

PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAIVE BAYES

Nurul Khasanah^{1)*}, Agus Salim²⁾, Nurul Afni³⁾, Rachman Komarudin⁴⁾, Yana Iqbal Maulana⁵⁾

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, Fakultas Teknologi Informasi, Universitas Bina Sarana Informatika

Email: nurul.nuk@nusamandiri.ac.id, agus.salim@bsi.ac.id, nurul.nrf@bsi.ac.id, rachman.rck@bsi.ac.id,
yana.yim@bsi.ac.id

Abstrak

Perguruan tinggi memiliki kewajiban untuk menghasilkan lulusan yang kompeten. Hal tersebut dapat dinilai dari tingkat kelulusan mahasiswanya. Selain itu, lulus tepat waktu merupakan impian para mahasiswa. Mahasiswa tidak perlu membayar biaya kuliah lagi dan bisa bekerja lebih cepat. Tetapi kondisi di lapangan, 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. Dari permasalahan yang ada, 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.

Keywords: Kelulusan, Naive Bayes, KNN

I. PENDAHULUAN

Perguruan tinggi memiliki kewajiban untuk menghasilkan lulusan yang kompeten. Hal tersebut dapat dinilai dari tingkat kelulusan mahasiswanya. Selain itu, lulus tepat waktu juga menjadi impian para mahasiswa. Mahasiswa tidak perlu membayar biaya kuliah lagi dan bisa bekerja lebih cepat. Tetapi kondisi di lapangan, mahasiswa belum tentu dapat menuntaskan masa studi tepat pada waktunya. 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. Selain itu, tingkat kelulusan juga berpengaruh pada akreditasi suatu perguruan tinggi sehingga perguruan tinggi berusaha untuk membantu mahasiswa agar lulus tepat waktu.

Tingkat kelulusan mahasiswa bisa diprediksi menggunakan suatu sistem. Namun,

beberapa perguruan tinggi belum memiliki sistem untuk memprediksi keterlambatan kelulusan mahasiswa, sehingga perguruan tinggi tidak dapat melakukan pencegahan akan hal tersebut.

Beberapa peneliti sudah membuat sebuah sistem untuk memprediksi tingkat kelulusan mahasiswa. Dalam penelitian (Pratama, Wihandika and Ratnawati, 2018) yang berjudul “Implementasi Algoritme Support Vector Machine (SVM) untuk Prediksi Ketepatan Waktu Kelulusan Mahasiswa” menghasilkan akurasi sebesar 80,55% dengan menggunakan data training sebesar 170 data. Sedangkan untuk penelitian (Putri and Waspada, 2018) yang berjudul “Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika” menghasilkan akurasi sebesar 62.44%. variabel yang digunakan antara lain: jenis kelamin, asal daerah, IPK, dan TOEFL. Hasil akurasi yang diperoleh dari penelitian sebelumnya masih kecil dan belum termasuk dalam kategori baik.

Dari permasalahan yang ada, perlu adanya sistem untuk memprediksi tingkat kelulusan mahasiswa berdasarkan variabel-variabel yang ada. Selain itu, diperlukan algoritma yang cocok sehingga bisa menghasilkan nilai akurasi yang baik. Dengan sistem yang dibuat diharapkan perguruan tinggi bisa membuat kebijakan sehingga mahasiswa dapat lulus tepat pada waktunya.

II. STUDI PUSTAKA

1. Dataset

Data *training* merupakan sekumpulan data yang memiliki atribut label/kelas yang digunakan oleh mesin untuk mengenal karaktersitik kumpulan data sehingga menghasilkan sebuah pola/model data. Sementara data testing adalah sekumpulan data yang juga memiliki label/kelas yang digunakan untuk menguji ketepatan pola/model dalam mengklasifikasikan data testing. Pada saat melakukan proses testing model, atribut label dari data testing disembunyikan selama proses klasifikasi berlangsung dan akan digunakan untuk membandingkan hasil klasifikasi sebagai tolak ukur seberapa besar ketepatan/akurasi model tersebut melakukan klasifikasi (Musu, Ibrahim and Heriadi, 2021).

Dalam praktek penemuan pola dari kumpulan data, seluruh data yang akan dianalisis dikumpulkan menjadi bentuk dataset yang terdiri dari atribut atribut yang mengandung informasi penting terhadap pola yang hendak di temukan (Fernández *et al.*, 2013). Dataset tersebut kemudian dibagi menjadi dataset untuk training dan dataset untuk testing. Hal-hal yang mempengaruhi tingkat akurasi pola klasifikasi yang dihasilkan selain karakteristik dari data yang dimiliki, komposisi atau presentasi pembagian jumlah data untuk training dan testing juga merupakan suatu hal penting yang harus menjadi pertimbangan, secara kusus ketika menggunakan algoritma-algoritma supervised learning. Ketidaktepatan dalam penentuan komposisi jumlah dataset training dan testing akan mempengaruhi akurasi terhadap pola klasifikasi yang di temukan.

Populasi adalah jumlah keseluruhan dari sampel yang digunakan dalam penelitian. Sedangkan sample adalah bagian dari populasi yang didapatkan dengan menggunakan metode tertentu yang kemudian dianggap menjadi wakil dari populasi tersebut yang menjadi fokus dalam metode penelitian sosial atau penelitian statistika.

2. Data mining

Data mining (Sikumbang, 2018) merupakan proses iterative dan interaktif untuk menemukan pola atau model baru yang sempurna, bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (massive database). Data mining berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambil keputusan diwaktu yang akan datang, pola-pola ini dikenali perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lain.

Tahap-tahap data mining ada 6 yaitu :

a. Pembersihan data (data cleaning)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak relevan. Pada umumnya data yang diperoleh, baik dari database memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Data- data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performansi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.

b. Integrasi data (data integration)

Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

c. Seleksi Data (Data Selection)

Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti factor kecenderungan orang membeli dalam kasus market basket analysis, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

d. Transformasi data (Data Transformation)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.

- e. Proses mining
Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
- f. Evaluasi pola (pattern evaluation)
Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasilnya berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai.

3. Klasifikasi

Klasifikasi merupakan suatu metode pengelompokan data yang akan mempelajari data latih dengan menggunakan algoritme pengklasifikasian (Pratama, Wihandika and Ratnawati, 2018). Beberapa algoritma yang digunakan untuk klasifikasi, antara lain k-Nearest Neighbor, DecisionTree, Algorithms, Naive Bayes dan Support Vector Machine. Proses klasifikasi didasarkan pada empat komponent (Arie Yandi Saputra, 2018) :

- a. Kelas Variabel dependen berupa kategori yang mempresentasikan “label” yang terdapat pada objek.
- b. Predictor Variabel independen yang direpresentasikan oleh karakteristik data.
- c. Training dataset Satu set data yang mempunyai nilai dari kedua komponen yang digunakan untuk menentukan kelas yang cocok berdasarkan predictor.
- d. Testing dataset Berupa data baru yang diklasifikasikan oleh model data yang telah di buat dan akurasi klasifikasi di evaluasi.

Tools yang digunakan dalam proses klasifikasi adalah RapidMiner 9.10.001. Rapidminer merupakan platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan yang bernama sama dengan pihak yang menyediakan lingkungan terintegrasi untuk persiapan data, penambahan teks, dan analisis prediktif. Rapidminer dilengkapi dengan satu prosesor

logika dan 10.000 baris data. Rapidminer yang tersedia berada di bawah lisensi AGPL (Elfaladonna and Rahmadani, 2019).

4. K-NN

Konsep dasar dari K-NN adalah mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga terdekatnya. Nilai dari jarak antara data uji dengan data latih diurutkan dari nilai terendah. Proses pengurutan tersebut dilakukan untuk memilih jarak minimum sebanyak K buah (Nasution, Khotimah and Chamidah, 2019).

Perhitungan K-Nearest Neighbor akan diterjemahkan kepada sistem dengan dibuat modelbase agar mampu mempresentasikan permasalahan ke dalam format kuantitatif sebagai dasar pengambilan keputusan (Novita, Harsani and Qur’ania, 2018). Perancangan modelbase bertujuan agar data yang diinputkan oleh user dapat diolah dan menghasilkan hasil keputusan yang baik. Alur perhitungan K-Nearest Neighbor ditunjukkan pada Gambar berikut



Gambar 1. Alur Perhitungan K-Nearest Neighbor
Sumber gambar : (Novita, Harsani and Qur’ania, 2018)

Berikut menunjukkan rumus mencari nilai jarak euclidian.

$$d = |x_i - y_i| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan :

x_i = Data Uji

y_i = Data Latih

i = Variabel Data

d = Jarak

Dalam metode K-NN, pendekatan sederhana dalam menentukan nilai k bisa dilakukan dengan cara berikut $k = \sqrt{n}$ dimana n merupakan banyaknya data yang dihitung. Nilai k yang tinggi akan mengurangi noise pada klasifikasi. Perancangan database dilakukan untuk menentukan isi dan pengaturan data yang

dibutuhkan untuk mendukung berbagai rancangan sistem. Tahap akhir perancangan adalah merancang interface dengan tujuan memberikan gambaran bagaimana user dan sistem berinteraksi atau dapat berkomunikasi dengan baik.

5. Evaluasi

Proses evaluasi hasil penelitian menggunakan beberapa hyperparameter, antara lain:

a. Confusion Matrix

Confusion matrix adalah tabel yang menyatakan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah (Normawati and Prayogi, 2021). Contoh tabel confusion matrix adalah sebagai berikut:

Tabel 1. Tabel confusion matrix			
		Kelas Prediksi	
		1	0
Kelas sebenarnya	1	TP	FN
	0	FP	TN

Sumber tabel : (Normawati and Prayogi, 2021)

Keterangan (Normawati and Prayogi, 2021)

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.

b. Akurasi

Akurasi yaitu menunjukkan tingkat akurasi model dalam mengklasifikasikan data dengan benar (M. Raihan Rafiiful Allaam, 2021).

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

Catatan : lihat keterangan di tabel 1.

c. Precision

Precision yaitu menunjukkan akurasi antara data yang diminta dengan hasil prediksi model (M. Raihan Rafiiful Allaam, 2021).

$$Presicion = \frac{TP}{TP + FP} \quad (3)$$

Catatan : lihat keterangan di tabel 1.

d. Recall

Recall menunjukkan keberhasilan model dalam menemukan kembali sebuah informasi (M. Raihan Rafiiful Allaam, 2021).

$$Recall = \frac{TP}{TP + FN}$$

Catatan : lihat keterangan di tabel 1.

e. Kurva ROC

Hasil perhitungan divisualisasikan dengan kurva ROC (Receiver Operating Characteristic) atau AUC (Area Under Curve). ROC memiliki tingkat nilai diagnosa yaitu (Wahyono and Nugroho, 2018):

- a) Akurasi bernilai 0.90 – 1.00 = excellent classification
- b) Akurasi bernilai 0.80 – 0.90 = good classification
- c) Akurasi bernilai 0.70 – 0.80 = fair classification
- d) Akurasi bernilai 0.60 – 0.70 = poor classification
- e) Akurasi bernilai 0.50 – 0.60 = failure

III. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini yaitu:

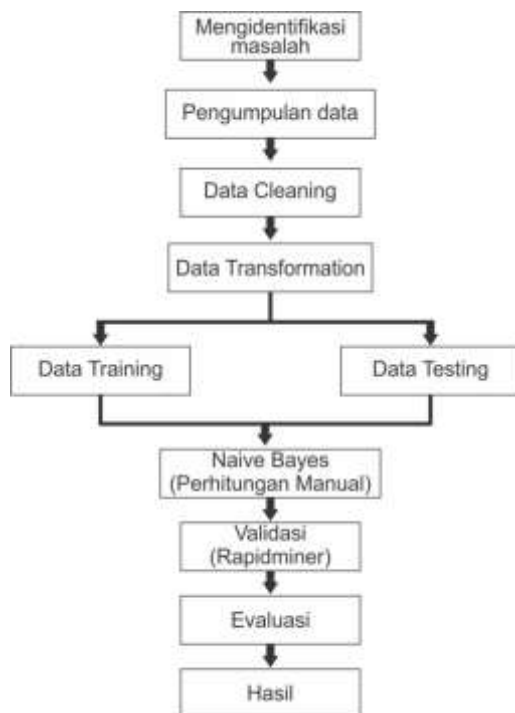
1. Teknik Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini, antara lain:

- a. Observasi, yaitu pengamatan langsung terhadap objek penelitian untuk memperoleh data penelitian.
- b. Studi Literatur, yaitu kegiatan untuk memperoleh kajian pustaka, pengetahuan, teori dan melibatkan penelitian terdahulu untuk menunjang hasil penelitian.

2. Tahapan penelitian

Dalam penelitian ini, tahapan penelitian yang digunakan yaitu (Marlina and Bakri, 2021)



Gambar 2. Tahapan Penelitian
Sumber gambar : Hasil penelitian

a. Identifikasi masalah

Dalam proses identifikasi masalah dilakukan riset langsung di lapangan. Proses yang dilalui, antara lain: identifikasi permasalahan yang ada, variabel-variabel yang berpengaruh terhadap masalah yang terjadi, solusi yang akan dilakukan dan rencana sistem yang akan dibuat untuk mengatasi masalah tersebut.

b. Pengumpulan data

Jenis data yang digunakan dalam penelitian ini adalah data kualitatif dan kuantitatif. Data yang digunakan dalam penelitian ini adalah data mahasiswa dengan jumlah 379 mahasiswa untuk diklasifikasi tingkat kelulusan tiap mahasiswa. Atribut yang digunakan dalam penelitian ini, antara lain: nama, jenis kelamin, status mahasiswa, umur, status perkawinan, IPS, IPK, dan status kelulusan.

c. Pembersihan data

Untuk membuang data yang tidak konsisten dan noise. Dalam proses ini, beberapa data yang noise atau kosong dan atribut yang tidak berpengaruh dalam perhitungan akan dihapus. Untuk atribut yang dihapus, antara lain: umur dan jenis kelamin. Sehingga atribut yang masih dipakai, antara

lain: nama, status mahasiswa, status perkawinan, IPS, IPK, dan status kelulusan.

d. Transformasi data

Untuk mengubah data menjadi bentuk yang sesuai untuk di mining. Dalam data asli, ada beberapa atribut yang perlu dirubah untuk bisa diolah dalam perhitungan data mining, antara lain: status mahasiswa dan status perkawinan. Perubahan yang dilakukan yaitu status mahasiswa: bekerja (0) dan mahasiswa (1), sedangkan untuk status perkawinan: belum menikah (1) dan menikah (0). Dalam proses transformasi data, data dibagi menjadi 2 yaitu data training dan data testing.

e. Perhitungan manual

Pada tahap ini, dilakukan perhitungan manual dengan menggunakan rumus algoritma Naive bayes. Tools yang digunakan dalam proses ini adalah Microsoft Excel.

f. Evaluasi

Aplikasi yang digunakan dalam perhitungan data mining untuk penelitian ini adalah Rapidminer 9.10. Proses evaluasi dilakukan dengan membandingkan perhitungan dengan aplikasi dan perhitungan manual. Pada tahap ini akan dilakukan perhitungan akurasi, precision dan recall menggunakan confusion matrix.

IV. HASIL DAN PEMBAHASAN

1. Pembersihan Data

Berikut kami tampilkan data asli yang digunakan dalam penelitian ini

Tabel 2. Tabel sample data penelitian

Tabel 2.1: Tabel sample data penelitian									
NAMA		JENIS KELAMIN	STATUS MAHASISWA			STATUS KELULUSAN	STATUS PERKAWINAN		
ANIK WIDAYANTI		PEREMPUAN	BEKERJA			TERLAMBAT	BELUM MENIKAH		
UMUR	IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK
28	2.76	2.8	3.2	3.17	2.98	3	3.03	0	3.07

Sumber tabel : data penelitian

Dalam proses ini dilakukan pembersihan data dan atribut yang tidak berpengaruh dalam perhitungan sehingga data yang dihasilkan sebagai berikut:

Tabel 3. Tabel sample data penelitian setelah dilakukan pembersihan data dan atribut

NAMA	STATUS MAHASISWA	STATUS KELULUSAN
ANIK WIDAYANTI	BEKERJA	TERLAMBAT

IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK
2.76	2.8	3.2	3.17	2.98	3	3.03	0	3.07

Sumber tabel : data penelitian

2. Transformasi Data

Proses transformasi data dilakukan agar data dapat diproses ke dalam aplikasi data mining. Hasil dari proses transformasi data, sebagai berikut:

Tabel 4. Hasil proses transformasi data

NAMA	STATUS MAHASISWA	STATUS KELULUSAN
ANIK WIDAYANTI	0	TERLAMBAT

IPS 1	IPS 2	IPS 3	IPS 4	IPS 5	IPS 6	IPS 7	IPS 8	IPK
2.76	2.8	3.2	3.17	2.98	3	3.03	0	3.07

Sumber tabel : data penelitian

Analisa menggunakan algoritma Naïve Bayes dimaksudkan untuk memperoleh hasil terbaik dalam mengklasifikasikan penduduk Desa Plumbon dalam pemberian bantuan Rutilahu (Rumah Tidak Layak Huni). Data yang digunakan dalam perhitungan ini sebanyak 379 data.

Data training (80%) = $80\% \times 379 = 303$ data

Data testing (20%) = $20\% \times 379 = 76$ data

Data training akan dihitung manual dengan sampel perhitungan yang diambil secara acak sejumlah 379 mahasiswa.

1. Data masukkan
Untuk data masukkan yang digunakan, penulis menggunakan 1 sampel data pada Tabel 1.1.
2. Menghitung probabilitas prior untuk label atau *class*

$$P(C_1) = \text{jumlah tepat/jumlah semua data}$$

$$= 249/478 \times 100\% = 66\%$$

$$P(C_2) = \text{jumlah terlambat/jumlah semua data}$$

$$= 130/478 \times 100\% = 34\%$$

3. Menghitung probabilitas bersyarat untuk setiap kelas $P(X_2|C)$, status mahasiswa (2) = 0 (bekerja); kelas = tepat, terlambat
 - a. $P(X_2=0, C_1=\text{tepat}) = (\text{status mahasiswa "0"}/\text{jumlah kelas tepat}) = 13\%$
 - b. $P(X_2=0, C_2=\text{terlambat}) = (\text{status mahasiswa "0"}/\text{jumlah kelas terlambat}) = 77\%$

4. Menghitung probabilitas bersyarat untuk setiap kelas $P(X_3|C)$, status nikah (3) = 1 (belum menikah); kelas = tepat, terlambat
 - a. $P(X_3=2, C_1=\text{tepat}) = (\text{status nikah "1"}/\text{jumlah kelas tepat}) = 98\%$
 - b. $P(X_3=2, C_1=\text{terlambat}) = (\text{status nikah "1"}/\text{jumlah kelas terlambat}) = 97\%$
5. Menghitung probabilitas bersyarat untuk setiap kelas $P(X_4|C)$, IPK (4) = 3 (2,5-3,4); kelas = tepat, terlambat
 - a. $P(X_4=3, C_1=\text{tepat}) = (\text{penghasilan "2"}/\text{jumlah kelas tepat}) = 78\%$
 - b. $P(X_4=3, C_1=\text{terlambat}) = (\text{penghasilan "2"}/\text{jumlah kelas terlambat}) = 42\%$
6. Mengalikan probabilitas untuk setiap kelas
 - a. $P(X|\text{Kategori}=\text{"Tepat"})$
 $P(\text{Kategori}=\text{"Tepat"})$
 $= P(\text{status mahasiswa} = 0 | X_1=\text{Tepat}) * (\text{status nikah} = 1 | X_1=\text{Tepat}) * (\text{IPK} = 3 | X_1=\text{Tepat}) = 7\%$
 - b. $P(X|\text{Kategori}=\text{"Terlambat"})$
 $P(\text{Kategori}=\text{"Terlambat"})$
 $= P(\text{status mahasiswa} = 0 | X_1=\text{Terlambat}) * (\text{status nikah} = 1 | X_1=\text{Terlambat}) * (\text{IPK} = 3 | X_1=\text{Terlambat}) = 3\%$
7. Membandingkan hasil probabilitas Tepat dan terlambat
 Berdasarkan perhitungan akhir di atas, terlihat bahwa nilai probabilitas pada class P (P|Tepat) lebih tinggi dari class P (P|Terlambat), sehingga dapat disimpulkan bahwa penduduk atas nama ANIK WIDAYANTI diprediksi bisa lulus dengan tepat waktu

3. Evaluasi Hasil

Berikut adalah perhitungan evaluasi hasil dengan RapidMiner 9.10.001

a. Confusion Matrix

Berikut adalah confusion matrix yang dihasilkan



Gambar 3. Confusion Matrix Hasil Penelitian
Sumber gambar : Hasil penelitian

Jumlah True Positive (TP) adalah 44 record diklasifikasikan sebagai mahasiswa TEPAT lulus dan False Negative (FN) sebanyak 6 record diklasifikasikan sebagai mahasiswa TEPAT tetapi TERLAMBAT lulus. Berikutnya 23 record untuk True Negative (TN) diklasifikasikan sebagai mahasiswa TERLAMBAT, dan 3 record False Positive (FP) diklasifikasikan sebagai mahasiswa TERLAMBAT tetapi TEPAT lulus.

b. Accuracy

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

$$\text{Accuracy} = \frac{44+23}{44+23+3+6}$$

$$\text{Accuracy} = 88,16\%$$

c. Presicion

$$\text{Presicion} = \frac{TP}{TP + FP}$$

$$\text{Presicion} = \frac{44}{44 + 3}$$

$$\text{Presicion} = 93,62\%$$

d. Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{Recall} = \frac{44}{44 + 6}$$

$$\text{Recall} = 88\%$$

V. SIMPULAN

Penelitian prediksi tingkat kelulusan mahasiswa menggunakan 379 data, dengan rincian data training 303 data dan data testing 76 data. Atribut yang digunakan, antara lain: nama, status mahasiswa, status perkawinan, IPS, IPK, dan status kelulusan. Tahapan yang dilakukan dalam penelitian, antara lain: 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%. Hasil klasifikasi termasuk dalam kategori good classification.

DAFTAR PUSTAKA

- Arie Yandi Saputra, Y. P. (2018) ‘Penerapan Teknik Klasifikasi Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma K-Nearest Neighbour’, 17(4), pp. 395–403.
- Elfaladonna, F. and Rahmadani, A. (2019) ‘Analisa Metode Classification-Decission Tree Dan Algoritma C.45 Untuk Memprediksi Penyakit Diabetes Dengan Menggunakan Aplikasi Rapid Miner’, *SINTECH (Science and Information Technology) Journal*, 2(1), pp. 10–17. doi: 10.31598/sintechjournal.v2i1.293.
- Fernández, A. *et al.* (2013) ‘Analysing the classification of imbalanced data-sets with multiple classes: Binarization techniques and ad-hoc approaches’, *Knowledge-Based Systems*, 42, pp. 97–110. doi: 10.1016/j.knosys.2013.01.018.
- M. Raihan Rafiiful Allaam, A. T. W. (2021) ‘KLASIFIKASI GENUS TANAMAN ANGGREK MENGGUNAKAN METODE CONVOLUTIONAL NEURAL NETWORK (CNN)’, in *e-Proceeding of Engineering*, pp. 3147–3179.
- Marlina, D. and Bakri, M. (2021) ‘Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4.5’, *Jurnal Teknologi dan Sistem Informasi (JTSI)*, 2(1), pp. 23–28.

- Musu, W., Ibrahim, A. and Heriadi (2021) ‘Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4 . 5’, *Prosiding Seminar Ilmiah Sistem Informasi Dan Teknologi Informasi*, X(1), pp. 186–195.
- Nasution, D. A., Khotimah, H. H. and Chamidah, N. (2019) ‘Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN’, *Computer Engineering, Science and System Journal*, 4(1), p. 78. doi: 10.24114/cess.v4i1.11458.
- Normawati, D. and Prayogi, S. A. (2021) ‘Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter’, *Jurnal Sains Komputer & Informatika (J-Sakti)*, 5(2), pp. 697–711.
- Novita, S., Harsani, P. and Qur’ania, A. (2018) ‘Penerapan K-Nearest Neighbor (KNN) untuk Klasifikasi Anggrek Berdasarkan Karakter Morfologi Daun dan Bunga’, *Komputasi*, 15(1), pp. 118–125.
- Pratama, A., Wihandika, R. C. and Ratnawati, D. E. (2018) ‘Implementasi algoritme support vector machine (SVM) untuk prediksi ketepatan waktu kelulusan mahasiswa’, *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 2(April), pp. 1704–1708.
- Putri, R. P. S. and Waspada, I. (2018) ‘Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika’, *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, 4(1), p. 1. doi: 10.23917/khif.v4i1.5975.
- Sikumbang, E. D. (2018) ‘Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori’, *Jurnal Teknik Komputer AMIK BSI (JTK)*, Vol 4, No.(September), pp. 1–4.
- Wahyono, W. and Nugroho, A. (2018) ‘Penerapan Algoritma C4.5 Untuk Prediksi Tingkat Kompetensi Karyawan Pt Multistrada Arah Sarana’, *Joutica*, 3(1), p. 145. doi: 10.30736/jti.v3i1.201.