suatu sistem yang menggunakan data dan informasi untuk menilai kelulusan mahasiswa dan memprediksi kelulusan mahasiswa dari awal, agar dapat menentukan kelulusan

mahasiswa diprediksi sejak awal untuk meminimalkan atauantisipasi jumlah kelulusan mahasiswa sehingga dapat mengatasi terjadinya kesulitan di bidang akademik.

Upaya dalam memanfaatkan data mahasiswa salah satunya untuk mengelola data dengan menggunakan teknik data mining yang merupakan proses dalam membuat teknik pengolahan data, teknik penggalian data, sehingga dapat dihasilkan pola-pola tertentu yang menjadi sebuah informasi berdasarkan metode dan algoritma tersebut. Metode klasifikasi dengan algoritma naïve bayes agar dapat menghasilkan suatu informasi berupa prediksi kelulusan mahasiswa. Data mahasiswa tersebut dapat dilakukan panambahan data atau biasa disebut dengan Knowledge Dsccovery in Database (KDD) yang merupakan proses eksternal informasi penting dari basis data yang besar. Maka dari itu berdasarkan data yang tersedia dimana dapat digunakan untuk memprediksi dimasa depan dengan menggunakan pendekatan pendekatan statistic, data yang kompleks dapat menerapkan kemampuan untuk mengekstrak data tersebut untuk diolah menjadi sebuah informasi yang lebih penting.

Pada saat ini teknologi informasi menjadi peran yang sangat penting untuk keperluan di setiap perusahaan, organisasi maupun di institusi pendidikan dalam mengelola proses bisnisnya. Karena berkembangnya teknologi yang semakin pesat didalam pekerjaan yang sebelumnya sulit untuk diprediksi dengan adanya teknologi pekerjaan tersebut menjadi mudah untuk diprediksi, maka dengan memanfaatkan teknologi terutama dalam bidang pendidikan dan dibantu adanya data proses prediksi tidak lagi menjadi bagian yang sulit[3]

Disetiap perguruan tinggi tentunya memiliki database untuk menyimpan semua data mahasiswa disetiap akademik, data mahasiswa tersebut terus menerus bertambah disetiap tahun dan menumpuk seperti data yang terabaikan kerana data tersebut tidak dimanfaatkan atau diolah kembali, sedangkan dari data mahasiswa tersebut dapat diolah agar menghasilkan sebuah informasi yaitu mengenai tingkat kelulusan mahasiswa sehingga dapat meminimalisir keterlambatan kelulusan.

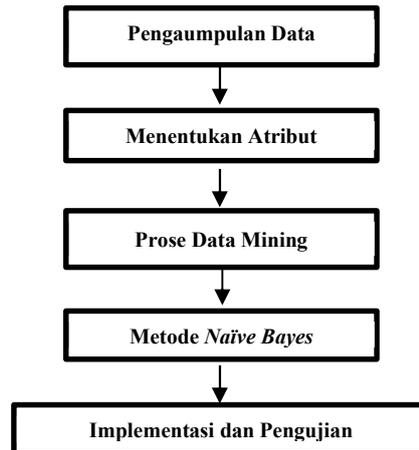
Dari hasil penelitian sebelumnya mengenai prediksi kelulusan mahasiswa menggunakan algoritma pohon keputusan C 4.5 dan software weka, hasil yang didapatkan dengan pengukuran akurasi confusion matrix sebesar 73% dikatakan tidak cukup baik, hal tersebut dikarenakan atribut pada data yang kurang menceritakan kondisi yang sebenarnya[4]. Penelitian dalam predikis klasifikasi masa studi mahasiswa menggunakan algoritma K-Nearest Neighbor dengan metode evaluasi *K-Fold cross validation* dengan melakukan pengujian 6 scenario terhadap atribut-atribut yang digunakan dan untuk mendapatkan akurasi terbaik dari algoritma yang digunakan sebesar 75.95%[5]. Berdasarkan penelitian sebelumnya untuk memprediksi kelulusan mahasiswa menggunakan model algoritma C 4.5 dan KNN. Maka dari itu untuk mendapatkan hasil akurasi dari model lain untuk prediksi kelulusan mahasiswa penelitian ini menggunakan model algoritma *naïve bayes* dan pengukuran akurasi hasil prediksi kelulusan persemester menggunakan *confusion matrix*.

Naïve bayes merupakan salah satu algoritma pembelajaran yang paling efektif dalam *data mining*, *naïve bayes* memiliki performa yang kompetitif di proses klasifikasi bahkan dengan asumsi tidak ada kaitan antara atribut[6]. *Naïve bayes* juga dapat memberikan sebuah kontribusi yang sama penting terhadap pemilihan kelas tertentu [7] juga merupakan algoritma yang dapat digunakan dalam mengklasifikasikan dari variable tertentu dengan metode probabilitas dan statistik untuk mencari peluang terbesar berdasarkan prediksi klasifikasi[8].

Berdasarkan penjelasan dari permasalahan di atas, prediksi kelulusan mahasiswa memanfaatkan data mining menggunakan algoritma naïve bayes metode klasifikasi di aplikasi rapid miner dimana bertujuan untuk dapat membantu mendapatkan informasi prediksi kelulusan mahasiswa Fakultas teknik Universitas Bina Darma sehingga dapat memberikan informasi bagi program studi untuk memprediksi status kelulusan serta dapat digunakan untuk menentukan langkah dan kebijakan bagi fakultas dalam membidik kelulusan mereka.

METODE PENELITIAN

Penelitian ini bertujuan untuk melakukan pengujian terhadap algoritma naïve bayes yang digunakan untuk mengklasifikasi prediksi kelulusan mahasiswa Fakultas Teknik Universitas Bina Darma, dengan memanfaatkan aplikasi rapid miner. Rapid miner adalah aplikasi data mining yang memiliki tampilan GUI (Graphical User Interface), open source dan dapat dijangkau di system manapun sehingga dapat membantu peneliti dalam menggunakannya[9] Tahapan proses penelitian ini dapat dilihat pada gambar 1.



Gambar 1 Tahapan Penelitian

A. Pengumpulan Data

Metode pengumpulan data dilakukan menggunakan data primer karena data yang digunakan merupakan data yang diperoleh secara langsung dari Universitas Bina Darma dimana peneliti melakukan observasi selain itu juga menggunakan studi pustaka dan literatur yang berkaitan dengan tema penelitian.

1) Pengamatan (*Observasi*)

Tahapan pengumpulan informasi terkait penelitian dengan cara mengamati secara langsung suatu obyek selama waktu tertentu. Yang dilakukan di ruangan pengolahan data Universitas Bina Darma untuk mendapatkan data yang dibutuhkan dalam penelitian. Data-data yang sudah dikumpulkan kemudian disimpan dalam bentuk excel agar dapat diolah menggunakan data mining.

2) Studi pustaka

Tahapan ini dilakukan pengambilan informasi berdasarkan buku, jurnal maupun hasil penelitian sebelumnya yang berkaitan dengan permasalahan, sehingga dapat dijadikan sumber informasi yang memiliki kaitan dengan penelitian.

3) Studi literatur

Studi literatur dimanfaatkan untuk menambah sumber informasi yang berhubungan dengan teori yang sesuai dengan permasalahan penelitian, metode ini digunakan untuk menelusuri, membandingkan masalah berdasarkan literatur yang didapatkan, dan menjadikannya bahan pokok penelitian, studi literatur ini memberikan orientasi yang luas dan menghindari terjadinya duplikasi penelitian yang telah dilakukan sebelumnya[6]

B. Menentukan Atribut

Himpunan data mahasiswa Fakultas teknik Universitas Bina Darma memiliki data primer yang diperoleh melalui pengamatan secara langsung di ruangan bagian pengolahan data. Atribut yang diperoleh dari data ini adalah 1 atribut tujuan dan atribut lain sebagai pendukung. Atribut tersebut dapat dilihat pada tabel 1.

TABEL I.
 ATRIBUT DATA MAHASISWA

No	Atribut	Type	Description
1	Nim	Teks	Nomor Induk Mahasiswa
2	Sks 1	Numeric	Sks semester 1
3	Sks 2	Numeric	Sks semester 2
4	Sks 3	Numeric	Sks semester 3
5	Sks 4	Numeric	Sks semester 4
6	lpk 1	Numeric	lpk Semester 1
7	lpk 2	Numeric	lpk Semester 2
8	lpk 3	Numeric	lpk Semester 3
9	lpk 4	Numeric	lpk Semester 4
10	Keterangan	Categorical	Status Kelulusan

C. Proses Data Mining

Data mining adalah teknik yang menggunakan statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengidentifikasi informasi yang berguna dalam sumber data yang besar[10]. Penelitian ini adalah menerapkan klasifikasi data mining untuk memprediksi kelulusan mahasiswa menggunakan algoritma Naïve bayes

untuk memprediksi lama semester. Klasifikasi dalam data mining adalah pembelajaran untuk memprediksi suatu nilai berdasarkan dari beberapa atribut dengan menggunakan rule sebagai label untuk dapat memprediksi kelas dari data yang akan diprediksi[2]. metode Knowledge Discovery in Database (KDD) yang mempunyai isi berbeda beda tetapi memiliki arti yang saling berhubungan antara tahap satu dengan tahap yang lainnya[11] *Knowledge Discovery in Database* adalah tahapan yang bertujuan untuk menggali sebuah informasi pada sejumlah data yang besar untuk mendapatkan pengetahuan yang berguna dan dapat mudah dipahami[12] Dan diolah menggunakan data mining algoritma Naïve Bayes dengan memanfaatkan aplikasi Rapid Miner untuk megatahui hasil prediksi dan tingkat akurasi kelulusan mahasiswa.

1) Data Selection

Data awal dalam penelitian ini berupa data angkatan 2014 dan 2019 dalam bentuk file *excel* yang terdiri dari NIM, Jenis kelamin, Asal sekolah, sks semester 1 – 4, IP semester 1-4 dan keterangan lulus. Selanjutnya dilakukan proses data selection untuk memilih data yang akan digunakan. Dalam penelitian ini data yang akan digunakan NIM, Sks, IP, dan keterangan lulus. Data yang digunakan dalam penelitian ini dapat dilihat pada tabel 2 untuk data angkatan 2014 dan data angkatan 2019 dapat dilihat pada tabel 3.

TABEL II
 DATA SAMPLE ANGGKATAN 2014

NIM	Tanggal Lahir	Alamat	Sks 1	Sks 2	Sks3	Sks 4	Ip 1	Ip 2	Ip 3	Ip 4	Keterangan Lulus
141710005	07/10/1995	JL. SWADAYA LR. CENDANA NO. 2034 RT. 010 RW. 11	16	17	16	16	3,25	3,14	2,95	2,62	9
141710006	11/11/1997	DESA KARANG MULIA,KEC LUBAI ULU, KAB MUARA ENIM	20	19	18	16	3,25	3,85	3,6	3,29	8
141710007	23/06/1997	jln tanjung pandan lrg nusantara 2 no.587 RT 06 RW 02	19	19	19	20	3,19	3,46	3,22	3,25	8
141710008	16/07/1995	JL. NURUL YAQIN RT 023 RW 005 NO 1546 PALEMBANG	20	21	21	19	3,54	3,9	3,79	3,61	8
141710009	13/06/1996	Desa Tanjung Kukuh, Kecamatan Semendawai Barat, Kabupaten OKU Timur	19	19	19	20	3	3,25	3,18	3,04	8
141710010	12/10/1996	JL.RAWASARI LR.PURNAMA RT/RW 48/10 PALEMBANG	20	21	19	21	3,85	3,55	3,81	3,65	8
141710011	29/05/1993	JL. RAMBUTAN DALAM LR. BUNTU RT 031 RW 011 KEL 30 ILIR KEC ILIR BARAT II KODE POS : 30144 PALEMBANG	20	19	19	20	3.01	3,14	3,47	3,82	8

TABEL III
 DATA SAMPLE ANGGKATAN 2019

NIM	Tanggal Lahir	Alamat	Sks 1	Sks 2	Sks3	Sks 4	Ip 1	Ip 2	Ip 3	Ip 4
191710002	28/08/1999	JL. SUKABANGUN 1 KOMP. SARJANA BLOK B NO 5 RT 023 RW 004	14	32	42	60	3,29	2,94	2,83	2,78
191710003	02/03/1996	Jl. Bukit Menur I/B 387 RT.006 RW.020 Kelurahan Sendangmulyo Kecamatan Tembalang	20	41	60	80	3,75	3,8	3,67	3,54
191710004	22/12/1995	JL. EKA JAYA RT 006 RW 003 DESA SUNGSANG I KECAMATAN BANYUASIN II	20	41	60	80	3,65	3,54	3,53	3,46

191710006	05/10/1989	JLN. INPRES NO. 01 RT.001 RW.001 KEL. CINTA MANIS BARU KEC. AIR KUMBANG KODE POS. 30763 KAB. BANYUASIN	20	41	60	80	3,35	3,54	3,38	3,28
191710007	08/07/1995	JL. SUKABANGUN 2 LR. CEMARA NO.97 RT.006 RW.002 KEL.SUKAJAYA KEC. SUKARAMI KOTA PALEMBANG	18	41	57	77	3,5	3,17	3,14	3,06
191710008	17/05/1995	JLN. KAPTEN ABDULAH LG SIMPANG 4 BAKARAN RT.005 RW.002 KEL. TALANG PUTRI PLAJU	20	41	60	80	3,15	3,12	3	3,03
191710009	18/05/2000	JL. TAPAK SIRING LG. BEDENG RT.024 RW.009 KEL. TALANG BETUTU KEC. SUKARAMI KODE POS. 20155	20	41	60	80	3,6	3,54	3,38	3,36

2) Pre-processing

Sebelumnya data angkatan 2014 terdiri dari 230 record dan 2019 terdiri dari 220 record. Setelah dilakukan *pre-processing* terdapat data yang memiliki *missing value*, duplikat data, dan *noise* pada data angkatan 2014 sebanyak 79 record dan 2019 sebanyak 100 record.

3) Transformation

Setelah melalui tahapan *pre-processing* maka didapatkan data sebanyak 151 record pada data angkatan 2014 dan 120 record pada angkatan 2019. Format data yang telah didapatkan sudah sesuai untuk dilakukan proses data mining pada aplikasi rapid miner.

4) Penambangan Data (Data mining)

Penambangan data pada aplikasi rapid miner menggunakan operator *read excel*, *cross validation*, *naïve bayes*, *apply model*, dan *performance*.

5) Interpretation/Evaluation

Memahami apakah pola atau informasi yang dihasilkan bertentangan dengan fakta. Memberikan informasi yang dihasilkan berdasarkan proses penambangan data dalam format yang mudah dimengerti oleh pihak yang membutuhkannya.

D. Naïve Bayes

Naïve Bayes adalah salah satu algoritma pembelajaran induktif yang paling efektif dalam pembelajaran mesin untuk penambangan data dan data mining[6]. dengan proses klasifikasi probabilistik dengan perhitungan dari sekumpulan probabilitas memanfaatkan jumlah frekuensi dan kombinasi nilai dari *dataset* yang sudah disediakan [13] Klasifikasi Naïve bayes menghitung probabilitas menjumlahkan nilai dari dataset yang sudah disediakan, menggunakan naive bayes memiliki ketentuan disetiap jenis atribut tidak memiliki ketergantungan antara satu sama lain[14]. Kelebihan dari algoritma naïve bayes adalah hanya memerlukan beberapa data latih (training) untuk menguji parameter *mean* serta *varians* dari atribut yang dibutuhkan untuk klasifikasi. Teorema naïve bayes mempunyai bentuk sebagai berikut [15]:

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

Keterangan

X : Data dengan *class* tidak diketahui

H : Hipotesis data dari suatu *class* spesifik

P(H|X) : Probabilitas *hipotesis* H berdasar kondisi X (*posterior probability*)

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

P(X|H) : Probabilitas X berdasarkan kondisi hipotesis H

P(X) : Probabilitas X

Menghitung masing masing class
Prior Probability

$P(C_i)$: Probabilitas Hipotesis C_i

Langkah awal yang akan dilakukan pengelompokkan data training untuk hasil yang dicapai yaitu Tepat waktu atau terlambat mahasiswa tersebut, lalu akan dilakukan perhitungan nilai probabilitas berdasarkan pengelompokkan data. nilai probabilitas tersebut.

Tepat waktu | class = 80 | 151 = 0,26
 Terlambat | class = 71 | 151 = 0,47

Berdarkan perhitungan di atas probabilitas tepat waktu sebanyak 80 dibagi dengan class atau jumlah keseluruhan data training sebanyak 151 dan mendapatkan hasil 0,26, sedangkan probabilitas terlambat sebanyak 71 dibagi dengan class jumlah keseluruhan data sebanyak 151 mendapatkan hasil 0,47, hasil tersebut akan digunakan untuk perhitungan nilai *posterior probability* dengan rumus berikut:

Posterior Probability

$P(X|C_i)$: Probabilitas X menurut class

Dari hasil perhituangan yang sudah dilakukan *prior probability* untuk menentukan jumlah *class* lalu akan dilakukan pengujian menggunakan contoh kasus berdasarkan data yang sudah ada.

Jika nilai X memiliki jumlah Sks disemester 4 sebanyak 66 dengan Ip sementara 3,05?

X = Sks 66 dan Ip sementara 3,05 ?
 Sks 66 | Tepat Waktu = 4/80
 Sks 74 | Terlambat = 6/71
 Ip > 3 | Tepat Waktu = 73/80
 Ip < 3 | Terlambat = 43/71

Perhitungan *posterior probability*

Tepat waktu = $80/151 \times 4/80 \times 73/80$
 $= 0,26 \times 0,05 \times 0,91$
 $= 0,011$

Berdasarkan perhitungan class tepat waktu menggunakan sks 66 lulus tepat waktu, dan memiliki nilai ipk lebih dari 3 mendapatkan hasil 0,011.

Terlambat = $71/151 \times 6/71 \times 43/71$
 $= 0,47 \times 0,08 \times 0,6$
 $= 0,022$

Berdasarkan perhitungan yang telah dilakukan pada class terlambat menggunakan 66 terlambat dan memiliki nilai ipk kurang dari 3, mendapatkan hasil 0,022. Selanjutnya akan dilakukan perhitungan untuk memaksimalkan nilai dari class yang sudah didapatkan dengan menggunakan rumus :

$P(X|C_i) P(C_i)$

Fungsi dari rumus ini yaitu untuk memaksimalkan dari nilai *posterior probability* berdasarkan class tepat waktu dan terlambat , menghitung nilai X dengan class yang sudah didapatkan,

Tepat waktu
 $P(X | \text{Tepat waktu}) \times P(\text{Tepat waktu} : \text{class})$
 $= 0,011 \times 80/151$
 $= 0,0058$

Hasil yang telah dilakukan dari pengujian nilai *posterior* dan nilai class tepat waktu yaitu 0,058. Selanjutnya dilakukan pada pengujian posterior terlambat dan *class* terlambat. sebagai berikut:

Terlambat
 $P(X | \text{Terlambat}) \times P(\text{Terlambat} : \text{class})$
 $= 0,022 \times 71/151$
 $= 0,0103$

Hasil perhitungan yang telah dilakukan terhadap class terlambat yaitu 0,0103, berdasarkan hasil perhitungan untuk memaksimalkan nilai dari *posterior probability* dari class tepat waktu dan terlambat nilai class tepat waktu lebih kecil dari nilai class terlambat. Maka kasus dari data tersebut diprediksi masuk ke kelas terlambat.

Pada penelitian ini untuk menghitung akurasi prediksi dari algoritma *naïve bayes* menggunakan Confusion Matrix. Confusion matrix memberikan informasi dari perbandingan hasil klasifikasi yang telah dilakukan untuk memprediksi dengan bentuk klasifikasi yang sebenarnya, hasil pada confusion matrix menunjukkan jumlah data yang di uji diklasifikasikan benar maupun data uji diklasifikasi salah [16].

Rumus perhitungan Confusion Matrix

$$(f1_{score}) : 2 \times \frac{precision \times recall}{precision + recall}$$

Dimana :

Precision = nilai prediksi positif tiap semester

Recall = nilai sensitifitas tiap semester

HASIL DAN PEMBAHASAN

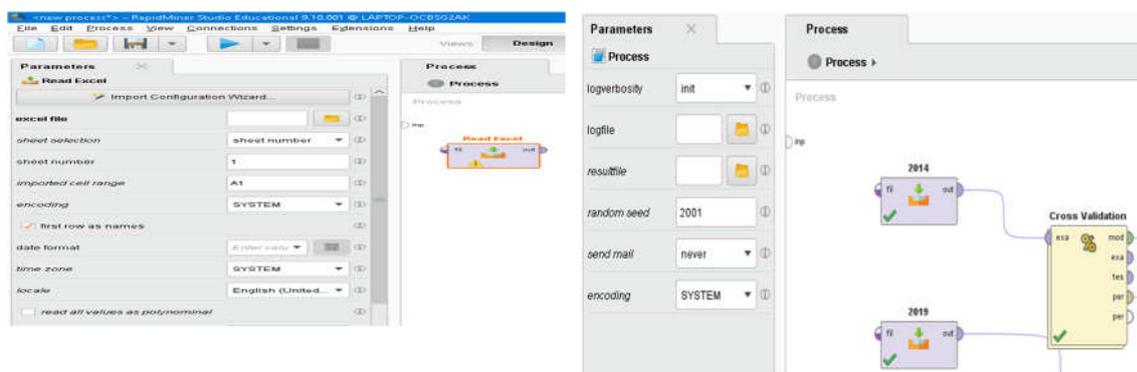
A. Implementasi dan Pengujian

Rapid miner merupakan alat yang dikembangkan oleh Markus Hofman yang memiliki tampilan GUI (*Graphical User Interface*), *Open Source*, dan dibuat menggunakan program Java dibawah lisensi Publik dapat berjalan pada sistem apapun sehingga dapat mempermudah bagi pengguna dalam pengoperasiannya[17] Penelitian ini menggunakan aplikasi rapid miner untuk melakukan pengujian prediksi kelulusan mahasiswa dengan memanfaatkan data training (data latih) dari alumni tahun 2014 yang berjumlah 151 data berisi 10 atribut di antaranya Nim, Sks semester 1 sampai 4, dan tahun kelulusan sebagai label. Data testing (data uji) dari angkatan tahun 2019 berjumlah 120 dengan atribut Nim, Sks dan Ipk. Data tersebut diuji menggunakan algoritma *naïve bayes* pada aplikasi rapid miner 9.10

B. Pembentukan Model

Di dalam rapid miner memiliki operator – operator yang dapat digunakan dalam proses datamining, berikut operator – operator yang digunakan pada proses penelitian ini, antara lain :

- 1) Operator *read excel*, operator ini melakukan pengambilan data *excel*, membaca file *excel* dan menampilkan data seperti table yang telah dipilih.
- 2) *Cross validation* proses ini digunakan untuk memprediksi seberapa besar akurasi pada suatu model dalam *data mining*. Proses tersebut dapat dilihat pada gambar 2



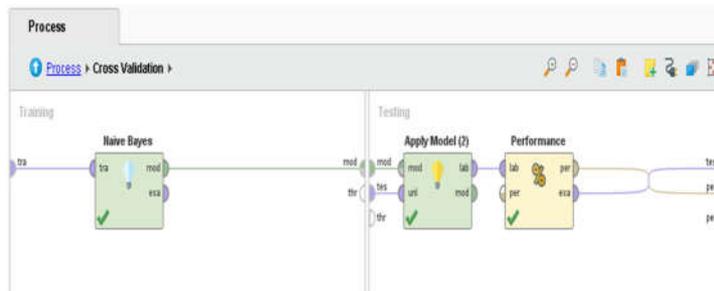
Gambar 2. Cross Validation

Operator cross validation merupakan operator yang bersarang yang memiliki dua subprocess yaitu subprocess training yang digunakan untuk melatih model, pada proses training digunakan untuk memilih metode algoritma datamining, pada penelitian ini menggunakan algoritma *naïve bayes* maka pada subprocess *training* terdapat operator *naïve bayes* dan subprocess testing untuk melakukan pengujian model sekaligus mengukur kinerja pada model yang digunakan[18].

- 3) *Naive Bayes* proses ini akan dilakukan sesuai dengan metode yang digunakan pada penelitian.

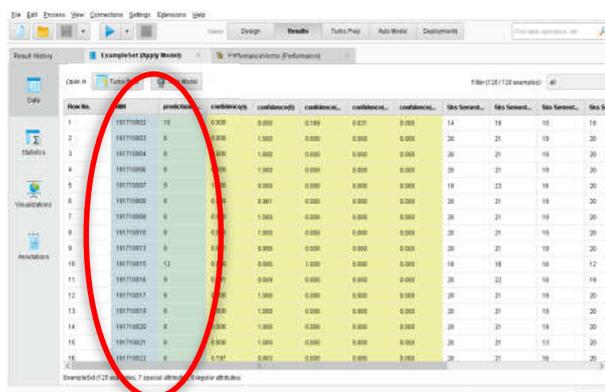
- 4) *Apply Model* tujuan dari menambahkan operator *apply model* untuk memperoleh prediksi pada data *training* dan pada data *testing*. Serta untuk penghubung antara model algoritma *naive bayes* dengan model *performance*. proses ini merupakan proses yang paling penting karena harus memiliki persamaan urutan, variable, dll antara data *training* dan data *testing*.
- 5) *Performance* , diproses ini berfungsi untuk menghitung jumlah akurasi pada suatu model yang akan di uji.

Pada subproses dari *cross validation* yaitu proses *Naive bayes* di training, serta *apply model* dan ditambah *performance* di testing. Proses tersebut dapat dilihat pada gambar 3.



Gambar 3. Model Subproses *Naive Bayes*

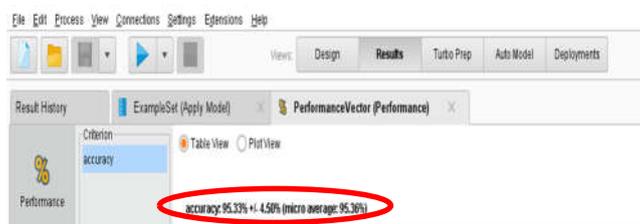
Hasil prediksi yang telah dilakukan pada aplikasi Rapidminer untuk memprediksi kelulusan mahasiswa dengan memanfaatkan data training (latih) dan data testing (uji) menggunakan model algoritma *Naive bayes*, dapat dilihat pada gambar 4.



Case #	Actual	Predicted	Confidence	Cost (actual)	Cost (predicted)	Cost (error)	Cost (weight)	Cost (total)	Cost (weight)	Cost (total)	Cost (weight)	Cost (total)
1	10	10	0.900	0.000	0.100	0.000	0.000	14	14	10	14	
2	8	8	0.900	1.000	0.000	0.000	0.000	20	20	13	20	
3	6	6	0.900	1.000	0.000	0.000	0.000	20	20	19	20	
4	9	9	0.900	1.000	0.000	0.000	0.000	20	20	19	20	
5	9	9	0.900	0.000	0.000	0.000	0.000	18	23	18	20	
6	8	8	0.900	0.000	0.000	0.000	0.000	20	20	18	20	
7	6	6	0.900	1.000	0.000	0.000	0.000	20	20	19	20	
8	9	9	0.900	0.000	0.000	0.000	0.000	20	20	19	20	
9	12	12	0.900	0.000	1.000	0.000	0.000	18	18	16	12	
10	9	9	0.900	0.000	0.000	0.000	0.000	20	20	18	20	
11	9	9	0.900	0.000	0.000	0.000	0.000	20	20	18	20	
12	9	9	0.900	1.000	0.000	0.000	0.000	20	20	19	20	
13	9	9	0.900	1.000	0.000	0.000	0.000	20	20	18	20	
14	8	8	0.900	1.000	0.000	0.000	0.000	20	20	18	20	
15	9	9	0.900	1.000	0.000	0.000	0.000	20	20	18	20	
16	9	9	0.900	1.000	0.000	0.000	0.000	20	20	18	20	
17	12	12	0.900	0.000	0.000	0.000	0.000	20	20	18	20	
18	9	9	0.900	0.000	0.000	0.000	0.000	20	20	18	20	

Gambar 4. Hasil Prediksi Kelulusan Semester

Data yang digunakan yaitu data uji dengan 120 examples dan 8 atribut diantaranya Nim, Ip semester 1 sampai dengan 4 dan Ipk Semester 1 sampai dengan 4 untuk menentukan hasil prediksi mahasiswa dengan hasil kelulusan prediksi bahwa yang lulus disemester 8 sebanyak 78, lulus di semester 9 sebanyak 24, sedangkan yang lulus disemester 10 sebanyak 3, dan mahasiswa yang lulus disemester 12 sebanyak 15 orang sementara dari hasil untuk pengujian kedua data tersebut dengan hasil akurasi yang sempurna sebesar 95,36%. Dapat dilihat pada gambar 5.

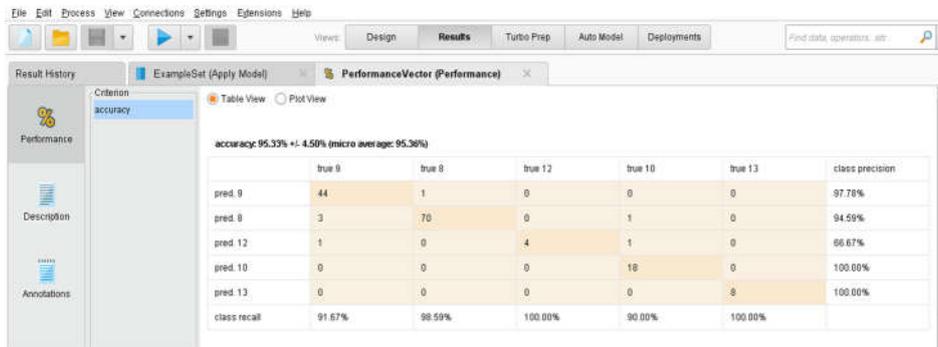


Criterion	Value
accuracy	accuracy: 95.33% +/- 4.50% (micro average: 95.36%)

Gambar 5. Hasil Akurasi Dari Pengujian Data

C. Pengujian *Confusion Matrix Multiclass*

Merupakan proses evaluasi dengan pengujian *confusion matrix*. *Confusion matrix* itu table yang memberikan tampilan hasil dari suatu kerja berdasarkan suatu klasifikasi. Menghitung akurasi adalah hasil yang paling terpenting pada suatu penelitian agar penelitian tersebut dapat diketahui keberhasilan dan kegagalan terhadap penelitian.



	true 8	true 9	true 12	true 10	true 13	class precision
pred. 8	44	1	0	0	0	97.70%
pred. 9	3	70	0	1	0	94.59%
pred. 12	1	0	4	1	0	66.67%
pred. 10	0	0	0	16	0	100.00%
pred. 13	0	0	0	0	8	100.00%
class recall	91.67%	98.59%	100.00%	90.00%	100.00%	

Gambar 6. Hasil Confusion Matrix Multi Class

Langkah perhitungan *Confusion Matrix Multiclass*

$$\text{Prediksi 8 (f1 score)} : 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

$$= 2 \times \frac{94,59\% \times 98,59\%}{94,59\% + 98,59\%} = 2 \times \frac{0,93}{1,88}$$

$$= 2 \times 0,49$$

$$= 0,98$$

$$= 98\%$$

Berdasarkan hasil perhitungan dari *confusion matrix multiclass* untuk mengukur kinerja klasifikasi prediksi kelulusan yaitu sebesar 98% yang diprediksi lulus di semester 8. Hasil perhitungan akurasi dari semester 8 – 12 dapat dilihat pada tabel 2.

TABEL 4. HASIL PERHITUNGAN AKURASI TIAP SEMESTER

Prediksi	Hasil Akurasi
Semester 8	98%
Semester 9	96%
Semester 10	100%
Semester 12	98%

F1 score merupakan perhitungan Confusion Matrix Multiclass untuk membandingkan nilai rata-rata presisi (*nilai prediksi positif*) dan recall (*sensitivitas*) mendapatkan ukuran akurasi. Berdasarkan dari perhitungan yang telah dilakukan menggunakan nilai *precision* dan *recall* dengan hasil *f1-score* dengan akurasi prediksi lulus di semester 8 sebesar 98%, nilai akurasi *f1-score* dari prediksi lulus di semester 9 dengan score 96%, hasil akurasi *f1-score* untuk prediksi lulus di semester 10 dengan hasil 100% dan hasil akurasi *F1-Score* prediksi lulus di semester 12 sebesar 98%. Berdasarkan penelitian terkait pada penelitian ini menggunakan lagoritma naïve bayes untuk prediksi kelulusan mahasiswa mendapatkan nilai akurasi 88,16% dengan data *training303* record dan data *testing* 76 record, adapun atribut yang digunakan nama, status mahasiswa, status perkawinan, Ip sementara, IPK, dan status kelulusan.

KESIMPULAN

Berdasarkan dari hasil pengujian yang sudah dilakukan dan telah diuraikan dalam penerapan data mining untuk memprediksi kelulusan mahasiswa pada Fakultas Teknik Universitas Bina Darma , maka penulis dapat menarik kesimpulan sebagai berikut:

1. Penerapan data mining dengan menggunakan model algoritma Naive Bayes yang sudah dilakukan dapat menghasilkan sebuah informasi mengenai memprediksi kelulusan mahasiswa fakultas teknik yang lulus disemester 8 sebanyak 78, lulus di semester 9 sebanyak 24, sedangkan yang lulus disemester 10 sebanyak 3, dan mahasiswa yang lulus disemester 12 sebanyak 15 orang dengan tingkat akurasi prediksi sebesar 95,33%. Sudah cukup akurat untuk dapat digunakan memprediksi kelulusan mahasiswa.
2. Menghasilkan informasi mengenai data kelulusan mahasiswa di Fakultas Teknik Universitas Bina Darma. Yang dihasilkan dari penerapan data mining dalam aplikasi rapid miner menggunakan algoritma naïve bayes.
3. Perhitungan yang sudah dilakukan secara teoritis menghasilkan informasi nilai data klasifikasi yang telah ditentukan.

SARAN

Berdasarkan hasil kesimpulan yang telah diuraikan di atas, maka penulis ada beberapa saran yang ingin penulis sampaikan, yaitu :

1. Pada penelitian selanjutnya disarankan dapat menggunakan dataset yang berbeda dan memiliki jumlah data yang lebih besar serta variabel atau atribut sehingga data kelulusan mahasiswa dan tingkat akurasi yang lebih baik.
2. pada penelitian seterusnya dapat mencoba untuk menggunakan teknik, dan model algoritma maupun aplikasi data mining yang lain sehingga dapat menghasilkan sebuah informasi yang lebih variatif.

DAFTAR PUSTAKA

- [1] M. F. Abdullah dan M. R. Arief, "Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Menggunakan Metode Support Vector Machine," vol. 14, no. 1, hlm. 10, 2022.
- [2] S. Alim, "IMPLEMENTASI ORANGE DATA MINING UNTUK KLASIFIKASI KELULUSAN MAHASISWA DENGAN MODEL K-NEAREST NEIGHBOR, DECISION TREE SERTA NAIVE BAYES ORANGE DATA MINING IMPLEMENTATION FOR STUDENT GRADUATION CLASSIFICATION USING K-NEAREST NEIGHBOR, DECISION TREE AND NAIVE BAYES MODELS," vol. 6, no. 2, hlm. 12, 2021.
- [3] M. Y. Putra dan D. I. Putri, "Pemanfaatan Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Klasifikasi Jurusan Siswa Kelas XI," vol. 16, no. 2, hlm. 12, 2022.
- [4] I. Mulia, "Model Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree C4.5 dan Software Weka," vol. 5, hlm. 8.
- [5] I. A. Nikmatun dan I. Waspada, "IMPLEMENTASI DATA MINING UNTUK KLASIFIKASI MASA STUDI MAHASISWA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR," vol. 10, no. 2, hlm. 12, 2019.
- [6] D. Anugrah Putra dan M. Kamayani, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes di Program Studi Teknik Informatika UHAMKA," *Prosidi Sem Nas Teknoka*, vol. 5, hlm. 34-40, Des 2020, doi: 10.22236/teknoka.v5i.331.
- [7] D. Darwis, N. Siskawati, dan Z. Abidin, "PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN REVIEW DATA TWITTER BMKG NASIONAL," *JTK*, vol. 15, no. 1, hlm. 131, Feb 2021, doi: 10.33365/jtk.v15i1.744.
- [8] T. Imandasari, E. Irawan, A. P. Windarto, dan A. Wanto, "Algoritma Naive Bayes Dalam Klasifikasi Lokasi Pembangunan Sumber Air," *SENARIS*, vol. 1, hlm. 750, Sep 2019, doi: 10.30645/senaris.v1i0.81.
- [9] D. Ardiansyah dan W. Walim, "ALGORITMA C4.5 UNTUK KLASIFIKASI CALON PESERTA LOMBA CERDAS CERMAT SISWA SMP DENGAN MENGGUNAKAN APLIKASI RAPID MINER," vol. 1, no. 2, hlm. 8, 2019.
- [10] Z. Nabila, A. R. Isnain, dan Z. Abidin, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," *Jurnal Teknologi dan Sistem Informasi*, vol. 2, no. 2, hlm. 9.
- [11] K. H. L. Malelak, I. M. D. Ardiada, dan G. Feoh, "IMPLEMENTASI KLASIFIKASI NAIVE BAYES DALAM MEMREDIKSI LAMA STUDI MAHASISWA (STUDI KASUS: UNIVERSITAS DHYANA PURA)," *SINTECH Journal*, vol. 4, no. 2, hlm. 202-209, Okt 2021, doi: 10.31598/sintechjournal.v4i2.964.
- [12] D. Maulana dan E. L. Nurjanah, "ANALISA TINGKAT KEPUASAN PELANGGAN TERHADAP PENJUALAN BEAUTY PRODUK PADA ONLINE SHOP DENGAN MENGGUNAKAN METODE NAIVE BAYES," vol. 10, hlm. 8, 2019.
- [13] D. Heryana, "Data mining untuk memprediksi kelulusan mahasiswa pendidikan matematika uin Raden Intan Lampung Menggunakan Naive Bayes." repository UIN Raden Intan Lampung, 2020. [Daring]. Tersedia pada: <http://repository.radenintan.ac.id/6430/>
- [14] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, hlm. 160-165, Agu 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [15] M. Silvi, "PENERAPAN METODE NAIVE BAYES CLASSIFIER UNTUK MEMREDIKSI LAMA MASA STUDI MAHASISWA." UIN Suska Riau Repositori, 2020. [Daring]. Tersedia pada: <http://repository.uinsuska.ac.id/27494>
- [16] M. I. Fikri, T. S. Sabrila, dan Y. Azhar, "Perbandingan Metode Naive Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *SMATIKA*, vol. 10, no. 02, hlm. 71-76, Des 2020, doi: 10.32664/smatika.v10i02.455.
- [17] B. G. Sudarsono, M. I. Leo, A. Santoso, dan F. Hendrawan, "ANALISIS DATA MINING DATA NETFLIX MENGGUNAKAN APLIKASI RAPID MINER," *JBASE*, vol. 4, no. 1, Apr 2021, doi: 10.30813/jbase.v4i1.2729.
- [18] Azka Agustine Kania dan Andri, "PREDIKSI KELULUSAN PMPA DI SEKOLAH MENENGAH ATAS," 2020.