

ANALISIS PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAIVE BAYES CLASSIFIER

Masungging Dwi Cahyo¹, Sri Redjeki², Erna Hudianti Pujiarini^{3*}

^{1,2,3} Universitas Teknologi Digital Indonesia, *Penulis Koresponden
e-mail: ¹masungging.dwi@students.utdi.ac.id, ²dzeky@utdi.ac.id, ³ernahudi@utdi.ac.id,

ABSTRACT

Graduating on time is an indicator of student success and the quality of higher education. Predicting graduation on time needs to be done to increase the efficiency of the education system in higher education. The Naive Bayes Classifier method is a classification method that can be used to predict whether students will graduate on time or not. The Naive Bayes Classifier method is based on Bayes' theorem with the assumption of independence between features. The research data used was 247 students who had graduated. The attributes used include gender, regional origin, Semester Achievement Index (IPS) 1 to 4, and competency course grades with four category target variables. For training data, 197 (80%) were taken, while 50 (20%) of the test data were taken. The research results show that the Naive Bayes Classifier model has high accuracy in predicting student graduation, namely 94%. The model precision values for category 1 and category 3 respectively reached 94% and 100%, while the model precision for category 2 and category 4 was 0. The recall value for category 1 and category 3 reached 100%, while for category 2 and category 4 it was 0. The F1-Score value shows a fairly good balance between precision and recall, namely 49%.

Keywords: Graduation, Classification, Naive Bayes Classifier

INTISARI

Kelulusan dalam waktu yang ditetapkan merupakan salah satu indikator keberhasilan mahasiswa dan kualitas Perguruan tinggi. Prediksi kelulusan tepat waktu perlu dilakukan guna meningkatkan efisiensi sistem pendidikan pada Perguruan Tinggi. Metode Naive Bayes Classifier merupakan metode klasifikasi yang dapat digunakan memprediksi mahasiswa akan lulus tepat waktu atau tidak. Metode Naive Bayes Classifier didasarkan pada teorema Bayes dengan asumsi independensi antar fitur. Data penelitian yang digunakan sebanyak 247 mahasiswa telah lulus. Atribut yang digunakan meliputi jenis kelamin, asal daerah, Indeks Prestasi Semester (IPS) 1 hingga 4, dan nilai mata kuliah kompetensi dengan variabel target empat kategori. Untuk data latih diambil sebanyak 197 (80%) sedangkan data uji diambil sebanyak 50 (20%). Hasil penelitian menunjukkan bahwa model Naive Bayes Classifier memiliki akurasi yang tinggi dalam memprediksi kelulusan mahasiswa, yaitu sebesar 94%. Nilai presisi model untuk kategori 1 dan kategori 3, masing-masing mencapai 94% dan 100% sedangkan presisi model untuk kategori 2 dan kategori 4 adalah 0. Nilai recall untuk kategori 1 dan kategori 3 mencapai 100% sedangkan untuk kategori 2 dan kategori 4 adalah 0. Nilai F1-Score menunjukkan keseimbangan yang cukup baik antara presisi dan recall, yaitu sebesar 49%.

Kata kunci: Kelulusan, Klasifikasi, Naive Bayes Classifier

1. PENDAHULUAN

Perguruan tinggi merupakan penyelenggara pendidikan tinggi (Setiyani, 2020). Sebagai penyelenggara pendidikan tinggi diharapkan mampu menghasilkan sumber daya manusia yang berkompeten (Nathan A. J., 2012). Salah satu program studi yang tersedia di Politeknik Kesehatan Bhakti Setya Indonesia adalah Program Studi Diploma III Teknologi Bank Darah. Pendidikan Diploma III Teknologi Bank Darah mempunyai visi untuk menghasilkan tenaga kesehatan sebagai teknisi pelayanan darah profesional dengan sebutan Ahli Madya Kesehatan (Kemenkes, 2018).

Lulusan Program Studi Diploma III Teknologi Bank Darah diharapkan mampu berkomunikasi, memberikan informasi, dan edukasi terkait pelayanan darah (Riawati, 2020). Salah satu aspek pengukuran kualitas perguruan tinggi dalam keberhasilan penyelenggaraan pendidikan tinggi adalah tingkat kelulusan mahasiswa (Setiyani, 2020).

Sebagai bagian dari upaya meningkatkan mutu pendidikan, Politeknik Kesehatan Bhakti Setya Indonesia telah menetapkan standar kelulusan bagi mahasiswa, yaitu lulus tepat waktu dalam enam semester atau tiga tahun dengan nilai IPK (Indeks Prestasi Kumulatif) minimal 3,00. Namun, data dari Program Studi Teknologi Bank Darah dalam tiga tahun terakhir menunjukkan bahwa 11% mahasiswa tidak berhasil lulus sesuai standar tersebut. Perlu ada sistem yang digunakan untuk memprediksi kelulusan mahasiswa sesuai dengan standar.

Naïve Bayes merupakan metode pengklasifikasi yang dapat memprediksi peluang untuk menjadi anggota kelas. Metode Naive Bayes dipilih karena dapat mengklasifikasikan data dengan tipe yang beragam (kategorikal dan numerik). Sehingga tidak diperlukan transformasi data sebelumnya, serta tidak memerlukan dataset yang terlalu besar.

Penelitian sejenis telah dilakukan (Diska, 2023), dengan menggunakan data training sebanyak 94 untuk data testing sebanyak 46 data. Atribut yang digunakan yaitu IPS semester 1 hingga 6, dan total sks. Diperoleh hasil accuracy 82,61%, precision 91,66%, dan recall 61,11%. Selanjutnya penelitian sejenis telah dilakukan (Riyadi, 2020) menggunakan data sebanyak 200 data dengan atribut kelamin, sks1, sks2, sks3, sks4, ipk1, ipk2, ipk3, dan ipk4. Data uji sebanyak 60% dan data latih sebanyak 40% dan diperoleh akurasi 91,86%. Penelitian sejenis juga telah dilakukan (Khasanah, 2022), menggunakan data sebanyak 379 data, dengan data training 303 data dan data testing 76 data. Diperoleh hasil penelitian akurasi = 88,16%, precision = 93,62% dan recall = 88%. Penelitian (Salmu, 2017) diperoleh hasil penelitian akurasi sebesar 80,72%, presisi 28,13%, recall 28,57%, dan F1-Score 32,53%. Data training yang digunakan sebanyak 1662 data dan data testing sebanyak 587 data. Penelitian (Qisthiano, 2021), dataset dibagi 70% untuk data training dan 30% sebagai data testing. Hasil penelitian ini merupakan akurasi dari model prediksi yang dilakukan, dimana hasil akurasi yang didapatkan adalah 0.8103 atau sebesar 81,03%. Penelitian (Purwati, 2021), atribut yang digunakan dalam penelitian ini yaitu meliputi jurusan, biaya, jenis kelamin, beasiswa, jumlah mata kuliah yang disetujui semester pertama, nilai rata-rata semester pertama, jumlah mata kuliah yang disetujui semester 2, dan nilai rata-rata semester 2. Menggunakan dataset sebanyak 500 data, kemudian dibagi menjadi 80% data training dan 20% data testing. Penelitian yang dilakukan mendapatkan hasil akurasi sebesar 95%, presisi 95,16%, recall 95%, dan F1-Score 95%. Penelitian yang dilakukan untuk prediksi kelulusan mahasiswa Program Studi DIII Teknologi Bank Darah di Politeknik Kesehatan Bhakti Setya Indonesia berbasis web menggunakan metode Naive Bayes Classifier (NBC).

2. METODE PENELITIAN

Penelitian ini menggunakan metode Naive Bayes Classifier untuk memprediksi kelulusan mahasiswa pada Program Studi Diploma III Teknologi Bank Darah di Poltekkes BSI Yogyakarta. Secara garis besar, alur penelitian ini dapat dibagi menjadi beberapa tahap mengacu pada metode pengembangan data mining yang digunakan untuk menganalisis data dalam penerapan data mining menggunakan proses tahapan *knowledge discovery in databases* (KDD).

2.1 Tahap pengumpulan data

Teknik pengumpulan data dengan mengambil arsip data sekunder yang diperoleh dari basis data akademik Politeknik Kesehatan Bhakti Setya Indonesia, Program Studi Diploma III Teknologi Bank Darah yang lulus dalam tiga tahun terakhir, yaitu tahun akademik 2020/2021, 2021/2022, dan 2022/2023. Dataset yang diambil sebanyak 258 data mahasiswa yang lulus.

2.2 Tahap Preprocessing Data

Dimulai dengan pembersihan data meliputi proses penghapusan atau perbaikan data yang tidak valid, duplikat, atau tidak lengkap. Juga dilakukan penanganan data kosong (missing values) dengan mengisi nilainya menggunakan median dari masing-masing atribut, serta menghapus data yang banyak memiliki nilai kosong, diperoleh dataset yang akan diolah sebanyak 247.

Tabel 1. Dataset yang Sudah Bersih

NO	NIM	NAMA	PROGRAM STUDI	JENIS KELAMIN	ASAL DAERAH	IPS_1	IPS_2	IPS_3	IPS_4	IPS
241	20114083	ACHMAD SAIF	D3 Teknologi Bank Darah	Pria	LUAR DIY	2.86	0.64	1.84	0.99	0.4
242	20114084	NUR RAHYANI RASYID	D3 Teknologi Bank Darah	Wanita	LUAR DIY	2.60	2.69	2.93	3.28	3.1
243	20114085	SEPTIAN YUDIYANSYAH OLI	D3 Teknologi Bank Darah	Pria	LUAR DIY	3.00	2.56	2.98	3.12	2.1
244	20114086	WA SARIANI	D3 Teknologi Bank Darah	Wanita	LUAR DIY	2.68	2.80	3.34	2.94	2.1
245	20114087	HARNIYANTI ODE	D3 Teknologi Bank Darah	Wanita	LUAR DIY	2.74	2.84	3.40	3.04	2.1
246	20114088	SUHARJO	D3 Teknologi Bank Darah	Pria	LUAR DIY	2.47	2.55	3.20	3.04	2.1
247	20114089	ISKANDAR ALI TUASIKAL	D3 Teknologi Bank Darah	Pria	LUAR DIY	2.55	0.36	1.26	2.83	2.1

Selanjutnya, tahap seleksi data atau seleksi fitur (feature selection) dilakukan untuk memilih variabel-variabel yang paling berpengaruh dalam proses prediksi. Pada tahap ini, atribut yang tidak digunakan akan dihapus, seperti No, Nim, Nama, Program Studi, IPS5, IPS6, IPK Lulus, Dan Nilai Karya Tulis. Atribut yang digunakan yaitu jenis kelamin, asal daerah, IPS1, IPS2, IPS3, IPS4, nilai Mata Kuliah Kompetensi, yaitu meliputi Serologi Golongan Darah I, Serologi Golongan Darah II, Serologi Golongan Darah III, Serologi Golongan Darah IV, Infeksi Menular Lewat Transfusi Darah I, Infeksi Menular Lewat Transfusi Darah II, Infeksi Menular Lewat Transfusi Darah III, Infeksi Menular Lewat Transfusi Darah IV, Penyadapan Darah, dan Komponen Darah. Dan label kelasnya ada 4 kategori yaitu :

- label kelas = 1 untuk mahasiswa lulus dalam 3 tahun dan $IPK \geq 3,00$
- label kelas = 2 untuk mahasiswa lulus dalam 3 tahun dan $IPK < 3,00$
- label kelas = 3 untuk mahasiswa lulus lebih dari 3 tahun dan $IPK \geq 3,00$
- label kelas = 4 untuk mahasiswa lulus lebih dari 3 tahun dan $IPK < 3,00$

Dataset hasil seleksi seperti tampak pada Tabel 2.

Tabel 2. Dataset Hasil Diseleksi

JENIS KELAMIN	ASAL DAERAH	IPS_1	IPS_2	IPS_3	IPS_4	SEROLOGI GOLONGAN DARAH I	INFEKSI MENULAR LEWAT TRANSFUSI DARAH I	SEROLOGI GOLONGAN DARAH II	INFEKSI MENULAR LEWAT TRANSFUSI DARAH II	PENYADAPAN DARAH	SEROLOGI GOLONGAN DARAH III	INFEKSI MENULAR LEWAT TRANSFUSI DARAH III	KOMPONEN DARAH	SEROLOGI GOLONGAN DARAH IV	INFEKSI MENULAR LEWAT TRANSFUSI DARAH IV	LABEL KELAS
Wanita	DI YOGYAKARTA	3.50	2.93	3.43	2.98	C	B	B	A	A	B	B	B	C	A	3
Pria	LUAR DIY	3.30	3.50	2.94	2.86	B	C	B	B	A	A	B	B	B	B	3
Pria	LUAR DIY	2.09	0.65	1.53	2.98	B	B	B	B	A	B	B	B	B	A	1
Wanita	LUAR DIY	1.32	0.91	1.95	1.96	B	B	B	B	A	B	B	B	B	A	1
Wanita	LUAR DIY	3.32	2.76	2.42	2.00	C	C	C	B	B	C	B	B	C	A	4
Pria	LUAR DIY	3.41	3.09	3.38	1.58	C	B	B	B	A	B	C	B	B	A	1
Wanita	LUAR DIY	2.86	2.86	2.88	1.60	C	B	B	B	B	C	C	B	B	B	2
Wanita	LUAR DIY	3.36	3.00	3.81	2.16	B	B	A	B	A	B	A	B	A	A	1

2.3 Tahap Perhitungan Naïve Bayes

Metode Naive Bayes untuk memprediksi kelulusan mahasiswa berdasarkan dataset yang telah dikumpulkan. Metode Naive Bayes Classifier dilakukan dengan mencari nilai probabilitas paling tinggi untuk mengklasifikasikan sebuah data uji dengan kategori yang tepat. Teknik prediksi probabilitas yang sederhana didasarkan pada penerapan teorema bayes yang secara umum dinyatakan sebagai berikut:

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots\dots\dots(1)$$

dimana :

- X : Data dengan kelas yang belum diketahui
- H : Hipotesis data X merupakan suatu label kelas tertentu
- P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X (Probabilitas Posterior)
- P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H (Probabilitas Likelihood)
- P(H) : Probabilitas hipotesis H (Probabilitas Prior)
- P(X) : Probabilitas X (Probabilitas Evidence)

Probabilitas posterior menunjukkan kemungkinan seorang mahasiswa dengan karakteristik data (X) untuk masuk ke dalam masing-masing label kelas (1, 2, 3, atau 4). Data mahasiswa akan diklasifikasikan ke dalam kelas dengan probabilitas posterior yang tertinggi.

2.4 Tahap Evaluasi Model

Tahap terakhir dalam pengembangan model adalah evaluasi model prediksi dengan menggunakan matriks konfusi (confusion matrix) untuk menghitung nilai akurasi, presisi, recall, dan F1-score dari model yang dibangun. Nilai akurasi untuk mengukur seberapa sering model klasifikasi membuat prediksi yang benar dari semua prediksi yang dibuat.

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

Nilai presisi untuk mengukur seberapa banyak prediksi positif yang benar dari semua prediksi positif yang dibuat oleh model.

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

Nilai recall untuk mengukur seberapa banyak kasus yang sebenarnya positif yang benar diprediksi oleh model.

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

F1-Score adalah rata-rata harmonis antara presisi dan recall, yang memberikan penilaian seimbang terhadap kedua metrik tersebut.

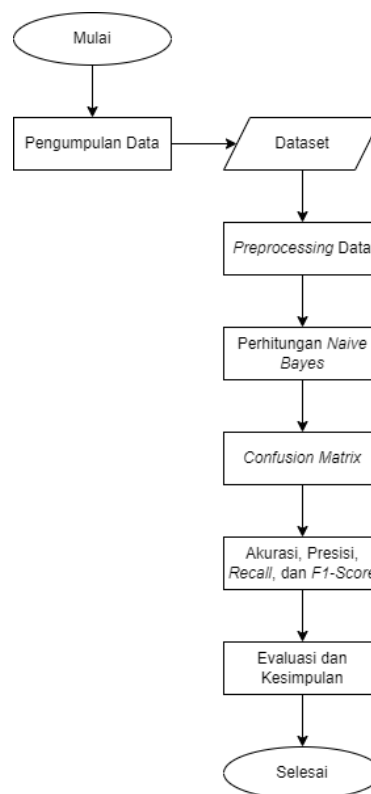
$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \dots\dots\dots(5)$$

Karena label kelas dalam penelitian ini menggunakan 4 label kelas, maka confusion matrix merupakan matriks 4x4. Setiap baris mewakili jumlah instance kelas sebenarnya, dan setiap kolom mewakili jumlah instance kelas yang diprediksi. Bentuk tabel confusion matrix 4x4 seperti tampak pada Tabel 3.

Tabel 3. Struktur Confusion Matrix 4 × 4

Prediksi Aktual	A	B	C	D
A	TP	FN	FN	FN
B	FP	TP	FN	FN
C	FP	FP	TP	FN
D	FP	FP	FP	TP

Tahapan yang dilakukan dalam penelitian terlihat pada Gambar 1.



Gambar 1. Desain Penelitian

3. HASIL DAN PEMBAHASAN

Dataset yang diambil sebanyak 258 data mahasiswa yang lulus dibagi data menjadi data latih (training data) sebanyak 197 data (80%) dan data uji (testing data) sebanyak 50 data (20%). Fitur dipakai untuk klasifikasi adalah

Jenis Kelamin, Daerah Asal, IPS1, IPS2, IPS3, IPS4, Nilai matakuliah Serologi Golongan Darah I, Nilai matakuliah Serologi Golongan Darah II, Nilai matakuliah Serologi Golongan Darah III, Nilai matakuliah Serologi Golongan Darah IV, Nilai matakuliah Infeksi Menular Lewat Transfusi Darah I, Nilai matakuliah Infeksi Menular Lewat Transfusi Darah II, Nilai matakuliah Infeksi Menular Lewat Transfusi Darah III, Nilai matakuliah Infeksi Menular Lewat Transfusi Darah IV, Nilai matakuliah Penyesapan Darah, dan Nilai matakuliah Komponen Darah. Dan atribut kelas nya adalah Kelulusan, dengan kategorinya ada 4, kategori 1(Lulus dalam 3 tahun dan IPK \geq 3,00), kategori 2(Lulus dalam 3 tahun dan IPK $<$ 3,00), kategori 3(Lulus lebih dari 3 tahun dan IPK \geq 3,00), kategori 4(Lulus lebih dari 3 tahun dan IPK $<$ 3,00).

Tabel 3. Jumlah Atribut Pada Setiap Kelas Dalam *Data Training*

Atribut		Label Kelas				Total
		1	2	3	4	
Jenis Kelamin	Laki-laki	41	7	2	2	52
	Perempuan	135	4	2	4	145
Asal Daerah	DIY	14	0	1	2	17
	Luar DIY	164	11	3	5	180
IPS 1	\geq 3.00	142	1	4	1	148
	$<$ 3.00	34	10	0	5	49
IPS 2	\geq 3.00	102	0	1	0	103
	$<$ 3.00	74	11	3	6	94
IPS 3	\geq 3.00	153	6	2	0	161
	$<$ 3.00	23	5	2	6	36
IPS 4	\geq 3.00	96	1	1	0	98
	$<$ 3.00	80	10	3	6	99
Serologi Golongan Darah I	A	22	1	0	0	23
	B	127	2	1	0	130
	C	27	8	3	6	44
	D	0	0	0	0	0
Infeksi Menular Lewat Transfusi Darah I	A	35	0	0	0	35
	B	124	4	2	2	132
	C	17	7	2	4	30
	D	0	0	0	0	0
Serologi Golongan Darah II	A	26	0	0	0	26
	B	134	6	3	2	145
	C	16	5	1	4	26
	D	0	0	0	0	0
Infeksi Menular Lewat Transfusi Darah II	A	31	0	1	0	32
	B	129	2	2	5	138
	C	16	9	1	1	27
	D	0	0	0	0	0
Penyesapan Darah	A	132	5	3	0	140
	B	41	5	0	2	48
	C	3	1	1	4	9
	D	0	0	0	0	0
Serologi Golongan Darah III	A	37	1	1	0	39
	B	116	6	2	2	126
	C	23	4	1	4	32
	D	0	0	0	0	0
Infeksi Menular Lewat Transfusi Darah III	A	49	0	0	0	49
	B	102	4	4	2	112
	C	25	7	0	4	36
	D	0	0	0	0	0
Komponen Darah	A	57	1	1	0	59
	B	102	6	2	1	111
	C	17	4	1	5	27
	D	0	0	0	0	0
Serologi Golongan Darah IV	A	42	0	0	0	42
	B	93	6	3	0	102

Atribut		Label Kelas				Total
		1	2	3	4	
	C	41	5	1	6	53
	D	0	0	0	0	0
Infeksi Menular Lewat Transfusi Darah IV	A	123	4	3	2	132
	B	45	5	1	3	54
	C	8	2	0	1	11
	D	0	0	0	0	0

Tahap terakhir dalam pengembangan model adalah evaluasi model prediksi dengan menggunakan matriks konfusi (*confusion matrix*) untuk menghitung akurasi, presisi, *recall*, dan *F1-score* dari model yang dibangun.

Tabel 4. *Confusion Matrix*

Prediksi \ Aktual	Kelas 1	Kelas 1	Kelas 1	Kelas 1
Kelas 1	46	0	0	0
Kelas 2	3	0	0	0
Kelas 3	0	0	1	0
Kelas 4	0	0	0	0

3.1 Akurasi

Akurasi menunjukkan kemampuan model dalam memprediksi data dengan benar secara keseluruhan. Akurasi dihitung dengan membagi jumlah prediksi benar (*true positive* dan *true negative*) dengan total jumlah data.

$$Akurasi = \frac{TP_1 + TP_2 + TP_3 + TP_4}{Total Instance} = \frac{46 + 0 + 1 + 0}{50} = 0,94$$

3.2 Presisi

Presisi menunjukkan tingkat ketepatan model dalam memprediksi data positif. Presisi dihitung dengan membagi jumlah *true positive* dengan jumlah *true positive* dan *false positive*.

$$Presisi\ Kelas\ 1 = \frac{TP_1}{TP_1 + FP_1} = \frac{46}{46 + 3 + 0 + 0} = 0,94$$

$$Presisi\ Kelas\ 2 = \frac{TP_2}{TP_2 + FP_2} = \frac{0}{0 + 0 + 0 + 0} = 0$$

$$Presisi\ Kelas\ 3 = \frac{TP_3}{TP_3 + FP_3} = \frac{1}{0 + 0 + 1 + 0} = 1$$

$$Presisi\ Kelas\ 4 = \frac{TP_4}{TP_4 + FP_4} = \frac{0}{0 + 0 + 0 + 0} = 0$$

3.3 Recall

Recall menunjukkan tingkat kelengkapan model dalam mendeteksi data positif. Recall dihitung dengan membagi jumlah *true positive* dengan jumlah *true positive* dan *false negative*.

$$Recall\ Kelas\ 1 = \frac{TP_1}{TP_1 + FN_1} = \frac{46}{46 + 0 + 0 + 0} = 1$$

$$Recall\ Kelas\ 2 = \frac{TP_2}{TP_2 + FN_2} = \frac{0}{3 + 0 + 0 + 0} = 0$$

$$Recall\ Kelas\ 3 = \frac{TP_3}{TP_3 + FN_3} = \frac{1}{0 + 0 + 1 + 0} = 1$$

$$Recall\ Kelas\ 4 = \frac{TP_4}{TP_4 + FN_4} = \frac{0}{0 + 0 + 0 + 0} = 0$$

3.4 F1-Score

F1-Score adalah rata-rata harmonis antara presisi dan *recall*, yang memberikan penilaian seimbang terhadap kedua metrik tersebut.

$$F1 - Score\ Kelas\ 1 = 2 \times \frac{Presisi_{Kelas\ 1} \times Recall_{Kelas\ 1}}{Presisi_{Kelas\ 1} + Recall_{Kelas\ 1}} = 2 \times \frac{0,94 \times 1}{0,94 + 1} = 0,97$$

$$F1 - \text{Score Kelas 2} = 2 \times \frac{\text{Presisi}_{\text{Kelas 2}} \times \text{Recall}_{\text{Kelas 2}}}{\text{Presisi}_{\text{Kelas 2}} + \text{Recall}_{\text{Kelas 2}}} = 2 \times \frac{0 \times 0}{0 + 0} = 0$$

$$F1 - \text{Score Kelas 3} = 2 \times \frac{\text{Presisi}_{\text{Kelas 3}} \times \text{Recall}_{\text{Kelas 3}}}{\text{Presisi}_{\text{Kelas 3}} + \text{Recall}_{\text{Kelas 3}}} = 2 \times \frac{1 \times 1}{1 + 1} = 1$$

$$F1 - \text{Score Kelas 4} = 2 \times \frac{\text{Presisi}_{\text{Kelas 4}} \times \text{Recall}_{\text{Kelas 4}}}{\text{Presisi}_{\text{Kelas 4}} + \text{Recall}_{\text{Kelas 4}}} = 2 \times \frac{0 \times 0}{0 + 0} = 0$$

F1-Score Rata-rata :

$$\begin{aligned} &= \frac{F1 - \text{Score}_{\text{Kelas 1}} + F1 - \text{Score}_{\text{Kelas 2}} + F1 - \text{Score}_{\text{Kelas 3}} + F1 - \text{Score}_{\text{Kelas 4}}}{4} \\ &= \frac{0,97 + 0 + 1 + 0}{4} \\ &= 0,49 \end{aligned}$$

Analisis hasil evaluasi model bertujuan untuk menganalisis dan menarik kesimpulan dari hasil evaluasi model, serta memberikan rekomendasi untuk perbaikan atau pengembangan model selanjutnya.

- Pengujian akurasi sistem dilakukan untuk mengetahui seberapa akurat sistem dalam memprediksi kelulusan mahasiswa. Pengujian ini dilakukan dengan membandingkan hasil prediksi sistem menggunakan algoritma Naive Bayes Classifier. Hasil perhitungan akurasi model adalah 0,94 atau 94%, menunjukkan bahwa model secara keseluruhan mampu memprediksi kelulusan mahasiswa dengan sangat baik, dengan tingkat kesalahan prediksi 6%.
- Presisi model untuk Kelas 1 dan Kelas 3 sangat baik, masing-masing mencapai 94% dan 100%. Hal ini mengindikasikan bahwa model dapat memprediksi dengan tepat mahasiswa yang termasuk dalam kelas-kelas tersebut. Namun, presisi model untuk Kelas 2 dan Kelas 4 adalah 0, yaitu menunjukkan bahwa model belum dapat memprediksi dengan benar kelas-kelas ini. Rendahnya presisi juga kemungkinan menunjukkan bahwa model memiliki kesulitan dalam membedakan karakteristik mahasiswa di Kelas 2 dan Kelas 4, sehingga perlu dilakukan perbaikan pada model.
- Dari sisi recall, model juga menunjukkan performa yang sangat baik untuk Kelas 1 dan Kelas 3, dengan nilai recall mencapai 100% untuk kedua kelas tersebut. Ini berarti model dapat mendeteksi dengan tepat mahasiswa yang termasuk dalam Kelas 1 dan Kelas 3. Namun recall model untuk Kelas 2 dan Kelas 4 adalah 0, artinya model belum dapat mendeteksi dengan benar kelas-kelas ini. Rendahnya recall juga kemungkinan menunjukkan bahwa model memiliki kesulitan dalam mengenali karakteristik mahasiswa di Kelas 2 dan Kelas 4, sehingga perlu dilakukan perbaikan pada model.
- Nilai rata-rata F1-score sistem yang diperoleh melalui perhitungan metrik yaitu sebesar 49%. F1-Score merupakan ukuran keseimbangan antara presisi dan recall, yang menggambarkan kinerja sistem secara keseluruhan. Nilai F1-score sebesar 49% termasuk dalam kategori sedang. Hal ini menunjukkan bahwa masih terdapat ruang untuk meningkatkan kinerja model agar dapat memprediksi dengan lebih baik secara keseluruhan.

4. KESIMPULAN

Metode Naive Bayes Classifier telah berhasil diterapkan untuk memprediksi kelulusan mahasiswa. Hasil dari implementasi ini menunjukkan bahwa model yang dibangun mampu mengklasifikasikan mahasiswa ke dalam kategori label kelas yang tersedia berdasarkan data historis akademik mereka.

Model Naive Bayes yang diimplementasikan memiliki performa yang cukup baik, dengan akurasi model 0,94 atau 94%, dalam memprediksi kelulusan mahasiswa.

Model ini berhasil memprediksi dengan tepat mahasiswa yang termasuk dalam Kelas 1 dan Kelas 3, dengan nilai presisi yang tinggi, yaitu 94% untuk Kelas 1 dan 100% untuk Kelas 3.

Dari sisi recall, model juga menunjukkan performa yang sangat baik untuk Kelas 1 dan Kelas 3, dengan nilai recall mencapai 100% untuk kedua kelas ini. Namun, model masih mengalami kesulitan dalam memprediksi dan mendeteksi mahasiswa yang termasuk dalam Kelas 2 dan Kelas 4, dengan nilai presisi dan recall yang rendah, yaitu 0 untuk kedua kelas tersebut.

Nilai F1-score keseluruhan dari model sebesar 49% menunjukkan bahwa keseimbangan antara presisi dan recall berada pada kategori sedang. Ini berarti bahwa sistem masih memiliki ruang untuk peningkatan, terutama dalam meningkatkan kemampuan deteksi pada Kelas 2 dan Kelas 4.

Secara keseluruhan, sistem prediksi yang dibangun telah menunjukkan kinerja yang baik, terutama dalam memprediksi Kelas 1 (lulus dalam 3 tahun dengan $IPK \geq 3,00$) dan Kelas 3 (lulus lebih dari 3 tahun dengan $IPK \geq 3,00$). Namun, perlu adanya perbaikan lebih lanjut untuk meningkatkan prediksi pada Kelas 2 (lulus dalam 3 tahun dengan $IPK < 3,00$) dan Kelas 4 (lulus lebih dari 3 tahun dengan $IPK < 3,00$) agar hasil prediksi menjadi lebih akurat dan menyeluruh.

UCAPAN TERIMA KASIH

Ucapan terimakasih disampaikan kepada Politeknik Kesehatan Bhakti Setya Indonesia, khususnya Program Studi Diploma III Teknologi Bank Darah. Pendidikan Diploma III Teknologi Bank Darah yang telah berpartisipasi dalam penelitian ini hingga selesai dilaksanakan dengan baik dan lancar. Ucapan terima kasih kami sampaikan juga kepada Universitas Teknologi Digital Indonesia pemberi dana penelitian.

DAFTAR PUSTAKA

- Diska, K. R. (2023). Sistem Informasi Prediksi Kelulusan Menggunakan Metode Naive Bayes Classifier (Studi Kasus: Prodi Pendidikan Teknik Informatika). . *Jurnal Pendidikan Tambusai*, 936–943.
- Gunawan, M. Z. (2021). Analisis Komparasi Algoritma Naive Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu. . *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 513–523.
- Kemenkes, R. (2018). *Kurikulum Inti Pendidikan Diploma III Teknologi Bank Darah*. Jakarta: Kementerian Kesehatan RI.
- Khasanah, N. S. (2022). Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes. *Technologia* , 207-215.
- Nathan A. J., S. A. (2012). Model Algoritma K-nearest Neighbor untuk memprediksi kelulusan mahasiswa. *Foreign Affairs*, 1–9.
- Purwati, N. &. (2021). Prediksi Tingkat Kelulusan Mahasiswa dengan Algoritma Naive Bayes. *Pepadun*, 126–137.
- Qisthiano, M. R. (2021). Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naive Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 987–994.
- Rahmatullah, S. (2019). Prediksi Tingkat Kelulusan Tepat Waktu Dengan Metode Naive Bayes Dan K-Nearest Neighbor. *Jurnal Informasi Dan Komputer*, 7–16.
- Riawati, D. (2020). Analisis Nilai Praktikum Edukator Pelayanan Darah Terhadap Nilai Ujian Akhir Semester (UAS) Komunikasi Informasi Edukasi (KIE) Dan Promosi Kesehatan. *Avicenna : Journal of Health Research* , 72–79.
- Riyadi, F. A. (2020). Implementasi Metode Naive Bayes Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu Prodi Informatika (Studi Kasus : Universitas Teknologi Yogyakarta). . <http://eprints.uty.ac.id/4863/>, 1-9.
- Salmu, S. &. (2017). Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu Menggunakan Naive Bayes : Studi Kasus UIN Syarif Hidayatullah Jakarta. *Seminar Nasional Multidisiplin Ilmu (SENMI)*, 701–709.
- Setiyani, L. W. (2020). Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Data Mining Naive Bayes. *Systematic Review Faktor Exacta*, 35–43.