

Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa Teknik Informatika dengan Orange Data Mining

Iqlimah Attyyatullatifah¹, Mia Kamayani²

iqlimahattyatullatifah0372@gmail.com, mia.kamayani@uhamka.ac.id

Universitas Muhammadiyah Prof. DR. HAMKA

Informasi Artikel	Abstrak
Diterima : 4 Mar 2024 Direview : 13 Mar 2024 Disetujui : 8 Mar 2024	Penyelesaian studi tepat waktu merupakan indikator penting dalam menilai kompetensi lulusan. Meskipun demikian, muncul tantangan karena tidak semua mahasiswa dapat menyelesaikan studi mereka sesuai jadwal yang telah ditentukan. Penelitian ini mengembangkan model prediksi status kelulusan mahasiswa menggunakan empat algoritma klasifikasi: <i>Decision Tree</i> , <i>Naïve Bayes</i> , K-NN, dan SVM. Data penelitian mencakup 500 data mahasiswa angkatan 2018-2020 di Universitas Muhammadiyah Prof. Dr. Hamka, dengan 60% data latihan dan 40% data uji. Analisis dilakukan menggunakan perangkat lunak <i>Orange Data Mining</i> , dengan evaluasi menggunakan <i>K-Fold Cross Validation</i> ($k=5$), <i>Confusion Matrix</i> , dan ROC. Hasil analisis menunjukkan bahwa model K-NN memiliki performa tertinggi dengan akurasi 92%, recall 90%, dan presisi 92%. <i>Decision Tree</i> menempati posisi kedua dengan akurasi 90%, presisi 87%, dan recall 90%. SVM mencapai akurasi sebesar 84%, dengan presisi 90% dan recall 73%. Sementara itu, model <i>Naïve Bayes</i> menunjukkan akurasi 83%, presisi 80%, dan recall 83%.
Kata Kunci	
Prediksi Kelulusan, Perbandingan Algoritma, Klasifikasi, Orange, Data Mining	

Keywords	Abstrak
Graduation Prediction, Algorithm Comparison, Classification, Orange, Data Mining	<i>Timely completion of studies is an important indicator in assessing the competence of graduates. However, the challenge arises because not all students can complete their studies according to the predetermined schedule. This research develops a prediction model for student graduation status using four classification algorithms: Decision Tree, Naïve Bayes, K-NN, and SVM. The research data includes 500 student data for the 2018-2020 batch at Universitas Muhammadiyah Prof. Dr. Hamka, with 60% training data and 40% test data. The analysis was conducted using Orange Data Mining software, with evaluation using K-Fold Cross Validation ($k=5$), Confusion Matrix, and ROC. The analysis results show that the K-NN model has the highest performance with 92% accuracy, 90% recall, and 92% precision. Decision Tree came second with 90% accuracy, 87% precision, and 90% recall. SVM achieved 84% accuracy, with 90% precision and 73% recall. Meanwhile, the Naïve Bayes model showed 83% accuracy, 80% precision, and 83% recall.</i>

A. Pendahuluan

Dalam perubahan teknologi yang berlangsung cepat, peran mahasiswa Teknik Informatika semakin krusial, karena mereka menjadi agen perubahan utama yang akan membentuk dan memimpin dunia digital di masa mendatang. Sangat penting bagi mereka untuk menguasai ilmu teknik informatika dengan baik dan menyelesaikan studi sebagaimana ketetapan pendidikan tinggi yaitu delapan semester atau rentang waktu 4 tahun [1].

Meskipun begitu, kenyataannya, tingkat kelulusan mahasiswa di Program Studi Teknik Informatika tidak selalu mencapai tingkat yang diharapkan. Alasan di balik rendahnya tingkat kelulusan mahasiswa bisa sangat bervariasi, mulai dari kendala akademik hingga faktor eksternal yang memengaruhi kemampuan mereka untuk mengejar jadwal studi yang ditetapkan. Permasalahan ini meliputi tingkat kelulusan yang rendah di Program Studi Teknik Informatika, yang menjadi tantangan signifikan dalam dunia pendidikan tinggi.

Universitas Muhammadiyah Prof. Dr. Hamka berdiri sebagai institusi pendidikan tinggi swasta yang berlokasi di Jakarta. Universitas ini berdedikasi untuk memberikan pendidikan berkualitas tinggi, terbukti dengan beragamnya program studi di seluruh cabang keilmuan, termasuk Program Studi Teknik Informatika. Dalam dunia pendidikan tinggi yang terus berkembang, tantangan terus terjadi seiring berjalannya waktu dan cenderung kompleks. Isu-isu seperti tingkat kelulusan mahasiswa, peningkatan kualitas pendidikan, dan adaptasi terhadap kemajuan teknologi informasi telah menjadi perhatian utama Universitas Muhammadiyah Prof. Dr. Hamka.

Menurut hasil penelitian yang dipublikasikan oleh Anugrah Putra dan Kamayani, mengungkapkan tantangan serius dalam mencapai tingkat kelulusan mahasiswa Universitas Muhammadiyah Prof. Dr. Hamka sesuai rencana. Data tersebut menunjukkan bahwa tingkat kelulusan pada tahun angkatan 2013 adalah sekitar 22.36%, dimana hanya 17 dari total 76 mahasiswa aktif yang berhasil menyelesaikan studinya tepat waktu. Pada tahun 2014, tingkat kelulusan mencapai sekitar 11.23%, tetapi hanya 10 dari 89 mahasiswa yang dapat menyelesaikan studi tepat waktu. Sementara pada mahasiswa angkatan 2015, tingkat kelulusan hanya sekitar 7.69%, di mana hanya 8 dari total 104 mahasiswa yang berhasil menyelesaikan studi tepat waktu. Ketidaksesuaian tingkat kelulusan ini dengan target yang telah ditetapkan, yaitu mencapai 50% pada program studi Informatika [2]. Jika hal tersebut tidak diatasi dengan benar, maka penumpukan jumlah mahasiswa yang tidak lulus dalam jangka waktu yang diharapkan atau terlambat akan menjadi masalah yang semakin serius.

Mengacu pada situasi tersebut, peneliti bermaksud untuk menjalankan penelitian terkait dengan klasifikasi data prediksi kelulusan mahasiswa. Dalam kerangka penelitian ini, pengembangan prediksi status kelulusan mahasiswa akan dilakukan dengan membandingkan empat model klasifikasi, seperti *Decision Tree*, *Naïve Bayes*, K-NN dan SVM. Penelitian ini akan melibatkan penambahan variabel-variabel seperti asal sekolah, kota dan total sks untuk meningkatkan akurasi prediksi tingkat kelulusan mahasiswa.

Penelitian-penelitian sebelumnya telah mengeksplorasi penerapan model algoritma klasifikasi dalam memprediksi kelulusan mahasiswa. Penelitian yang dilakukan oleh Hozairi, Anwari, dan Syariful Alim, yang melibatkan 35 data uji,

menunjukkan tingkat akurasi algoritma K-NN sebesar 77%, *Decision Tree* sebesar 74%, dan *Naïve Bayes* sebesar 89% [3]. Penelitian Muhammad Gunawan, Muhammad Zarlis, dan Roslina, yang menganalisis 30 data uji menggunakan perangkat lunak Weka, menghasilkan tingkat akurasi yang sama antara algoritma KNN dan *Naïve Bayes* yaitu sebesar 90% [4]. Selain itu, penelitian Sri Widaningsih dalam studi kasus yang sama dengan menggunakan perangkat lunak *Rapidminer*, menunjukkan hasil bahwa algoritma C4.5 memiliki akurasi sebesar 75%, *Naïve Bayes* sebesar 76%, K-NN sebesar 68%, dan SVM sebesar 74% [5].

Penerapan algoritma klasifikasi untuk menganalisis data historis mahasiswa diharapkan dapat membantu mengevaluasi apakah mahasiswa akan lulus tepat waktu atau mengalami keterlambatan [6], serta dapat memberikan kontribusi yang lebih substansial terhadap pemahaman faktor-faktor yang mempengaruhi tingkat kelulusan mahasiswa. Dengan demikian, langkah ini diharapkan dapat membantu perguruan tinggi untuk merancang strategi yang lebih efektif dalam meningkatkan tingkat kelulusan serta mengurangi angka *drop-out* [7].

B. Metode Penelitian

Adapun alur dari penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

1. Pengumpulan Data

Pengumpulan data untuk penelitian ini melibatkan interaksi langsung dengan staf akademik untuk mendapatkan wawasan praktis tentang manajemen data akademik. Selain itu, tinjauan pustaka berkontribusi dalam membangun landasan teori yang kuat untuk menerapkan algoritma klasifikasi pada 500 dataset mahasiswa UHAMKA periode angkatan 2018-2020.

2. Preprocessing Data

Tahapan selanjutnya yang dilakukan adalah pemilihan data tanpa menghilangkan data atau menghapus atribut yang ada untuk memudahkan dalam memprediksi status kelulusan mahasiswa Teknik Informatika [8].

3. Perancangan Model Learning

Pada tahap ini, model klasifikasi dikonfigurasi dan dirancang menggunakan aplikasi *Orange Data Mining* [9], mencakup model *Decision Tree*, *Naïve Bayes*, K-NN dan SVM.

3.1 Decision Tree

Decision Tree merupakan algoritma struktur pohon yang mewakili aturan klasifikasi. Struktur ini memvisualisasikan atribut sebagai simpul dan kelas sebagai daun, dengan setiap pohon memiliki akar di bagian atasnya. Kejelasan visual dan interpretabilitas bawaannya membuat *Decision Tree* sangat ahli dalam tugas prediksi, menawarkan pemahaman intuitif tentang aturan klasifikasi [10],[11]. Dalam penghitungan nilai gain, nilai entropi sangat diperlukan dan dapat ditentukan dengan menggunakan rumus berikut:

$$\text{Entropy (S)} = \sum_{i=1}^n p_i * \log_2 p_i \quad (1)$$

Penjelasan

P_i : Proposi dari S_i terhadap S
 n : Jumlah patisi S
 A : Atribut
 S : Himpunan Kasus

Adapun rumus matematis untuk menghitung information gain adalah sebagai berikut:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (2)$$

Penjelasan

$|S|$: Jumlah kasus dalam himpunan S
 $A|S_i$: Jumlah kasus dalam partisi ke- i
 P_i : Pernyataan dari S_i terhadap seluruh himpunan S
 N : Jumlah patisi dalam himpunan S
 A : Atribut
 S : Himpunan Kasus

3.2 Naive Bayes

Naïve Bayes berdisi sebagai pendekatan klasifikasi probalistik langsung [12]. Algoritma ini menghitung berbagai peluang dalam menganalisis itensitas serta permutasi nilai yang ada di dalam himpunan data yang telah disediakan [13],[14]. *Naïve Bayes* dapat digunakan dalam berbagai tugas seperti prediksi kelulusan, penyaringan spam, analisis sentimen, dan kategorisasi dokumen. Kekuatan utamanya terletak pada perolehan parameter mean dan variance yang penting untuk klasifikasi dengan jumlah data pelatihan yang relatif sedikit. Rumus untuk teorema Bayes, sebuah komponen fundamental dari *Naïve Bayes*, dinyatakan sebagai berikut [15]:

$$P(X|H) = \frac{P(X|H).P(H)}{P(X)} \quad (3)$$

Penjelasan

$P(X)$: akumulasi *probability* X
 $P(X|H)$: *Probability* X terhadap hipotesa H
 $P(H)$: *Probability* hipotesa H sebelum mempertimbangkan data (Prior Probabilitas)
 $P(H|X)$: *Probability* hipotesa H saat keadaan X (Posteriori Probabilitas)
 H : Hipotesis yang menunjukkan afiliasi data dengan kelas tertentu
 X : Data dengan kelas yang tidak teridentifikasi

3.3 K-Nearest Neighbor (K-NN)

Algoritma K-NN merupakan pendekatan terawasi yang bergantung pada data pelatihan untuk mengkategorikan objek berdasarkan kedekatan jaraknya. Konsep operasional K-NN berkisar pada identifikasi tetangga terdekat dengan menilai interval terdekat antara data dan k tetangga [16],[17]. Algoritma K-NN sangat penting untuk keberhasilan aplikasinya dalam berbagai domain, seperti pengenalan pola, klasifikasi gambar, dan sistem rekomendasi. Teorema K-NN, yang menguraikan metode untuk mengukur jarak, secara umum dapat dinyatakan sebagai berikut:

$$d_i = \sqrt{\sum_{j=1}^n (x_{ij} - p_j)^2} \quad (4)$$

Penjelasan

- n : Total Sampel
 P_j : Data Input dari variabel ke-j
 x_{ij} : Data Sampel Pengetahuan
 d_i : Interval Sampel

3.4 Support Vector Machine (SVM)

Algoritma SVM adalah jenis algoritma pembelajaran terawasi yang digunakan untuk tugas-tugas yang melibatkan klasifikasi atau regresi data. SVM beroperasi dengan mengidentifikasi hyperplane yang paling efektif di dalam ruang input, yang bertindak sebagai pemisah optimal antara dua kelas yang berbeda [18]. SVM bertujuan untuk menemukan garis pemisah atau permukaan yang secara maksimal memisahkan dua kelas dalam dataset [5]. Kemampuan SVM dalam mengidentifikasi batas keputusan yang kompleks membuatnya menjadi alat yang serbaguna dan diterapkan secara luas di berbagai bidang seperti pembelajaran mesin, pengenalan pola, dan analisis data.

4. Pengujian Model *Learning*

Pada tahap pengujian model, setelah persiapan data dan desain model, algoritma klasifikasi dijalankan dengan menerapkan validasi silang *K-Fold Cross Validation* menggunakan aplikasi *Orange Data Mining*.

5. Evaluasi

Dalam tahap ini, dilakukan perhitungan akurasi dengan *Confusion Matrix* yang berfungsi sebagai instrumen penilaian untuk mengukur keefektifan model dengan menyandingkan hasil klasifikasi dengan nilai aktual yang ada pada data uji. Rumus dari *Confusion Matrix* dapat dilihat pada Gambar 2, dimana memberikan penilaian komprehensif terhadap model klasifikasi. [19], [20].

		Prediction Values	
		Positive	Negative
Actual Values	Positive	TP (True Positive)	FN (False Negative)
	Negative	FP (False Positive)	TN (True Negative)

Gambar 2. *Confusion matrix*

Pengimplementasian *Confusion Matrix* memudahkan dalam melakukan perhitungan metrik evaluasi kinerja seperti akurasi, presisi dan recall. Nilai akurasi mengukur sejauh mana prediksi model sesuai dengan nilai sebenarnya, presisi mengevaluasi keakuratan model dalam mengklasifikasikan contoh positif, sementara recall mengukur keberhasilan model dalam mengidentifikasi dengan benar semua contoh positif yang seharusnya dikenali [21], [22].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (5)$$

$$Presisi = \frac{TP}{TP+FP} \times 100\% \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (7)$$

C. Hasil dan Pembahasan

1. Pengumpulan Data

Data yang dikumpulkan dalam penelitian ini berisi berbagai atribut kunci seperti NIM, Nama, Jenis Kelamin, Asal Sekolah, Kota, Angkatan, Jumlah SKS, IPS1-4, IPK dan Keterangan untuk status kelulusan mahasiswa. Seluruh data disimpan dalam format Microsoft Excel (.xlsx) seperti pada Gambar 3.

	D	E	F	G	H	I	J	K	L	M	N	O	R	S
	NIM	Nama	Jenis Kelamin	Asal Sekolah	Kota	Angkatan	SKS	IPS1	IPS2	IPS3	IPS4	IPK	Masa Studi	Keterangan
1	UHI899001	MUHAMMAD QIDAM	Laki-laki	SMK KEMALA BHAYANGKARI DELOS	Luar Kota	2018	154	3,95	2,80	2,85	2,85	3,19	4 Tahun, 11 Bulan, 4 Hari	TIDAK TEPAT WAKTU
2	UHI899002	ANDIKA SAPUTRA	Laki-laki	SMAN 5 TANGERANG	Luar Kota	2018	154	2,85	2,57	2,80	2,20	3,01	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
3	UHI899004	ILHAM NUR HIDAYAT	Laki-laki	SMK Islam Al Hikmah	Luar Kota	2018	154	3,00	2,52	3,45	3,25	3,42	4 Tahun, 4 Bulan, 2 Hari	TIDAK TEPAT WAKTU
4	UHI899005	DEDE ISMAIL	Laki-laki	SMKN 3 KUNINGAN	Jakarta	2018	154	2,75	3,62	3,50	3,88	3,6	3 Tahun, 11 Bulan, 12 Hari	TEPAT WAKTU
5	UHI899007	ALFANDI SAFIRA	Laki-laki	SMA ANGASA 2 JAKARTA	Jakarta	2018	154	3,60	3,13	3,67	2,50	3,28	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
6	UHI899014	PRISTA AFRIKAH	Perempuan	SMA 1 TOBOLAN	Jakarta	2018	154	3,70	3,61	3,46	3,92	3,7	3 Tahun, 9 Bulan, 17 Hari	TEPAT WAKTU
7	UHI899018	FAHMI TRYADI FAUZI	Laki-laki	SMA Triguna Utama	Jakarta	2018	154	3,15	2,91	3,10	3,52	3,31	3 Tahun, 11 Bulan, 12 Hari	TEPAT WAKTU
8	UHI899019	NIKKO ESA PRIDANA	Laki-laki	SMK TELEKOMUNIKASI TELESANDI	Luar Kota	2018	154	3,00	3,00	3,25	3,50	3,34	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
9	UHI899026	DINAS RAHMAT RAMADHAN	Laki-laki	SMA Yapika	Luar Kota	2018	154	2,10	2,53	2,67	3,15	3,08	4 Tahun, 11 Bulan, 4 Hari	TIDAK TEPAT WAKTU
10	UHI899039	DWI SETIONO	Laki-laki	PKBM HIKMAH AINUL QORI (HAGI)	Jakarta	2018	154	3,20	2,74	2,90	2,20	3,04	4 Tahun, 4 Bulan, 2 Hari	TIDAK TEPAT WAKTU
11	UHI899053	RIZKY ARDIANSYAH	Laki-laki	SMKN 01 Cibinong	Luar Kota	2018	154	3,60	3,65	3,50	4,00	3,68	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
12	UHI899069	FAHRI AHMAD FACHRUDIN	Laki-laki	SMA NEGERI 22 JAKARTA	Jakarta	2018	154	3,25	3,65	3,67	3,86	3,77	3 Tahun, 11 Bulan, 12 Hari	TEPAT WAKTU
13	UHI899071	HARIZH FAHUDDIN	Laki-laki	SMAS ADI LUNIR	Jakarta	2018	154	2,45	2,56	3,00	3,88	3,31	4 Tahun, 4 Bulan, 2 Hari	TIDAK TEPAT WAKTU
14	UHI899074	BAGAS PRASOJO	Laki-laki	SMK PGRI 20 JAKARTA	Luar Kota	2018	154	2,50	2,81	2,80	2,95	3,24	4 Tahun, 4 Bulan, 2 Hari	TIDAK TEPAT WAKTU
15	UHI899079	MUHAMMAD ADIL HAKIM	Laki-laki	SMAN 9 TANGERANG	Luar Kota	2018	154	3,85	2,87	2,35	3,06	3,13	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
16	UHI899089	MUHAMMAD FITRA	Laki-laki	SMA Terampil 2	Jakarta	2018	154	2,25	3,28	2,54	2,80	3,07	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
17	UHI899090	TRIYANTI	Perempuan	SMA NEGERI 1 LAKSOK	Luar Kota	2018	154	3,35	2,78	2,90	2,81	3,52	3 Tahun, 11 Bulan, 12 Hari	TEPAT WAKTU
18	UHI899097	MARUF DICRY PRAYOGA	Laki-laki	SMK HUTAMA	Luar Kota	2018	154	2,90	2,76	2,95	3,25	3,25	3 Tahun, 11 Bulan, 12 Hari	TEPAT WAKTU
19	UHI899103	GAURI AGUNG SUKMAWATI	Laki-laki	SMA BINA DHARMA	Luar Kota	2018	154	3,30	3,57	3,46	3,50	3,51	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
20	UHI899107	MEYCA ULUL ALBAB	Laki-laki	SMKN 43 JAKARTA	Jakarta	2018	154	2,40	3,22	3,25	3,71	3,49	4 Tahun, 11 Bulan, 4 Hari	TIDAK TEPAT WAKTU
21	UHI899111	FAHRY OCTOBBRA SEPTIANO	Laki-laki	SMAN 04 Jakarta	Luar Kota	2018	154	3,30	3,30	3,58	3,54	3,58	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
22	UHI899118	MUHAMMAD ZAKI FARHAN	Laki-laki	SMAN 43 JAKARTA	Jakarta	2018	154	3,60	2,59	2,75	2,71	3,27	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
23	UHI899126	MUHAMMAD RAHMAN	Laki-laki	SMA PGRI 3 BOGOR	Jakarta	2018	154	2,05	2,67	2,65	3,24	3,09	4 Tahun, 11 Bulan, 4 Hari	TIDAK TEPAT WAKTU
24	UHI899128	MUHAMMAD RAFFIE RIZZANY	Laki-laki	isma islam harsani ibu	Luar Kota	2018	154	2,90	2,19	2,56	3,15	3,13	4 Tahun, 11 Bulan, 4 Hari	TIDAK TEPAT WAKTU
25	UHI899130	ALI ABUL IKRI	Laki-laki	SMK AL BAHRI	Luar Kota	2018	154	2,80	2,52	3,10	3,63	3,3	4 Tahun, 4 Bulan, 2 Hari	TIDAK TEPAT WAKTU
26	UHI899131	MUHAMMAD ARDI ANDREANSYAH	Laki-laki	SMK AL BAHRI	Jakarta	2018	154	2,30	2,89	2,50	3,00	3,14	4 Tahun, 4 Bulan, 2 Hari	TIDAK TEPAT WAKTU
27	UHI899133	RISMAWATI	Perempuan	SMA AL-FALAH HMM	Luar Kota	2018	154	3,40	3,09	3,25	3,25	3,44	3 Tahun, 11 Bulan, 12 Hari	TEPAT WAKTU
28	UHI899139	WILAN NUR ANJANI ATIKAH	Perempuan	SMAN 1 DEPOK	Jakarta	2018	154	3,20	3,30	3,38	3,33	3,48	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
29	UHI899141	AGA AGUSTINA	Perempuan	SMAN 1 Wangi-wangi	Luar Kota	2018	154	3,60	3,30	3,38	3,67	3,52	3 Tahun, 11 Bulan, 12 Hari	TEPAT WAKTU
30	UHI899152	HISYAM SEPTIO MUHAMMAD	Laki-laki	SMK BINA INFORMATIKA	Jakarta	2018	154	2,65	2,62	2,60	3,33	3,11	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
31	UHI899153	WAHYU PRATAMA	Laki-laki	SMA SANTO FRANSISKUS ASISI	Luar Kota	2018	154	2,40	2,72	2,00	2,56	2,99	4 Tahun, 5 Bulan, 17 Hari	TIDAK TEPAT WAKTU
32	UHI899156	MUHAMMAD ALAN ZAINUR	Laki-laki	SMAS ISLAMIC MUGHNI	Luar Kota	2018	154	3,05	3,04	2,83	3,40	3,25	4 Tahun, 4 Bulan, 2 Hari	TIDAK TEPAT WAKTU
33	UHI899175	SUPRIATMAN	Laki-laki	SMK Negeri Sukalayang	Luar Kota	2018	154	2,60	2,62	2,95	3,65	3,25	4 Tahun, 4 Bulan, 2 Hari	TIDAK TEPAT WAKTU
34	UHI899176	YOSIA AJ PRATAMA	Laki-laki	SMK N 52 JAKARTA	Luar Kota	2018	154	3,05	3,65	3,00	3,50	3,42	4 Tahun, 11 Bulan, 4 Hari	TIDAK TEPAT WAKTU

Gambar 3. Data Mahasiswa Teknik Informatika

Informasi mengenai atribut-atribut data yang akan digunakan dapat dilihat dalam Tabel 1.

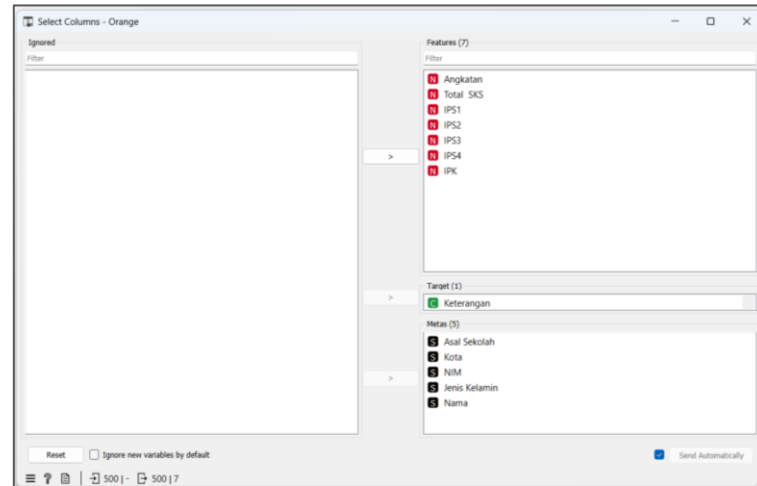
Tabel 1. Detail Atribut

No	Atribut	Type	Deskripsi
1	NIM	Teks	Nomor Induk Mahasiswa
2	Nama	Teks	Nama Lengkap Mahasiswa
3	Jenis Kelamin	Teks	Laki-Laki (L), Perempuan (P)
4	Asal Sekolah	Teks	Sekolah Asal Mahasiswa
5	Kota	Teks	Jakarta, Luar Kota
6	Angkatan	Numeric	Tahun Masuk Mahasiswa
7	Jumlah SKS	Numeric	Total SKS yang diambil
8	IPS 1	Numeric	IP Semester 1
9	IPS 2	Numeric	IP Semester 2
10	IPS 3	Numeric	IP Semester 3
11	IPS 4	Numeric	IP Semester 4
12	IPK	Numeric	Nilai Kumulatif
13	Keterangan	Categorical	Status Kelulusan

2. Preprocessing Data

Berdasarkan Gambar 4 merupakan proses pemilihan data yang dilakukan oleh widget dalam aplikasi Orange Data Mining. Atribut data yang dipilih sebagai *Features* untuk proses ini meliputi Angkatan, Jumlah SKS, IPS 1-4 dan IPK. Keterangan dipilih sebagai atribut target. Selain itu, terdapat atribut tambahan

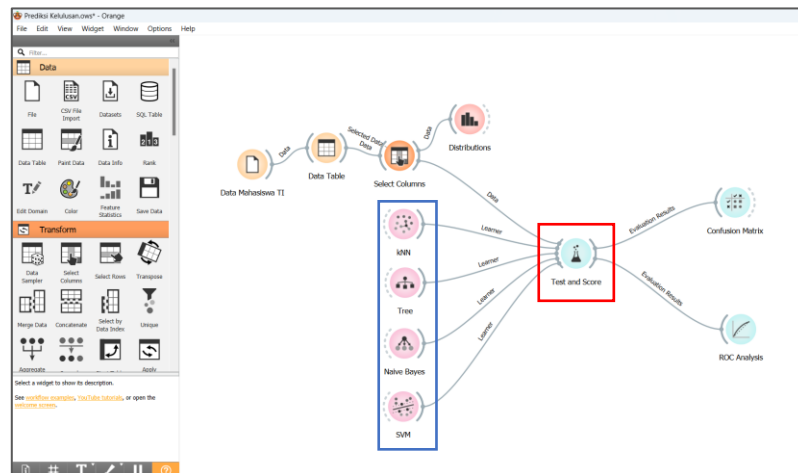
pada bagian *Meta*, yaitu Asal Sekolah, Kota, NIM, Jenis Kelamin, dan Nama, yang tidak digunakan sebagai fitur atau target tetapi disimpan sebagai informasi tambahan. Proses seleksi secara eksklusif hanya mencakup pemilihan titik data yang relevan tanpa perlu menangani nilai yang hilang.



Gambar 4. Pemilihan Atribut dengan *Orange*

3. Perancangan Model

Dalam perancangan model klasifikasi ini, perangkat lunak yang digunakan adalah *Orange Data Mining*. Tujuan utamanya adalah untuk menilai dan membandingkan algoritma yang berbeda untuk mendapatkan akurasi terbaik dalam memprediksi status kelulusan mahasiswa. Adapun rancangan empat model klasifikasi dalam penelitian ini tertera pada Gambar 5.

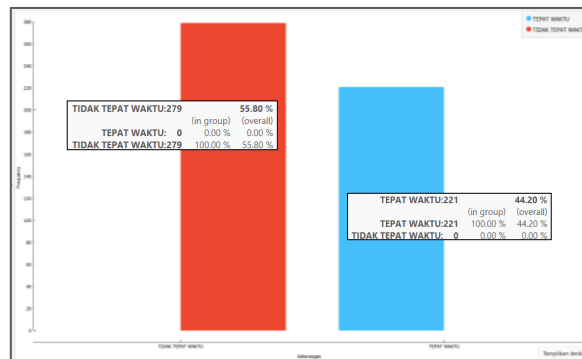


Gambar 5. Perancangan Model Learning

Gambar 5 menggambarkan konfigurasi widget dalam aplikasi *Orange* untuk mengimpor dataset yang telah diproses sebelumnya. Kotak biru mewakili bagian di mana dataset tes dimasukkan ke dalam proses klasifikasi untuk model *Decision Tree*, *Naïve Bayes*, K-NN dan SVM dalam memastikan hasil klasifikasi model kelulusan yang akurat. Kotak merah mewakili widget *test* dan *score*, yang digunakan untuk evaluasi akurasi melalui *Confusion Matrix* dan Analisis ROC.

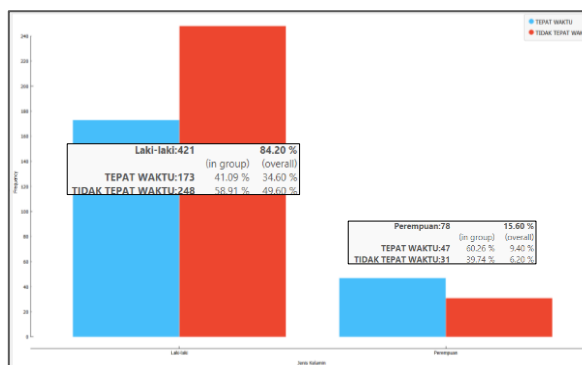
4. Distribusi Frekuensi

Pada Gambar 6, distribusi frekuensi dalam *Orange Data Mining* menggambarkan kelulusan mahasiswa yang menunjukkan bahwa 55,80% mahasiswa diperkirakan akan mengalami keterlambatan, sementara 44,20% diperkirakan akan lulus tepat waktu, ini menunjukkan bahwa sebagian besar mahasiswa mengalami keterlambatan, sehingga membutuhkan waktu lebih dari 4 tahun untuk menyelesaikan studi.



Gambar 6. Distribusi Frekuensi Kelulusan Mahasiswa

Pada pengujian lebih lanjut yang ditunjukkan pada Gambar 7, data menunjukkan bahwa di antara mahasiswa laki-laki, 41,09% berhasil lulus tepat waktu, sementara 58,91% mengalami keterlambatan. Jika dilihat dari keseluruhan populasi mahasiswa, laki-laki yang lulus tepat waktu hanya mencakup 34,60%, sedangkan yang mengalami keterlambatan mencapai 49,60%. Sebaliknya, di antara mahasiswa perempuan, 60,26% berhasil lulus tepat waktu, sedangkan 39,74% mengalami keterlambatan. Dalam keseluruhan populasi, perempuan yang lulus tepat waktu mencakup 9,40%, sementara yang mengalami keterlambatan adalah 6,20% dari total mahasiswa.

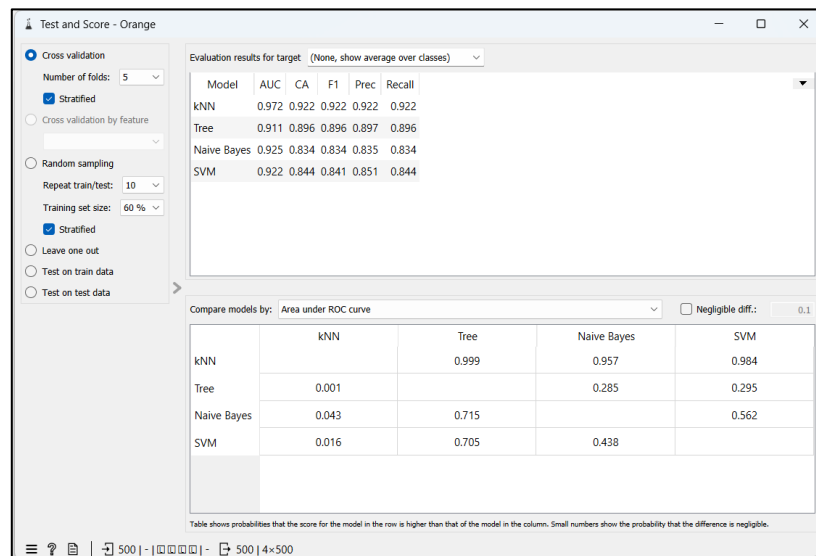


Gambar 7. Distribusi Frekuensi Berdasarkan Jenis Kelamin

5. Hasil Simulasi 4 Model Klasifikasi

Pada Gambar 8 menampilkan hasil dari perhitungan matriks kunci seperti akurasi, presisi, dan recall untuk setiap model algoritma. Algoritma K-NN menunjukkan akurasi tertinggi, mencapai 92%, disertai dengan nilai AUC sebesar 0,972. Nilai AUC (*Area Under the Curve*) berperan sebagai parameter evaluasi yang menilai kinerja pemilihan dengan memperkirakan probabilitas

keluaran untuk sampel yang diambil secara acak. Nilai AUC yang tinggi menandakan kualitas yang unggul dalam hasil klasifikasi yang diperoleh.



Gambar 8. Hasil test dan score

6. Hasil Evaluasi menggunakan *Confusion Matrix*

Pada Gambar 9 menampilkan hasil evaluasi *Confusion Matrix* untuk algoritma *Decision Tree*. Selain itu, Gambar 10 menunjukkan hasil untuk algoritma *Naïve Bayes*, sedangkan Gambar 11 untuk algoritma K-NN, dan Gambar 12 untuk algoritma SVM.

		Predicted		
		TEPAT WAKTU	TIDAK TEPAT WAKTU	Σ
Actual	TEPAT WAKTU	198	23	221
	TIDAK TEPAT WAKTU	29	250	279
Σ		227	273	500

Gambar 9. *Confusion Matrix* model *Decision Tree*

Pada Gambar 9, Algoritma *Decision Tree* menampilkan nilai numerik seperti 198 untuk TP (*True Positive*), 250 untuk TN (*True Negative*), 29 untuk FP (*False Positive*), dan 23 untuk FN (*False Negative*). Berikut adalah hasil perhitungan nilai *Accuracy*, *Precision*, dan *Recall* pada algoritma *Decision Tree*:

$$Accuracy = \frac{198 + 250}{198 + 250 + 29 + 23} \times 100\% = 90\%$$

$$Presisi = \frac{198}{198 + 29} \times 100\% = 87\%$$

$$Recall = \frac{198}{198 + 23} \times 100\% = 90\%$$

		Predicted		Σ
		TEPAT WAKTU	TIDAK TEPAT WAKTU	
Actual	TEPAT WAKTU	183	38	221
	TIDAK TEPAT WAKTU	45	234	279
Σ		228	272	500

Gambar 10. Confusion Matrix model Naïve Bayes

Pada Gambar 10, Algoritma Naïve Bayes menampilkan nilai numerik seperti 183 untuk TP (*True Positive*), 234 untuk TN (*True Negative*), 45 untuk FP (*False Positive*), dan 38 untuk FN (*False Negative*). Berikut adalah hasil perhitungan nilai *Accuracy*, *Precision*, dan *Recall* pada algoritma Naïve Bayes:

$$Accuracy = \frac{183 + 234}{183 + 234 + 45 + 38} \times 100\% = 83\%$$

$$Presisi = \frac{183}{183 + 45} \times 100\% = 80\%$$

$$Recall = \frac{183}{183 + 38} \times 100\% = 83\%$$

		Predicted		Σ
		TEPAT WAKTU	TIDAK TEPAT WAKTU	
Actual	TEPAT WAKTU	199	22	221
	TIDAK TEPAT WAKTU	17	262	279
Σ		216	284	500

Gambar 11. Confusion Matrix model K-NN

Pada Gambar 11, Algoritma K-NN menampilkan nilai numerik seperti 199 untuk TP (*True Positive*), 262 untuk TN (*True Negative*), 17 untuk FP (*False Positive*), dan 22 untuk FN (*False Negative*). Berikut adalah hasil perhitungan nilai *Accuracy*, *Precision*, dan *Recall* pada algoritma K-NN:

$$Accuracy = \frac{199 + 262}{199 + 262 + 17 + 22} \times 100\% = 92\%$$

$$Presisi = \frac{199}{199 + 17} \times 100\% = 92\%$$

$$Recall = \frac{199}{199 + 22} \times 100\% = 90\%$$

		Predicted		Σ
		TEPAT WAKTU	TIDAK TEPAT WAKTU	
Actual	TEPAT WAKTU	161	60	221
	TIDAK TEPAT WAKTU	18	261	279
Σ		179	321	500

Gambar 12. Confusion Matrix model SVM

Pada Gambar 12, Algoritma SVM menampilkan nilai numerik seperti 161 untuk TP (*True Positive*), 261 untuk TN (*True Negative*), 18 untuk FP (*False Positive*), dan 60 untuk FN (*False Negative*). Berikut adalah hasil perhitungan nilai *Accuracy*, *Precision* dan *Recall* pada algoritma SVM:

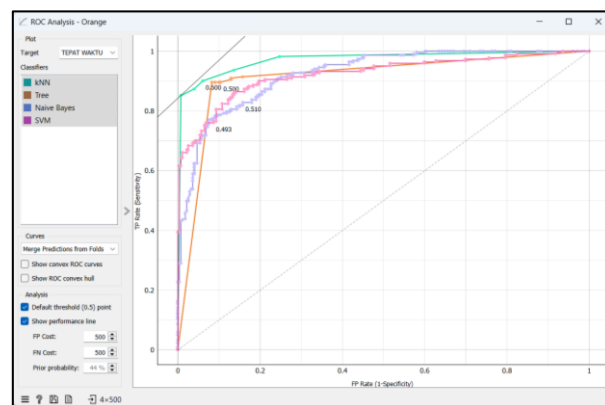
$$Accuracy = \frac{161 + 261}{161 + 261 + 18 + 60} \times 100\% = 84\%$$

$$Presisi = \frac{161}{161 + 18} \times 100\% = 90\%$$

$$Recall = \frac{161}{161 + 60} \times 100\% = 73\%$$

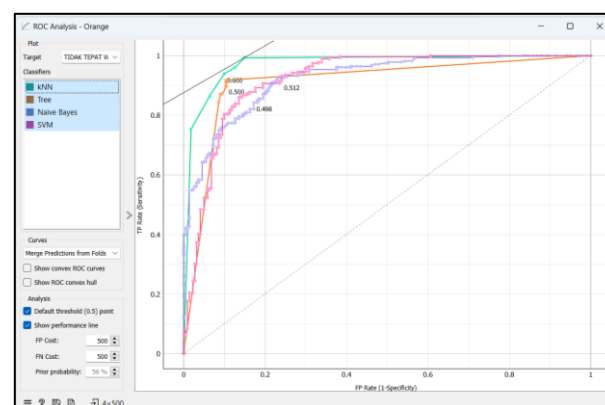
7. Hasil ROC Curve

Pendekatan pragmatis untuk menilai secara visual keakuratan setiap model klasifikasi melibatkan penggunaan kurva ROC yang dihasilkan dari *Confusion Matrix*. Pada Gambar 13 dan 14 di bawah ini memberikan representasi grafis dari hasil ROC keempat algoritma klasifikasi.



Gambar 13. ROC Curve ‘TEPAT WAKTU’

Pada Gambar 13 menampilkan hasil ROC untuk kelulusan mahasiswa Teknik Informatika yang dikategorikan dalam kelompok ‘TEPAT WAKTU’, menunjukkan nilai untuk masing-masing algoritma: K-NN sekitar 0.600, *Decision Tree* sekitar 0.500, *Naïve Bayes* sekitar 0.510 dan SVM sekitar 0.493.



Gambar 14. ROC Curve ‘TIDAK TEPAT WAKTU’

Pada Gambar 14, hasil ROC untuk kelulusan mahasiswa Teknik Informatika yang dikategorikan dalam kelompok 'TIDAK TEPAT WAKTU', menunjukkan nilai untuk masing-masing algoritma: K-NN sekitar 0.600, *Decision Tree* sekitar 0.500, *Naive Bayes* sekitar 0.498 dan SVM sekitar 0.512. Oleh karena itu, dalam penelitian ini K-NN muncul sebagai model algoritma klasifikasi dengan nilai akurasi yang paling baik, seperti yang ditunjukkan oleh kurva analisis ROC yang mendekati titik kiri atas 0,1.

8. Hasil Evaluasi Perbandingan

Di bawah ini adalah Tabel 2, yang menyajikan hasil evaluasi komparatif dari keempat algoritma seperti nilai *Accuracy*, *Precision* dan *Recall*.

Tabel 2. Hasil Evaluasi Perbandingan

Metode	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
Decision Tree	90%	87%	90%
Naive Bayes	83%	80%	83%
K-NN	92%	92%	90%
SVM	84%	90%	73%

Berdasarkan tabel 2, terlihat jelas bahwa hasil akurasi algoritma K-NN mengungguli algoritma lain seperti *Decision Tree*, *Naive Bayes* dan SVM.

D. Simpulan

Dalam penelitian ini menggunakan empat model klasifikasi, yaitu *Decision Tree*, *Naive Bayes*, K-NN dan SVM untuk memprediksi status kelulusan mahasiswa. Penggunaan dataset melibatkan sebanyak 500 data mahasiswa dari prodi Teknik Informatika, UHAMKA, dengan pembagian 60% data latihan dan 40% data uji. Hasil penelitian menunjukkan bahwa 55,80% mahasiswa diperkirakan akan mengalami keterlambatan kelulusan, sementara 44,20% diperkirakan akan lulus tepat waktu.

Dalam analisis Confusion Matrix, algoritma K-NN menonjol sebagai algoritma terbaik dengan tingkat akurasi dan presisi tertinggi, masing-masing sebesar 92% dan recall sebesar 90%. Algoritma *Decision Tree* menempati posisi kedua, dengan akurasi 90%, presisi 87%, dan recall 90%. SVM memperoleh akurasi sebesar 84%, dengan presisi 90% dan recall sebesar 73%. Sementara itu, algoritma *Naive Bayes* menunjukkan akurasi sebesar 83%, dengan presisi 80% dan recall 83%.

Secara keseluruhan, algoritma K-NN terbukti menjadi pilihan yang optimal untuk memprediksi status kelulusan mahasiswa karena unggul di seluruh metrik utama. Algoritma-algoritma tersebut menunjukkan nilai akurasi, presisi, dan recall yang tinggi, menjadikannya sebagai prediktor terbaik dalam penelitian ini jika dibandingkan dengan algoritma lainnya. Keberhasilan klasifikasi pada penelitian ini juga dapat dipengaruhi oleh berbagai faktor, seperti jumlah dataset yang digunakan dan penambahan variabel tambahan selama tahap pengujian.

E. Referensi

- [1] M. R. Qisthiano, T. B. Kurniawan, E. S. Negara, and M. Akbar, "Pengembangan Model Untuk Prediksi Tingkat Kelulusan Mahasiswa Tepat Waktu dengan Metode Naive Bayes," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 987, Jul. 2021, doi: 10.30865/mib.v5i3.3030.

-
- [2] D. Anugrah Putra and M. Kamayani, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naive Bayes di Program Studi Teknik Informatika UHAMKA," *Pros. Semin. Nas. Teknoka*, vol. 5, no., pp. 34–40, Dec. 2020, doi: 10.22236/teknoka.v5i.331.
- [3] Hozairi, 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.
- [4] M. Gunawan, M. Zarlis, and Roslina, "Analisis Komparasi Algoritma Naïve Bayes dan K-Nearest Neighbor Untuk Memprediksi Kelulusan Mahasiswa Tepat Waktu," *J. Media Inform. Budidarma*, vol. 5, no. 2, pp. 513–523, Apr. 2021, doi: 10.30865/mib.v5i2.2925.
- [5] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [6] I. Ramdhani, "PERBANDINGAN METODE DATA MINING MODEL KLASIFIKASI NAIVE BAYES, DECISION TREE DAN K-NEAREST NEIGHBOUR DALAM MEMPREDIKSI KETEPATAN KELULUSAN MAHASISWA PRODI TEKNIK INFORMATIKA DI UNIVERSITAS PAMULANG," *J. Ilmu Komput.*, vol. VI, no. 01, pp. 88–94, 2023, [Online]. Available: <https://jurnal.pranataindonesia.ac.id/index.php/jik/article/view/156>
- [7] A. Budiyantra, I. Irwansyah, E. Prengki, P. A. Pratama, and N. Wiliani, "KOMPARASI ALGORITMA DECISION TREE, NAIVE BAYES DAN K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI MAHASISWA LULUS TEPAT WAKTU," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 265–270, 2020, doi: 10.33480/jitk.v5i2.1214.
- [8] D. Safitri, S. S. Hilabi, and F. Nurapriani, "Analisis Penggunaan Algoritma Klasifikasi dalam Prediksi Kelulusan menggunakan Orange Data Mining," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 1, pp. 75–81, Jan. 2023, doi: 10.36341/rabit.v8i1.3009.
- [9] Shedriko, "PERBANDINGAN ALGORITMA SVM DAN KNN DALAM MENGLASIFIKASI KELULUSAN MAHASISWA PADA SUATU MATA KULIAH," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.)*, vol. 6, no. 2, pp. 115–122, 2021.
- [10] R. Maulana and D. Kumalasari, "ANALISIS KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI STATUS KELULUSAN MAHASISWA AKADEMI BINA SARANA INFORMATIKA," *J. Inform. Kaputama*, vol. 3, no. 2, pp. 79–86, 2019, doi: 10.59697/jik.v3i2.368.
- [11] A. P. Permana, K. Ainayah, and K. F. H. Holle, "Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 6, no. 3, pp. 178–188, 2021, doi: 10.14421/jiska.2021.6.3.178-188.
- [12] E. Novianto, A. Hermawan, and D. Avianto, "KLASIFIKASI ALGORITMA K-NEAREST NEIGHBOR, NAIVE BAYES, DECISION TREE UNTUK PREDIKSI STATUS KELULUSAN MAHASISWA S1," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 8, no. 2, pp. 146–154, Jul. 2023, doi: 10.36341/rabit.v8i2.3434.
- [13] B. Krithiga, P. Sabari, I. Jayasri, and I. Anjali, "Early detection of coronary heart

- disease by using naive bayes algorithm,” in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 12040. doi: 10.1088/1742-6596/1717/1/012040.
- [14] Suyanto, *Data mining untuk klasifikasi dan klasterisasi data : edisi revisi*, Cetakan Pe. Bandung: Informatika Bandung, 2019.
- [15] E. Sutoyo and A. Almaarif, “Educational Data Mining for Predicting Student Graduation Using the Naïve Bayes Classifier Algorithm,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 1, pp. 95–101, Feb. 2020, doi: 10.29207/RESTI.V4I1.1502.
- [16] S. H. Rukmawan, F. R. Aszhari, Z. Rustam, and J. Pandelaki, “Cerebral Infarction Classification Using the K-Nearest Neighbor and Naive Bayes Classifier,” in *Journal of Physics: Conference Series*, IOP Publishing, Feb. 2021, p. 012045. doi: 10.1088/1742-6596/1752/1/012045.
- [17] H. Saadatfar, S. Khosravi, J. H. Joloudari, A. Mosavi, and S. Shamshirband, “A new k-nearest neighbors classifier for big data based on efficient data pruning,” *Mathematics*, vol. 8, no. 2, Feb. 2020, doi: 10.3390/MATH8020286.
- [18] R. Yunita and M. Kamayani, “Indonesian Journal of Computer Science,” *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2023, doi: <https://doi.org/10.33022/ijcs.v12i1.3135>.
- [19] M. Kamil and W. Cholil, “Analisis Perbandingan Algoritma C4.5 dan Naive Bayes pada Lulusan Tepat Waktu Mahasiswa,” *J. Inform.*, vol. 7, no. 2, pp. 97–106, 2020, doi: 10.31294/ji.v7i2.7723.
- [20] S. Alfaris and Kusnawi, “Komparasi Metode KNN dan Naive Bayes Terhadap Analisis Sentimen Pengguna Aplikasi Shopee,” *Indones. J. Comput. Sci.*, vol. 12, no. 2, pp. 284–301, 2023, doi: <https://doi.org/10.33022/ijcs.v12i1.3135>.
- [21] K. L. Kohsasih and Z. Situmorang, “Analisis Perbandingan Algoritma C4.5 dan Naïve Bayes Dalam Memprediksi Penyakit Cerebrovascular,” *J. Inform.*, vol. 9, no. 1, pp. 13–17, 2022, doi: 10.31294/inf.v9i1.11931.
- [22] S. Lonang, A. Yudhana, and M. K. Biddinika, “Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting,” *J. Media Inform. Budidarma*, vol. 7, no. 4, pp. 2109–2117, 2023, doi: 10.30865/mib.v7i4.6553.