

Klasifikasi Untuk Memprediksi Tingkat Kelulusan Mahasiswa STMIK Widuri Menggunakan Algoritma *Naïve Bayes*

Alvian David Imanuel¹, Nur Nawaningtyas Pusparini², Asrul Sani³

^{1,2,3}STMIK Widuri, Jalan Palmerah Barat No. 353, Jakarta Selatan 12210, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 28 November 2023

Revisi Akhir: 05 Maret 2024

Diterbitkan Online: 12 Maret 2024

KATA KUNCI

Kelulusan mahasiswa

Prediksi

Klasifikasi

Algoritma *Naïve Bayes*

KORESPONDENSI

E-mail: alviandavidmanuel@kampuswiduri.ac.id

A B S T R A C T

Student delays in completing their studies are experienced by most higher education institutions, for example at STMIK Widuri. STMIK Widuri must be able to predict student graduation early to prevent graduation that is not on time and maintain a good name and the accreditation assessment that has been obtained. For this reason, this research was conducted to predict the graduation of STMIK Widuri students using the classification method with the *Naïve Bayes* algorithm. *Naïve Bayes* is a classification algorithm that uses probability and statistics to predict a class. The dataset used is lecture activities of STMIK Widuri students class of 2021 from 2021-2022 odd to even 2022-2023 academic year and processed using the Rapidminer application. The dataset is processed through the stages of Knowledge Discovery in Database, including selection, pre-processing, transformation, data mining and evaluation stages. From the evaluation results using the confusion matrix on the distribution of training data 50% and data testing 50%, this study resulted in an Accuracy 93,10%, Precision 95,24%, and Recall 90%. In this way, it is hoped that STMIK Widuri can utilize attributes of the data stored in the database to be processed more optimally, for example using existing techniques in data mining.

1. PENDAHULUAN

Perguruan tinggi adalah salah satu lembaga penyedia pendidikan yang diperuntukkan untuk mahasiswa dan berkewajiban menghasilkan lulusan berkompeten. Salah satu hal terpenting dalam menilai keberhasilan program studi pada sebuah perguruan tinggi yaitu tingkat keberhasilan mahasiswa itu sendiri [1]. Suatu program studi pada perguruan tinggi harus mempunyai tindakan-tindakan yang responsif mengenai tingkat kelulusan mahasiswa sebagai cara pencegahan kemungkinan terjadinya keterlambatan kelulusan mahasiswa [2]. Kelulusan mahasiswa dari mata kuliah yang ditempuh adalah sebuah nilai yang dapat dihitung parameter keberhasilannya artinya mahasiswa yang mendapat nilai akhir baik, menggambarkan mahasiswa tersebut mengikuti alur perkuliahan dengan baik [3].

Permasalahan yang timbul selama kegiatan perkuliahan yaitu tidak semua mahasiswa berpotensi lulus tepat waktu karena mahasiswa tidak mengikuti alur perkuliahan dengan baik. Beberapa faktor yang mempengaruhi diantaranya masalah ekonomi dan juga kurangnya pemahaman mahasiswa dalam *mentransfer* ilmu yang diberikan tenaga pengajar/dosen.

Khasanah dkk [4] mengatakan bahwa tingkat kelulusan mahasiswa yang tidak tepat waktu bisa juga mempengaruhi kualitas dari perguruan tinggi misalnya berdampak pada penilaian akreditasi yang telah diraih dan juga nama baik perguruan tinggi. Ketidaksanggupan mahasiswa dalam menuntaskan studinya dihadapi juga oleh STMIK Widuri yang memiliki program studi Sistem Informasi dan Teknik Informatika. Terdapat beberapa mahasiswa STMIK Widuri yang menyelesaikan masa studinya lebih lama dari yang diharapkan yaitu 8 semester/4 tahun. Sebagai contoh terdapat sebagian mahasiswa STMIK Widuri angkatan 2018 yang seharusnya sudah lulus di tahun akademik 2021/2022 genap tetapi pada kenyataannya mahasiswa tersebut masih berstatus aktif kuliah karena ada beberapa mata kuliah yang tidak lulus/harus diulang maupun jumlah SKS yang masih dibawah rata-rata. Mengenai masalah diatas, untuk mengetahui kelulusan mahasiswa pastinya juga membutuhkan pengolahan dari data-data misalnya pada *dataset* aktivitas kuliah mahasiswa yang berisikan atribut-atribut prediktor seperti SKS, Jumlah mata kuliah, IPS/IPK. Oleh karena itu, data tersebut akan menguntungkan perguruan tinggi khususnya STMIK Widuri jika diolah lebih maksimal untuk memprediksi tingkat kelulusan mahasiswa lebih awal.

Penerapan teknik *data mining* dapat dipakai untuk menemukan pola kelulusan mahasiswa sehingga diketahui prediksi kelulusan

mahasiswa lebih awal [5]. *Data mining* ialah rangkaian-rangkaian proses untuk memperdalam nilai guna pada informasi yang sepanjang ini tidak didapati secara manual dari sebuah basis data, caranya yaitu dengan melaksanakan penambangan pola-pola pada data dengan maksud mengekstraksi dan menemukan pola-pola penting yang terdapat pada data di sebuah basis data dengan tujuan mengubah data menjadi suatu informasi agar didapatkan pengetahuan yang lebih bermanfaat untuk pengambilan suatu keputusan [6]. Salah satu metode dan algoritma pada *data mining* yang dapat memprediksi kelulusan mahasiswa adalah metode klasifikasi, metode ini menghasilkan model untuk memprediksi kelas atau label dari atribut-atribut di dalam basis data [7]. Algoritma *Naïve Bayes* merupakan algoritma pengklasifikasian statistik berjenis *supervised learning* yang memakai perhitungan probabilitas guna memprediksi keanggotaan suatu kelas [8].

Penelitian ini menggunakan algoritma *Naïve Bayes* untuk melakukan klasifikasi tingkat kelulusan mahasiswa karena dilihat dari penelitian terdahulu, algoritma *Naïve Bayes* dinilai menghasilkan akurasi yang lebih baik dibanding algoritma klasifikasi lain. Hozairi dkk [9] melakukan klasifikasi kelulusan mahasiswa dengan algoritma *K-Nearest Neighbor*, *Decision Tree* dan *Naïve Bayes*. Hasil penelitian ini, algoritma *Naïve Bayes* mendapat akurasi yang paling tinggi yaitu 89% sedangkan *KNN* (77%) dan *Decision Tree* (74%). Rovidatul dkk [10] memprediksi kelulusan mahasiswa dengan membandingkan algoritma *C4.5* dengan *Naïve Bayes*. Penelitian ini menghasilkan tingkat akurasi algoritma *Naïve Bayes* lebih besar yaitu 81,58% dibandingkan algoritma *C4.5* hanya 78,95%. Selanjutnya oleh Solikhah dkk [11] menganalisa perbandingan algoritma *Naïve Bayes* dan *C4.5* untuk memprediksi kelulusan perguruan tinggi. Hasil penelitian ini menunjukkan pengklasifikasian dengan memakai algoritma *Naïve Bayes* tingkat keakuratannya lebih besar dari pada algoritma *C4.5* yaitu 94%.

Berlandaskan uraian penjelasan diatas, penelitian ini dibuat dengan tujuan mengklasifikasikan tingkat kelulusan mahasiswa STMIK Widuri guna memprediksi tingkat kelulusan mahasiswa lebih awal menggunakan metode klasifikasi dengan algoritma *Naïve Bayes*. *Dataset* yang digunakan adalah *dataset* aktivitas kuliah mahasiswa STMIK Widuri angkatan 2021, *dataset* tersebut akan diolah menggunakan bantuan aplikasi *Rapidminer* untuk memberikan hasil prediksi klasifikasi kelulusan mahasiswa. Harapan dari penelitian ini adalah STMIK Widuri bisa mengetahui prediksi kelulusan mahasiswanya lebih awal sehingga dapat memberikan peringatan kepada mahasiswa yang diklasifikasikan lulus tidak tepat waktu ataupun mencegah terjadinya mahasiswa yang diklasifikasikan lulus tidak tepat waktu dengan mengadakan semester pendek ataupun memberikan remedial mata kuliah yang tidak lulus/harus diulang. Dengan cara tersebut juga, data-data akademik misalnya *dataset* aktivitas kuliah mahasiswa yang tersimpan dalam *database* siacad STMIK Widuri bisa diolah lebih maksimal lagi untuk mengetahui informasi penting yang tersembunyi.

2. TINJAUAN PUSTAKA

2.1 Kelulusan Mahasiswa

Seorang mahasiswa dikatakan lulus, jika sudah menuntaskan syarat minimal SKS sesuai dengan kurikulum masing-masing

perguruan tinggi dan memperoleh IPK/Indeks Prestasi Kumulatif minimal 2,00 dan mahasiswa tersebut telah menyelesaikan tugas akhir/skripsi lalu *publish* sebuah karya ilmiah bagi mahasiswa Strata 1 (S-1) [12].

2.2 Data Mining

Data mining merupakan satu diantara cabang ilmu yang mendalami untuk menemukan suatu pola atau pengetahuan dari sebuah basis data. *Data mining* ialah rangkaian-rangkaian proses untuk memperdalam nilai guna pada informasi yang sepanjang ini tidak didapati secara manual dari sebuah basis data, caranya yaitu dengan melaksanakan penambangan pola-pola pada data dengan maksud mengekstraksi dan menemukan pola-pola penting yang terdapat pada data di sebuah basis data dengan tujuan mengubah data menjadi suatu informasi agar didapatkan pengetahuan yang lebih bermanfaat untuk pengambilan suatu keputusan [5].

Pada umumnya, *data mining* mengolah data yang berasal dari suatu basis data dalam jumlah yang banyak dengan menggunakan ilmu statistik, matematika, *artificial intelligence*, hingga *machine learning* untuk mendapatkan sebuah informasi dan pengetahuan yang bermanfaat. Hasil dari pengolahan tersebut nantinya dapat digunakan sebagai pengambilan suatu keputusan hingga analisa-analisa yang diperlukan [13].

2.3 Klasifikasi

Mahanggara dan Laksito berpendapat bahwa [14], klasifikasi ialah cara untuk mengetahui model yang menerangkan atau membeda-bedakan kelas yang belum diketahui dari sebuah objek. Dalam klasifikasi ada 2 proses untuk mengklasifikasikan data adalah berikut ini :

1. *Menstraining* data, memakai *dataset training* yang mana telah diberikan label-label pada *dataset* tersebut untuk melatih algoritma dalam menemukan model yang tepat.
2. *Mentesting* data, digunakan untuk melakukan pengujian lalu mengetahui akurasi dari model yang akan diciptakan pada proses *testing*. Oleh sebab itu, diperlukan penggunaan *dataset testing* untuk melaksanakan perhitungan akurasi dari label-label yang telah diprediksikan.

Berdasarkan hal tersebut pengklasifikasian *dataset* ialah salah satu upaya untuk membuat *dataset* terstruktur. Atribut-atribut pada *dataset* yang memiliki ciri yang sama akan ditempatkan pada kelas-kelas yang sama juga.

2.4 Algoritma *Naïve Bayes*

Saputro & Sari [15] berpendapat bahwa *Naïve bayes* ialah algoritma yang dipergunakan untuk pengklasifikasian, memakai teori teorema *bayes* serta beranggapan bahwa nilai antar variabel saling bebas/independen terhadap suatu nilai *output* atau dengan maksud lain bisa dikatakan bahwa ada atau tidaknya dari suatu variabel tertentu tidak memiliki keterkaitan terhadap ada atau tidaknya variabel yang lainnya.

Algoritma *Naïve Bayes* termasuk ke dalam *supervised learning*, maksud nya adalah pembelajaran yang seperti dilakukan di sekolah yaitu memerlukan seorang “guru” untuk melatih data (dalam hal ini pada *data training* yang digunakan) sehingga bisa menemukan model prediksi klasifikasi yang sesuai dari pembelajaran *data training* yang digunakan tersebut dan juga dari pembelajaran *data training*, akan diuji untuk mengetahui akurasi

dari algoritma *Naive Bayes* terhadap model yang dihasilkan pada proses *testing* menggunakan *data testing*. *Supervised learning* ialah sebuah teknik dimana memiliki data yang dilatih (*data training*) dan pada *data training* tersebut sudah diberi label/kelas yang merupakan ciri-ciri dari *supervised learning* sehingga tujuannya yaitu mengelompokkan suatu *dataset* untuk diberikan label/target ke *dataset* yang sudah diberi label/target [16]. Dibawah ini adalah rumus umum pada algoritma *Naive Bayes* yaitu [15] :

$$P(c|d) = \frac{P(d|c) \times P(c)}{P(d)} \dots\dots\dots (1)$$

- P(c|d) = *posterior*/peluang kelas c diberikan kondisi d.
- P(d|c) = peluang kelas d diberikan kondisi c (*likelihood*).
- P(c) = *prior*/peluang awal munculnya kelas c.
- P(d) = *evidence*/peluang munculnya kelas d.

2.5 *Confusion Matrix*

Confusion matrix ialah sebuah cara/teknik yang umumnya dipergunakan untuk menghitung akurasi pada pembelajaran *data mining*. *Confusion matrix* merupakan perhitungan performa yang ditujukan untuk permasalahan pengklasifikasian pada *data mining*, dimana *output* dapat berupa 2 kelas/label bahkan lebih. *Confusion matrix* berisikan informasi-informasi terkait kelas sebenarnya/aktual dan kelas prediksi yang didapatkan dari proses klasifikasi [14].

Tabel 1. Tabel *Confusion Matrix*

Kelas aktual	Kelas Prediksi		
	True (+)	True Positif/TP	False (-)
	False (-)	False Positif/FP	True Negatif /TN

Melalui Tabel 1, bisa didapatkan suatu pengukuran yang sangat bermanfaat untuk mengetahui performa dari model yang dihasilkan, antara lain yaitu [17] :

1. *Accuracy* yaitu besarnya persentase banyaknya data yang diklasifikasikan benar dari total seluruh data.

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \dots\dots\dots (2)$$

2. *Precision* merupakan banyaknya data diprediksi benar oleh model dan ternyata sesuai dengan prediksi sebenarnya yang terdapat pada *dataset*.

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (3)$$

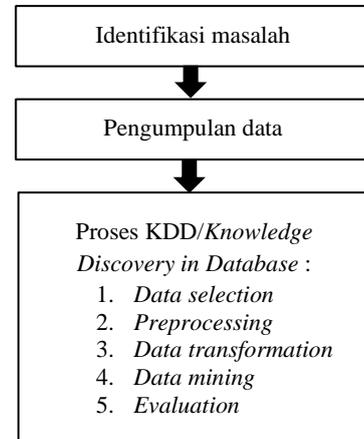
3. *Recall* memperlihatkan bahwa persentase/seberapa banyak data aktual (data sebenarnya) yang berada pada kelas positif/negatif dan diprediksikan benar oleh model.

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (4)$$

3. METODOLOGI

Teknik pengumpulan data pada penelitian ini memakai pendekatan kuantitatif yang dimana *dataset* yang digunakan pada penelitian dapat diolah menggunakan cara-cara statistik [18]. Sumber data primer berasal dari *dataset* aktivitas kuliah

mahasiswa STMIK Widuri program studi Sistem Informasi dan Teknik Informatika angkatan tahun akademik 2021 sedangkan sumber data sekunder berasal dari referensi-referensi yang berkaitan dengan penelitian ini seperti dari jurnal-jurnal terdahulu tentang klasifikasi menggunakan algoritma *Naive Bayes*. Langkah-langkah penelitian yang akan dilakukan penulis dari awal hingga akhir sehingga penelitian ini menjadi lebih baik lagi adalah sebagai berikut :

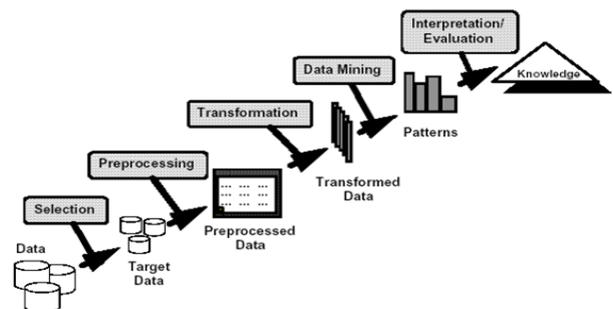


Gambar 1. Langkah-langkah Penelitian

Penulis melakukan identifikasi masalah dengan memberikan poin penting masalah yaitu masih terdapat mahasiswa STMIK Widuri yang lulus tidak tepat waktu sehingga bisa berdampak pada penilaian akreditasi yang telah diraih dan juga nama baik STMIK Widuri. Untuk itu STMIK Widuri memerlukan sebuah metode yang dapat memprediksi lebih awal tingkat kelulusan mahasiswanya.

Data yang dikumpulkan penulis yaitu *dataset* aktivitas kuliah mahasiswa STMIK Widuri angkatan 2021 yang terdiri dari prodi Sistem Informasi sebanyak 27 mahasiswa dan prodi Teknik Informatika sebanyak 34 mahasiswa. Atribut datanya adalah No. daftar, NIM, Nama, Tempat dan tanggal lahir, Jenis kelamin, No. telp, E-mail, Alamat, Status pekerjaan, Program studi, Kelas, IPS semester 1-4, SKS semester 1-4 yang lulus, Mata kuliah semester 1-4 yang lulus, Pembayaran uang kuliah semester 1-4.

Dataset aktivitas kuliah mahasiswa STMIK Widuri akan diolah terlebih dahulu sebelum masuk ke dalam tahap *data mining* menggunakan proses-proses yang ada pada KDD karena tidak semua data akan diolah dan juga untuk menghindari inkonsistensi data, redudansi data ataupun *missing value*. Adapun proses-proses nya sebagai berikut [19]:



Gambar 2. Proses *Knowledge Discovery in Database* (KDD)
Sumber : [19]

Data selection : melakukan pemilihan atribut dari *dataset* yang digunakan karena tidak semua atribut dipakai untuk proses klasifikasi. **Preprocessing** : melakukan pembersihan data jika terdapat data ganda, data yang tidak konsisten, data yang salah ketik (*typo*), dan *missing value*. **Data transformation** : transformasi dilakukan untuk mengubah *dataset* menjadi bentuk yang sesuai untuk *data mining*, dimana terdapat atribut-atribut dari *dataset* yang digunakan akan dikategorikan. Selain itu, *dataset* akan dipisah menjadi 2 yaitu *data training* dan *data testing*. **Data mining** : pada proses ini *dataset* aktivitas kuliah mahasiswa STMIK Widuri sudah siap untuk diolah dengan bantuan aplikasi *Rapidminer* menggunakan metode klasifikasi dengan algoritma *Naïve Bayes*. **Evaluation** : evaluasi dilakukan untuk mengetahui performa algoritma *Naïve Bayes* yang mempergunakan *confusion matrix* dalam mengklasifikasikan tingkat kelulusan mahasiswa STMIK Widuri.

4. HASIL DAN PEMBAHASAN

4.1. Data Selection

Penulis memilih atribut-atribut mana saja yang akan dipakai untuk proses klasifikasi karena tidak semua atribut pada *dataset* terpakai untuk proses klasifikasi. Atribut yang dipakai untuk proses klasifikasi *Naïve Bayes* antara lain Nama, Status pekerjaan, IPS semester 1-4, Jumlah SKS semester 1-4 yang lulus, Jumlah mata kuliah semester 1-4 yang lulus, Pembayaran uang kuliah semester 1-4, dan Kelulusan. Pada atribut Nama hanya dijadikan sebagai ID saja sehingga tidak akan terpakai untuk proses klasifikasi tetapi atribut Nama tersebut tetap muncul pada hasil klasifikasi.

Tabel 2. Penyeleksian Atribut

Atribut Prediktor	Atribut Target (Label)
Status pekerjaan	Kelulusan (tepat waktu dan tidak tepat waktu)
IPS semester 1-4	
Jumlah SKS semester 1-4 yang lulus	
Jumlah mata kuliah semester 1-4 yang lulus	
Pembayaran uang kuliah semester 1-4	

4.2. Preprocessing

Dataset aktivitas kuliah mahasiswa STMIK Widuri angkatan 2021 terdapat 2 *record* yang *missing value*. Penulis menggunakan operator *filter examples* pada aplikasi *Rapidminer* untuk membuang *record* yang *missing value*. Pada operator *filter examples*, penulis memilih kondisi yaitu *no_missing_attributes* sehingga atribut yang bersifat *missing values* akan dibuang.

Row No.	Nama	Status ...	IPS1	IPS2	IPS3	IPS4	SI
1	ABD H...	?	?	?	?	?	?
2	NUR R...	?	?	?	?	?	?

Gambar 3. Contoh *Missing Value*

4.3. Data Transformation

Transformasi dilakukan penulis supaya *dataset* pada penelitian ini menjadi bentuk yang lebih sesuai untuk proses *data mining*.

Tabel 3. *Data Transformation*

Atribut	Kategori
IPS semester 1-4 menjadi	0–1,99 (kurang)
IPK	2,00–2,49 (cukup)
	2,50–2,99 (baik)
	3,00–3,49 (memuaskan)
	3,50–4,00 (sangat memuaskan)
Pembayaran uang kuliah semester 1-4 menjadi status pembayaran	Tidak ada tunggakan, belum melunasi 4 semester, belum melunasi 3 semester, belum melunasi 2 semester, belum melunasi 1 semester
SKS semester 1-4 yang lulus menjadi total SKS lulus	
Mata kuliah semester 1-4 yang lulus menjadi total mata kuliah yang lulus	

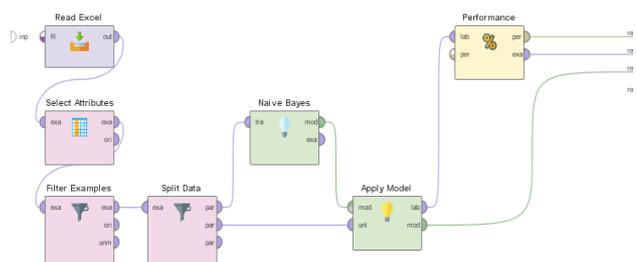
Tabel 4. *Dataset* Setelah Transformasi

Nama	Status Pekerjaan	IPK	Total SKS Lulus	Total Matkul Lulus	Status Pembayaran	Kelulusan
MHS 1	Tidak Bekerja	Memuaskan	82	31	Tidak ada tunggakan	Tepat waktu
MHS 2	Bekerja	Sangat Memuaskan	82	32	Tidak ada tunggakan	Tepat waktu
MHS 3	Bekerja	Sangat Memuaskan	84	32	Tidak ada tunggakan	Tepat waktu
MHS 4	Bekerja	Kurang	18	7	Belum melunasi 4 semester	Tidak tepat waktu
MHS 5	Bekerja	Memuaskan	84	32	Tidak ada tunggakan	Tepat waktu
MHS 6	Bekerja	Memuaskan	84	32	Tidak ada tunggakan	Tepat waktu
MHS 7	Tidak Bekerja	Kurang	2	1	Belum melunasi 4 semester	Tidak tepat waktu
MHS 8	Tidak Bekerja	Memuaskan	74	28	Tidak ada tunggakan	Tidak tepat waktu
MHS 9	Bekerja	Memuaskan	82	32	Tidak ada tunggakan	Tepat waktu
MHS 10	Tidak Bekerja	Sangat Memuaskan	84	32	Tidak ada tunggakan	Tepat waktu

Pada tahap *transformation* ini, *Dataset* aktivitas kuliah mahasiswa STMIK angkatan 2021 juga dibagi ke dalam 2 jenis yaitu *data training* (50%) dan *data testing* (50%) menggunakan operator *split data* di aplikasi *Rapidminer*. Pembagian tersebut menggunakan *stratified sampling* yaitu proses pengambilan sampel acak terstruktur, dilakukan dengan cara membagi anggota populasi menjadi beberapa sub-sub kelompok yang dinamai strata, kemudian suatu sampel diambil dari tiap-tiap strata [20].

4.4. Data Mining

Pada tahap ini dilakukan pemodelan klasifikasi *Naïve Bayes* menggunakan aplikasi *Rapidminer*. Operator yang digunakan yaitu *Read Excel* untuk menginput *dataset* aktivitas kuliah mahasiswa STMIK Widuri, *Select Attributes* untuk memilih atribut mana saja yang dipakai, *Filter Examples* untuk menghilangkan data dengan kondisi *missing value*, *Split Data* untuk membagi *data training* dan *data testing*, *Naïve Bayes* untuk menghasilkan model klasifikasi *Naïve Bayes*, *Apply Model* untuk menguji *data testing* menggunakan pembelajaran *data training*, dan *Performance* untuk mengetahui performa klasifikasi *Naïve Bayes* dengan *confusion matrix*.



Gambar 4. Pemodelan Pada Aplikasi *Rapidminer*

4.5. Evaluation

Berikut ini disajikan hasil *confusion matrix* dari pengujian *data testing* 50% terhadap pembelajaran *data training* 50%. *Confusion matrix* tersebut didapatkan dari penggunaan operator *performance* di aplikasi *Rapidminer*.

Tabel 5. Hasil *Confusion Matrix*

	True tepat waktu	True tidak tepat waktu
Pred. tepat waktu	19	2
Pred. tidak tepat waktu	0	8
Accuracy	93,10%	
Precision	95,24%	
Recall	90%	

Pada tabel *confusion matrix* diatas, terdapat 19 mahasiswa yang diprediksi berada pada kelas tepat waktu dan benar berada pada kelas tepat waktu dan ada 8 mahasiswa yang diprediksi lulus tidak tepat waktu dan benar lulus tidak tepat waktu. Selain itu ada perbedaan prediksi dari data aktual (data sebenarnya) dengan prediksi dari model yaitu sebanyak 2 mahasiswa yang sebenarnya berada pada kelas aktual tidak tepat waktu tetapi ternyata model memprediksi nya berada pada kelas tepat waktu.

Accuracy yang dihasilkan dari pengujian *data testing* 50% terhadap pembelajaran *data training* 50% adalah 93,10% artinya kesamaan prediksi yang berasal pada *dataset* STMIK Widuri angkatan 2021 dengan prediksi yang dihasilkan oleh model sebesar 93,10% dengan tingkat *error* sebesar 6,9% karena ada perbedaan prediksi. Adapun *precision* yang diperoleh yaitu sebesar 95,24% artinya banyaknya data yang diprediksikan benar oleh model dan ternyata sesuai dengan data aktual (data sebenarnya) yaitu sebesar 95,24% sedangkan *recall* yang didapatkan adalah sebesar 90% yang berarti dari keseluruhan prediksi kelulusan mahasiswa STMIK Widuri angkatan 2021 yang berasal dari data aktual (data sebenarnya) didapatkan 90% diprediksi benar oleh sistem (*Rapidminer*).

4.6. Analisa Hasil Klasifikasi

Hasil klasifikasi dengan algoritma *Naïve Bayes* yang bersumber dari *dataset* aktivitas kuliah mahasiswa STMIK Widuri angkatan 2021 dari tahun akademik 2021-2022 ganjil sampai 2022-2023 genap terhadap pembagian *data training* dan *data testing* sebesar 50:50, didapatkan 8 mahasiswa yang ternyata benar tidak lulus tepat waktu karena mendapatkan IPK yang kurang (0-1,99). Total SKS yang lulus paling banyak 20 SKS. Padahal mahasiswa angkatan 2021 di semester 2022-2023 genap seharusnya sudah menempuh SKS (lulus) sebanyak 82-84 SKS. Seperti hal nya dengan total SKS lulus, pada total mata kuliah yang lulus paling banyak 8 mata kuliah, seharusnya mahasiswa angkatan tersebut di semester 2022-2023 genap sudah lulus mata kuliah sebanyak 31-32 mata kuliah. Status pembayaran dari mahasiswa-mahasiswa yang diprediksi tidak tepat waktu, rata-rata belum melunasi pembayaran uang kuliah nya selama 3-4 semester dan mahasiswa-mahasiswa tersebut ada yang bekerja dan tidak bekerja, padahal mahasiswa yang tidak bekerja bisa lebih banyak waktu untuk kuliahnya tetapi mahasiswa tersebut diprediksikan tidak tepat waktu karena mendapatkan IPK, SKS lulus, mata kuliah lulus dibawah rata-rata.

Selain itu, ada 19 mahasiswa yang diprediksi benar tepat waktu karena mendapatkan IPK yang memuaskan dan juga sangat memuaskan (3,00-4,00). Total SKS yang lulus diatas 80 SKS dan total mata kuliah lulus diatas 30 mata kuliah. 19 mahasiswa tersebut tidak ada tunggakan pembayaran uang kuliah/sudah melunasi ke-4 (semester 1-4) semesternya.

Kemudian terdapat 2 mahasiswa yang diprediksi dari data aktual tidak tepat waktu tetapi model memprediksinya tepat waktu, hal ini bisa saja terjadi karena pada kedua mahasiswa tersebut mendapatkan IPK yang memuaskan (3,00-3,49). Total SKS lulus diatas 70 SKS dan total mata kuliah lulus diatas 25 mata kuliah. Dari kedua mahasiswa yang diprediksi akan lulus tepat waktu terdapat salah satu mahasiswa yang bekerja, untuk itu mahasiswa tersebut harus bisa mengatur waktunya supaya bisa lulus tepat waktu.

Tabel 6. Probabilitas Setiap Atribut Prediktor

Atribut	Parameter	Tepat waktu	Tidak tepat waktu
Status Pekerjaan	Tidak Bekerja	0,421	0,455
Status Pekerjaan	Bekerja	0,579	0,545
IPK	Memuaskan	0,421	0
IPK	Sangat Memuaskan	0,579	0
IPK	Kurang	0	0,909
IPK	Cukup	0	0,091
Total SKS Lulus	mean	82,105	19,636
Total SKS Lulus	standard deviation	2,767	20,091
Total Matkul Lulus	mean	31,579	7,909
Total Matkul Lulus	standard deviation	0,902	7,867
Status Pembayaran	Tidak ada tunggakan	1	0
Status Pembayaran	Belum melunasi 4 semester	0	0,364
Status Pembayaran	Belum melunasi 3 semester	0	0,364
Status Pembayaran	Belum melunasi 2 semester	0	0,182
Status Pembayaran	Belum melunasi 1 semester	0	0,091

Tabel 6 merupakan tabel perhitungan probabilitas setiap atribut prediktor dari aplikasi *Rapidminer* terhadap *data training* 50%. Berdasarkan Tabel 6, mahasiswa diprediksi tepat waktu, jika mahasiswa tersebut memperoleh IPK memuaskan ataupun sangat memuaskan, rata-rata total SKS lulus sebanyak 82 SKS, rata-rata total mata kuliah lulus kurang lebih 32 mata kuliah, serta tidak ada tunggakan pembayaran uang kuliah, dan juga mahasiswa yang diprediksi lulus tepat waktu ada yang bekerja dan tidak bekerja. Kemudian, mahasiswa yang diprediksi tidak tepat waktu, jika mahasiswa tersebut memperoleh IPK cukup bahkan kurang, rata-rata total SKS lulus sebanyak 20 SKS, rata-rata total mata kuliah lulus tidak lebih dari 8 mata kuliah, serta mahasiswa tersebut masih ada tunggakan pembayaran uang kuliah 1-2 semester bahkan yang paling lambat 3-4 semester, dan mahasiswa yang diprediksi tidak tepat waktu ada yang bekerja dan juga tidak bekerja.

Dengan demikian, dari hasil-hasil prediksi terhadap klasifikasi tingkat kelulusan STMIK Widuri angkatan 2021 pada semester 2021-2022 ganjil sampai 2022-2023 genap, STMIK Widuri bisa mengetahui lebih awal tingkat kelulusan mahasiswanya, sebelum mahasiswa pada angkatan tersebut akan menempuh semester selanjutnya. Oleh karena itu, STMIK Widuri bisa melakukan tindakan preventif/pencegahan terhadap mahasiswa-mahasiswa yang diprediksi lulus tidak tepat waktu, sehingga mahasiswa-

mahasiswa tersebut bisa mengejar ketertinggalan mata kuliahnya dengan mengikuti remedial ataupun tutorial sehingga pada saat mahasiswa tersebut sudah berada di semester 6 diharapkan mahasiswa sudah mencapai 100 SKS lulus untuk bisa mengambil mata kuliah Kerja Praktek dan di semester akhir, mahasiswa yang akan mengambil mata kuliah Skripsi, mahasiswa tersebut diharapkan juga sudah tidak ada lagi mata kuliah yang harus diulang sehingga kelulusan mahasiswa tersebut bisa tepat pada waktunya.

5. KESIMPULAN DAN SARAN

Metode klasifikasi menggunakan algoritma *Naive Bayes* berhasil diterapkan untuk memprediksi kelulusan mahasiswa STMIK Widuri angkatan 2021 dari penggunaan atribut-atribut prediktor seperti Status pekerjaan, IPK, Total SKS lulus, Total mata kuliah lulus, dan juga Status pembayaran, yang menghasilkan akurasi sebesar 93,10% pada pembagian *data training* dan *data testing* 50:50. Dari hasil penelitian ini, STMIK Widuri bisa mengetahui tingkat kelulusan mahasiswanya lebih awal sehingga dapat meminimalisir terjadinya mahasiswa yang lulus tidak tepat waktu dan juga kedepannya, penelitian ini bisa dijadikan acuan/referensi bagi STMIK Widuri untuk mengolah lebih maksimal lagi data-data yang tersimpan sehingga bisa didapatkan pengetahuan-pengetahuan penting lainnya menggunakan teknik-teknik *data mining* seperti yang dilakukan pada penelitian ini. Dengan begitu nama baik STMIK Widuri dapat terjaga dan STMIK Widuri dapat mempertahankan penilaian akreditasinya.

Untuk penelitian berikutnya, dapat dilakukan perbandingan algoritma klasifikasi lain seperti *C4.5*, *K-Nearest Neighbor*, dan sejenisnya sehingga dapat mengetahui algoritma terbaik untuk memprediksi kelulusan mahasiswa. Selain itu, bisa dikembangkan aplikasi *data mining* untuk memprediksi kelulusan mahasiswa menggunakan metode dan algoritma yang sesuai untuk memudahkan proses pengolahan data dan juga penggunanya.

DAFTAR PUSTAKA

[1] Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, "Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest," *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, Dec. 2022, doi: 10.33372/stn.v8i2.885.

[2] P. Sainanda *et al.*, "Penerapan Data Mining Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes Classifier (Studi Kasus STMIK Primakara)." *Progresif J. Ilm. Komput.*, vol. 16, no. 1, pp. 47–56, 2020, doi: 10.35889/progresif.v16i1.427.

[3] H. Hartatik, "Optimasi Model Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Naive Bayes," *Indones. J. Appl. Informatics*, vol. 5, no. 1, p. 32, Apr. 2021, doi: 10.20961/ijai.v5i1.44379.

[4] N. Khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, "Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes," *Technol. J. Ilm.*, vol. 13, no. 3, p. 207, 2022, doi: 10.31602/tji.v13i3.7312.

[5] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naive Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13,

no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.

[6] A. Z. M. Sigid Widodo, A. Pandu Kusuma, and W. Dwi Puspitasari, "Analisis Algoritma Naive Bayes Classifier (Nbc) Pada Klasifikasi Tingkat Minat Barang Di Toko Violet Cell," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 87–94, 2023, doi: 10.36040/jati.v7i1.5692.

[7] Heliyanti Susana, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet," *J. Ris. Sist. Inf. dan Teknol. Inf.*, vol. 4, no. 1, pp. 1–8, 2022, doi: 10.52005/jursistekni.v4i1.96.

[8] N. Purwati and A. Dwi Januanti, "APLIKASI DATA MINING DENGAN ALGORITMA NAIVE BAYES UNTUK MEMPREDIKSI TINGKAT KELULUSAN MAHASISWA," *J. Pepadun*, vol. 2, no. 1, pp. 123–137, Apr. 2021, doi: 10.23960/pepadun.v2i1.38.

[9] H. Hozairi, A. Anwari, and S. Alim, "Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes," *Netw. Eng. Res. Oper.*, vol. 6, no. 2, p. 133, 2021, doi: 10.21107/nero.v6i2.237.

[10] Rovidatul, Y. Yunus, and G. W. Nurcahyo, "Perbandingan algoritma c4.5 dan naive bayes dalam prediksi kelulusan mahasiswa," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 4, no. 1, pp. 193–199, Apr. 2023, doi: 10.37859/coscitech.v4i1.4755.

[11] F. Solikhah, M. Febianah, A. L. Kamil, W. A. Arifin, and Shelly Janu Setyaning Tyas, "Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," *TEMATIK*, vol. 8, no. 1, pp. 96–103, Jun. 2021, doi: 10.38204/tematik.v8i1.576.

[12] S. Wahyuni and Adinda, "PENERAPAN ALGORITMA NAIVE BAYES PADA PREDIKSI KELULUSAN MAHASISWA (STUDI KASUS: INSTITUT MEDIKA Drg. SUHERMAN)," *J. Inform. SIMANTIK*, vol. 6, no. 2, pp. 29–34, 2021.

[13] Tumini and L. Damayanti, "Aplikasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naive Bayes Studi Kasus Smp Negeri 11 Kotabumi Utara," *Inform. SIMANTIK*, vol. 3, no. 2, pp. 23–30, 2018.

[14] A. Mahanggara and A. D. Laksito, "Prediksi Pengunduran Diri Mahasiswa Universitas Amikom Yogyakarta Menggunakan Metode Naive Bayes," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 10, no. 1, pp. 273–280, 2019, doi: 10.24176/simet.v10i1.2967.

[15] I. W. Saputro and B. W. Sari, "Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Mahasiswa," *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 1, 2020, doi: 10.24076/citec.2019v6i1.178.

[16] G. A. Pradnyana and K. Agustini, "Konsep Dasar Data Mining," vol. 1, 2018, pp. 1–47.

[17] E. Etriyanti, D. Syamsuar, and N. Kunang, "Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa," *Telematika*, vol. 13, no. 1, pp. 56–67, 2020, doi: 10.35671/telematika.v13i1.881.

[18] A. Khaerunnisa, "Analisis Tingkat Kelulusan Mahasiswa di Unisba dengan menggunakan Algoritma K-Means Clustering," *J. Ris. Mat.*, pp. 67–76, 2022, doi: 10.29313/jrm.v2i1.1018.

[19] S. Alam, M. G. Resmi, and N. Masripah, "Classification of Covid-19 vaccine data screening with Naive Bayes algorithm using Knowledge Discovery in database method," *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 4, no. 2, pp. 177–185, 2022, doi: 10.47709/cnahpc.v4i2.1584.

[20] F. Sulianta, *Basic Data Mining from A to Z*. 2023.

BIODATA PENULIS



Alvia David Imanuel

Mahasiswa Program Studi Teknik Informatika
STMIK Widuri, Jakarta Selatan. Email:
alviandavidmanuel@kampuswiduri.ac.id



Nur Nawaningtyas Pusparini

Dosen Teknik Informatika STMIK Widuri,
Jakarta Selatan. Email:
tyaspusparini@kampuswiduri.ac.id



Asrul Sani

Dosen Teknik Informatika STMIK Widuri,
Jakarta Selatan. Email:
asrulsani@kampuswiduri.ac.id