Versi Online tersedia di : https://jurnal.buddhidharma.ac.id/index.php/algor/index

JURNAL ALGOR

|2715-0577 (Online)| 2715-0569 (Print)

IMPLEMENTASI ALGORITMA DATA MINING DALAM PREDIKSI KELULUSAN MAHASISWA BERBASIS WEB

Suwitno[1], Benny Daniawan[2], Andri Wijaya[3], Leona Fandini[4], Gesima Chintia Angel Sirait[5]

[1,2,3,4,5]Sistem Informasi, Universitas Buddhi Dharma, Banten, Indonesia

SUBMISSION TRACK

Recieved: Feb 12, 2026
Final Revision: Mar 13, 2026
Available Online: Mar 30, 2026

KEYWORD

Prediksi Kelulusan Mahasiswa, Support Vector Machine (SVM) , Naïve Bayes, Akurasi Klasifikasi

KORESPONDENSI

Phone: 081311190089
E-mail: suwitno@ubd.ac.id

A B S T R A K

Prediksi kelulusan mahasiswa merupakan salah satu upaya strategis dalam mendukung pengambilan keputusan akademik secara lebih dini. Penelitian ini bertujuan membandingkan kinerja algoritma Support Vector Machine (SVM) dan Naive Bayes dalam mengidentifikasi status kelulusan mahasiswa berdasarkan data akademik. Dataset terdiri dari 541 data mahasiswa dengan atribut jenis kelamin, Indeks Prestasi Semester (IPS 1–8), IPK, dan usia. Tahapan penelitian meliputi pembersihan data, penanganan missing value menggunakan nilai rata-rata, transformasi data kategorikal dengan label encoding, normalisasi menggunakan standard scaler, serta pembagian data menjadi 80% data latih dan 20% data uji. Berdasarkan hasil pengujian, algoritma Support Vector Machine memperoleh akurasi 95,41% dengan nilai precision, recall, dan F1-score yang relatif seimbang pada kedua kelas sehingga menunjukkan kemampuan generalisasi yang baik. Sementara itu, Naive Bayes juga mampu melakukan klasifikasi dengan baik dengan akurasi 90,82%, namun menghasilkan tingkat akurasi yang lebih rendah dibandingkan SVM akibat asumsi independensi antar atribut yang kurang sesuai dengan karakteristik data akademik. Hasil penelitian menunjukkan bahwa Support Vector Machine merupakan metode yang lebih efektif dan direkomendasikan untuk prediksi kelulusan mahasiswa dibandingkan Naive Bayes pada dataset yang digunakan.

PENDAHULUAN

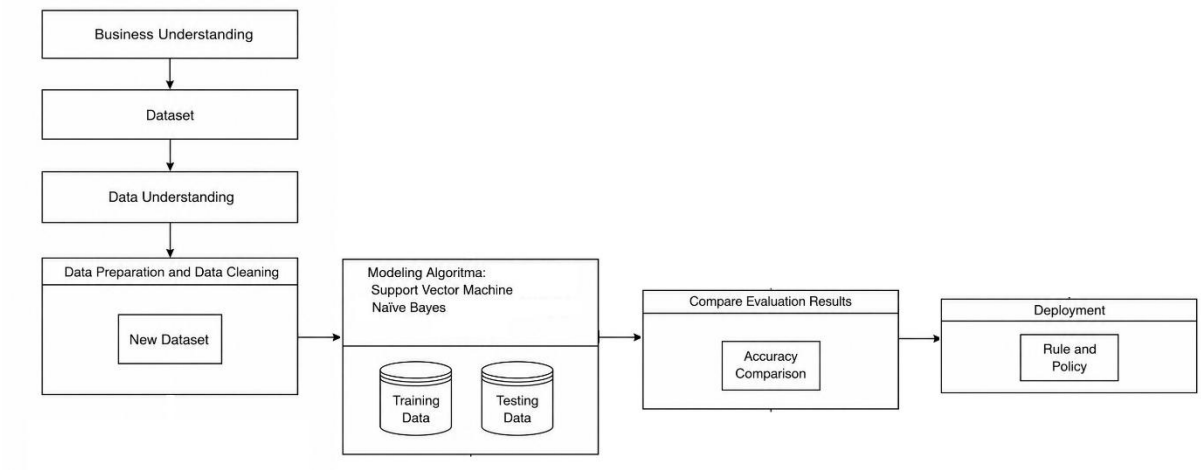
Perkembangan teknologi informasi, kecerdasan buatan (*Artificial Intelligence*), dan analisis data telah mendorong perubahan signifikan dalam berbagai sektor, termasuk dunia pendidikan tinggi. Perguruan tinggi saat ini tidak hanya berperan sebagai penyelenggara proses pembelajaran, tetapi juga sebagai pengelola data akademik yang sangat besar dan terus bertambah setiap tahunnya . Data tersebut mencakup informasi akademik mahasiswa, riwayat

perkuliahan, indeks prestasi, kehadiran, hingga data kelulusan yang apabila dikelola secara optimal dapat menghasilkan informasi strategis untuk mendukung pengambilan keputusan,[1] Pemanfaatan teknologi *data mining* menjadi salah satu pendekatan yang mampu mengekstraksi pola, hubungan, maupun pengetahuan tersembunyi dari kumpulan data tersebut sehingga menghasilkan informasi yang lebih bernilai[2]. Dengan demikian, institusi pendidikan dapat memanfaatkan data historis mahasiswa sebagai dasar dalam menyusun kebijakan akademik yang lebih efektif, meningkatkan kualitas layanan pendidikan, serta melakukan evaluasi terhadap proses pembelajaran secara berkelanjutan.

Salah satu permasalahan yang masih menjadi perhatian utama perguruan tinggi adalah ketepatan waktu kelulusan mahasiswa[3]. Tingkat kelulusan yang rendah atau keterlambatan penyelesaian studi tidak hanya berdampak pada mahasiswa, tetapi juga memengaruhi indikator kinerja institusi, kualitas akademik, serta penilaian akreditasi program studi [4]. Oleh karena itu, diperlukan suatu sistem prediksi yang mampu mengidentifikasi sejak dini mahasiswa yang berpotensi mengalami keterlambatan kelulusan sehingga pihak universitas dapat memberikan intervensi akademik secara tepat sasaran[5]. Penelitian ini mengusulkan komparasi algoritma klasifikasi pada *data mining*, yaitu Support Vector Machine (SVM) dan Naïve Bayes. Melalui perbandingan tingkat akurasi, presisi, *recall*, dan metrik evaluasi lainnya, penelitian ini diharapkan mampu menghasilkan model prediksi yang paling optimal dalam memprediksi kelulusan mahasiswa secara tepat waktu. Hasil penelitian ini diharapkan dapat menjadi salah satu solusi pendukung keputusan bagi perguruan tinggi dalam meningkatkan kualitas layanan akademik, efektivitas monitoring mahasiswa, serta mendukung terciptanya pengelolaan pendidikan yang berbasis data (*data-driven decision making*).

I. METODE

Adapun kerangka pemikiran penelitian ini ditunjukan pada Gambar 1.



Gambar 1: Model Kerangka Pemikiran yang diusulkan

Penelitian ini dimulai dari *Business Understanding* yaitu merupakan tahap yang berfokus pada identifikasi dan perumusan permasalahan yang terjadi dalam proses bisnis secara jelas, sistematis, dan tepat. Pada tahap ini, diperlukan pemahaman yang mendalam terhadap proses bisnis yang sedang berlangsung, termasuk berbagai regulasi dan kebijakan yang menjadi dasar dalam pengelolaan proses tersebut[6]. Data pada penelitian ini dikumpulkan dari Universitas dan dipelajari terkait mahasiswa yang lulus tepat waktu atau tidaknya, dengan tujuan untuk mengetahui faktor penyebab kemungkinan akan tidak lulus tepat waktunya.

Tahap berikutnya *Data Understanding* yaitu merupakan tahapan yang berfokus pada proses identifikasi dan pemahaman terhadap data yang tersedia, kemudian membandingkannya

dengan kebutuhan data yang diperlukan untuk mencapai tujuan penelitian[7]. Tahap ini mencakup analisis struktur, kualitas, dan karakteristik data guna memastikan bahwa data yang digunakan relevan, lengkap, dan layak untuk diproses pada tahapan selanjutnya. Pada tahap ini data yang di dapat sebanyak 564 mahasiswa. Atribut yang digunakan sebanyak 17 yang ditunjukan pada Tabel 1

Tabel 1. Atribut Data Penelitian

No	Atribut	Tipe Data	Keterangan
1	Jurusan	Polinomial	Jurusan Mahasiswa
2	Nim	Numerik	Nomor Induk Mahasiswa
3	Jenis Kelamin	Binomial	Laki-laki atau Perempuan
4	Tanggal Lahir	Numerik	Tanggal Lahir Mahasiswa
5	Status sipil	Binomial	Menikah atau Belum
6	Mahasiswa Kerja	Binomial	Status Bekerja Ya atau Tidak
7	IPS 1	Numerik	Indeks Prestasi Semester 1
8	IPS 2	Numerik	Indeks Prestasi Semester 2
9	IPS 3	Numerik	Indeks Prestasi Semester 3
10	IPS 4	Numerik	Indeks Prestasi Semester 4
11	IPS 5	Numerik	Indeks Prestasi Semester 5
12	IPS 6	Numerik	Indeks Prestasi Semester 6
13	IPS 7	Numerik	Indeks Prestasi Semester 7
14	IPS 8	Numerik	Indeks Prestasi Semester 8
15	IPK	Numerik	Indeks Prestasi Kumulatif
15	Semester Lulus	Numerik	Semester Mahasiswa Lulus
16	Total SKS Lulus	Numerik	Jumlah SKS Mahasiswa saat lulus
17	Status Kelulusan	Binomial	Terlambat atau Tepat (Label)

Tahap berikutnya *Data Preparation* merupakan fase yang dimulai dengan memilih, membersihkan, dan mengintegrasikannya ke dalam bentuk yang berkualitas dan berguna [8].
Data yang baik dan akurat berawal dari data preparation. Beberapa hal yang dilakukan meliputi pemeriksaan kembali untuk memastikan keakuratan data, penanganan data *outlier*, pembersihan data yang hilang (*missing values*), serta perbaikan data yang tidak konsisten [9] . Proses tersebut bertujuan untuk menghasilkan data yang berkualitas sehingga siap digunakan pada tahapan analisis berikutnya. Pada dataset yang digunakan, proses validasi data dilakukan dengan menghapus data yang tidak lengkap, seperti data mahasiswa yang hanya terdaftar pada satu semester tetapi tidak melanjutkan perkuliahan pada semester berikutnya. Penghapusan data tersebut bertujuan untuk menjaga kualitas data sehingga hasil pengolahan dan analisis menjadi lebih akurat. Sementara itu, mahasiswa yang telah mencapai semester 8 tetapi tidak memiliki nilai indeks prestasi semester tetap dipertahankan dalam dataset. Pada kondisi ini, nilai ips yang tidak tersedia (*missing value*) diisi dengan nilai nol karena mahasiswa masih berstatus aktif pada saat itu, namun tidak berhasil menyelesaikan kegiatan akademiknya sehingga tidak memperoleh nilai ips.
Tahap *Modelling* dilaksanakan setelah data melalui proses pembersihan sehingga dipastikan tidak mengandung nilai yang hilang (*missing values*) maupun data yang tidak konsisten. Kondisi ini penting untuk memastikan bahwa model yang dibangun dapat menghasilkan

analisis dan prediksi yang lebih akurat serta dapat diandalkan. Pada tahap modelling ini akan dilakukan klasifikasi menggunakan algoritma Support Vector Machine dan Naïve Bayes dengan pembagian data training dan data testing adalah 80 : 20. Pembagian tersebut bertujuan agar model memperoleh data yang cukup untuk mempelajari pola selama proses pelatihan, sementara data testing digunakan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data yang belum pernah digunakan sebelumnya. Rasio 80:20 dipilih karena memberikan keseimbangan antara proses pembelajaran dan evaluasi model serta merupakan salah satu pendekatan yang umum digunakan dalam penelitian data mining [10], [11].
Tahap *Evaluation* merupakan tahapan untuk memvalidasi performa dan akurasi model yang telah dikembangkan. Pada tahap ini parameter pengujian bertindak sebagai indikator valid standar untuk menentukan layak atau tidaknya suatu model diimplementasikan[12].
Tahap *Deployment* bertujuan untuk menyajikan hasil keseluruhan proses data mining dalam bentuk laporan yang memuat pengetahuan dan pola yang telah diperoleh. Selain itu, model yang dihasilkan juga divisualisasikan agar informasi yang diperoleh lebih mudah dipahami, diinterpretasikan, dan diterapkan sesuai dengan kebutuhan pengguna.

Support Vector Machine (SVM)
Support Vector Machine (SVM) merupakan salah satu algoritma dalam pembelajaran mesin (*machine learning*) yang banyak dimanfaatkan untuk menyelesaikan permasalahan klasifikasi maupun regresi[13] .

Naïve Bayes (NB)
NB adalah algoritma klasifikasi yang menggunakan teorema Bayes untuk prediksi[14]. Kelebihan metode NB meliputi kemudahan implementasi, kinerja yang baik dalam klasifikasi data besar, dan kemampuan menangani fitur-fitur yang tidak relevan.

II. HASIL

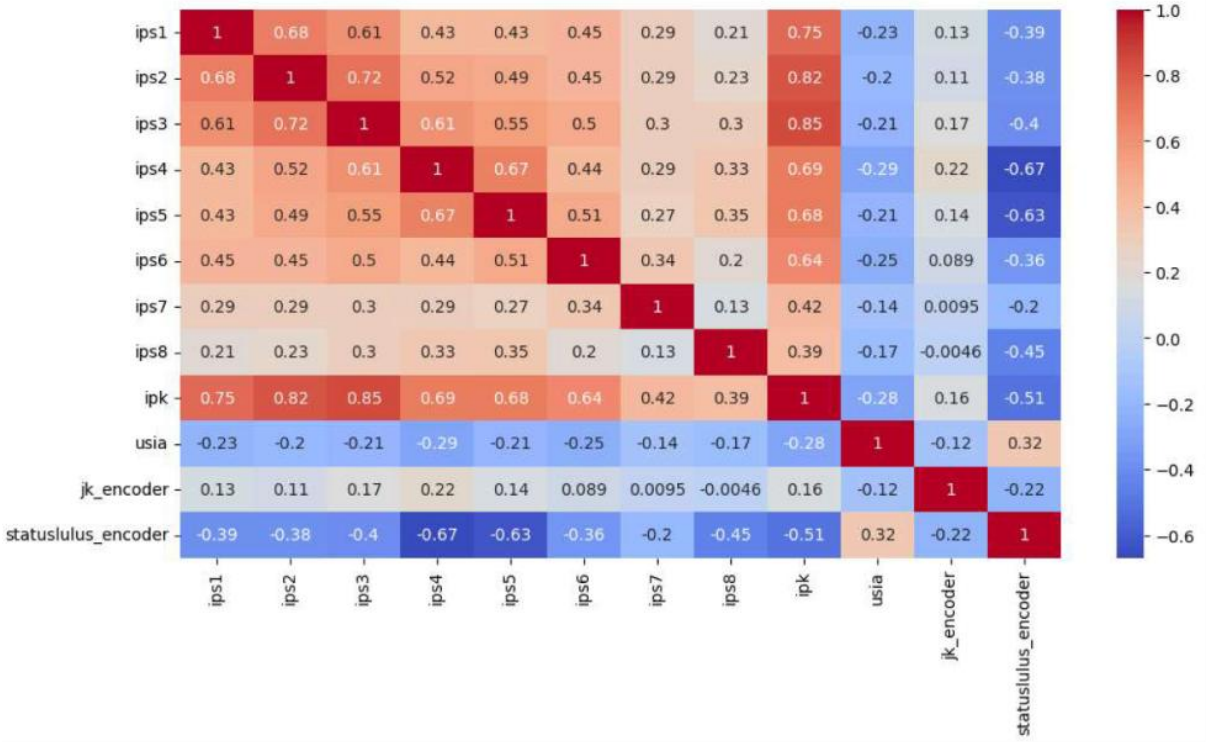
Penelitian ini menghasilkan model klasifikasi kelulusan mahasiswa menggunakan algoritma klasifikasi *Support Vector Machine* (SVM) dan Naive Bayes. Pengolahan data akan menggunakan Jupiter Notebook 6.5.12 dan Python 3.9.10. Data awal ditunjukan pada Tabel 2.

Tabel 2. data mahasiswa

No	Jrs	Nim	JK	Mhs Krj	Tgl Lahir	Sts Sipil	Ips 1	…	Ips 14	Sms Lulus	Ipk	Tsl
1	Ak	20150100205	P	-	02/11/1997	BM	0.33	…	NaN	13	2,94	148
2	Ak	20150100010	L	-	05/08/1994	BM	0.31	…	NaN	13	3.07	148
3	Ak	20160100096	L	-	15/06/1998	BM	3.39	…	NaN	11	3.24	147
4	Ak	20160101096	P	-	24/11/1998	BM	3.01	…	NaN	11	3.19	146
5	Ak	20160100137	L	-	08/09/1998	BM	3.09	…	NaN	13	3.09	146
…	…	…	..	-	…	…	…	…	…	…	….	….
561	TI	20181000071	L	-	08/12/1999	BM	3.23		NaN	8	3.77	144
562	TI	20181000074	L	-	18/01/2000	BM	3.33		NaN	8	3.56	144
563	TI	20181000080	P	-	19/03/2000	BM	3.20		NaN	8	3.40	144
564	TI	20181000081	L	-	21/05/2000	BM	3.12		NaN	8	3.27	144

Pada tabel diatas Jrs merupakan jurusan yang dipilih oleh mahasiswa. Nim adalah nomor induk mahasiswa. Jk adalah jenis kelamin dimana P adalaj Perempuan dan L adalah Laki-laki. Mhs Krj adalah status mahasiswa bekerja. Status sipil BM adalah belum menikah. Ips 1 adalah

indeks prestasi smster satu. Ips 14 adalah indeks prestasi semester keempat belas. Ipk adalah indeks prestasi kumulatif dan Tsl adalah total sks lulus.
Atribut seperti Nim dihapus, status tanggal lahir diubah menjadi usia pada saat tahun kelulusan, indeks prestasi semester kesembilan sampai keempat belas dihapus dan diganti dengan label lulus terlambat atau tepat. Status sipil dan status bekerja juga dihapus di karenakan pada data ini didapatkan bahwa semua mahasiswa belum bekerja dan belum menikah. Total data yang semula ada 564 menjadi 541 dikarenakan ada beberapa data yang redudant. Data jenis kelamin dan label status kelulusan yang semula berupa text diubah menjadi numerik menggunakan label encoder karena metode svm tidak mampu memproses teks biasa dalam bentuk data mentah, sehingga sangat penting untuk mengkodekan data mentah menjadi vektor numerik[15]. Kemudian data diolah dan ditampilkan dalam bentuk matriks korelasi seperti pada Gambar 2.



Gambar 2: Hasil Matriks Korelasi

Hasil matriks korelasi menunjukkan bahwa mendekati -1 artinya hubungan negatif yang kuat, jika mendekati 0 artinya tidak memiliki hubungan yang linear, dan jika mendekati + artinya memiliki hubungan yang kuat dan signifikan[16], [17]. Pada matriks diatas terlihat bahwa korelasi antar atribut paling tinggi berada di indek prestasi semester ketiga terhadap IPK sebesar 0,85 dan indeks prestasi semester kedua sebesar 0,82 terhadap IPK, yang artinya bahwa performa mahasiswa di semester awal berpengaruh terhadap hasil indeks prestasi kumulatif mahasiswa. Korelasi antar indeks prestasi semester yang tertinggi terjadi pada indeks ptestasi semester 2 dan 3 sebesar 0,72 yang artinya performa mahasiswa di semester ke 2 sangat berpengaruh pada performa di semester ke 3. Sedangkan atribut usia menunjukkan korelasi negatif terhadap hampir seluruh variabel akademik, yang artinya pengaruh usia terhadap prestasi akademik dapat dikatakan relatif kecil. Hasil beberapa korelasi atribut paling berpengaruh (kuat) ditunjukkan pada tabel 3.

Tabel 3. korelasi antar atribut ips

Faktor	Nilai Korelasi	Interprestasi
IPS2 – IPS3	0.72 (korelasi kuat)	Semaikin besar nilai ips2 maka mempengaruhi nilai ips3
IPS4 – IPS5	0.67 (korelasi kuat)	Semakin besar nilai ips4 maka mempengaruhi nilai ips5
IPS1 – IPS2	0.68 (korelasi kuat)	Semakin besar ips1 maka akan mempengaruhi nilai ips2
IPS3 – IPS4	0.61 (korelasi kuat)	Semakin besar ip3 maka akan mempengaruhi nilai ips4
IPS1 – IPS3	0.61 (korelasi kuat)	Hasil ips2 juga mempengaruhi hasil di ips3 nanti

Beberapa atribut ips juga sangat mempengaruhi hasil ipk juga, seperti ditunjukan pada tabel 4.

Tabel 4. korelasi ips terhadap ipk

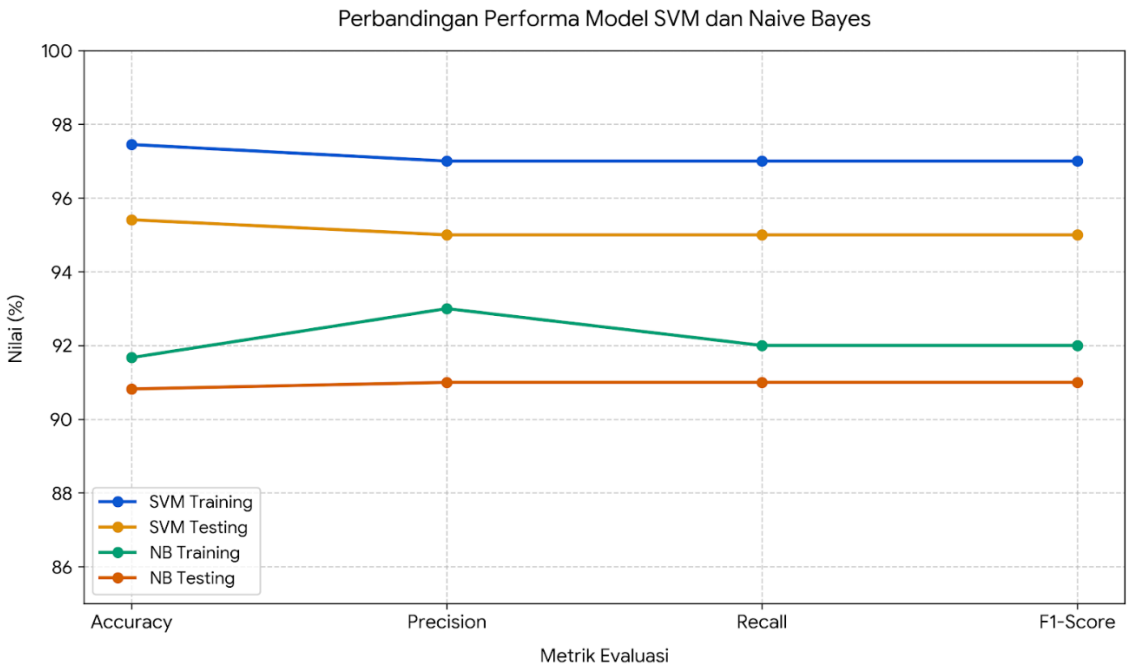
Atribut	Nilai Korelasi
IPS3	0.85
IPS2	0.82
IPS1	0.75
IPS4	0.69
IPS5	0.68
IPS6	0.64
IPS7	0.42
IPS8	0/38

Dapat dilihat bahwa IPS3 memiliki korelasi tertinggi terhadap IPK sebesar 0.85 yang termasuk kategori sangat kuat. Hal ini menunjukkan bahwa pencapaian akademik mahasiswa pada semester ketiga memiliki kontribusi yang sangat besar terhadap pembentukan IPK akhir. Kondisi tersebut cukup logis karena IPK merupakan akumulasi dari seluruh IPS, sehingga semakin tinggi nilai IPS pada setiap semester maka semakin tinggi pula IPK mahasiswa. Menariknya, korelasi mulai menurun pada semester tujuh dan delapan. Hal ini kemungkinan disebabkan karena pada semester akhir mahasiswa lebih banyak mengambil mata kuliah skripsi, magang, atau tugas akhir sehingga variasi nilai akademik menjadi lebih kecil dibandingkan semester awal.

Langkah berikutnya adalah data diolah dengan menggunakan metode klasifikasi *Support Vector Machine* (SVM) dan Naive Bayes. Hasil pengolahan dari metode tersebut disajikan pada tabel 5.

Tabel 5. hasil pengolahan klasifikasi

Algoritma	Data Set	Accuracy	Precision	Recall	F1=Score
Support Vector Machine (SVM)	Training	97.45	97	97	97
	Testing	95.41	95	95	95
Naive Bayes (NB)	Training	91.67	93	92	92
	Testing	90.82	91	91	91



Gambar 3: Perbandingan performa Model SVM dan Naive/ Bayes

Berdasarkan Tabel 5 dan Gambar 3 diatas dapat disimpulkan bahwa algoritma *Support Vector Machine* (SVM) memiliki performa yang jauh lebih unggul dan konsisten dibandingkan dengan algoritma Naive Bayes (NB) dalam mengklasifikasikan data. Keunggulan ini dibuktikan oleh seluruh metrik evaluasi yang diukur, yang meliputi tingkat akurasi, presisi, recall, hingga nilai F1-score pada kedua skenario pengujian. Pada data pengujian (testing), algoritma SVM berhasil mencapai nilai akurasi yang sangat tinggi sebesar 95.41%, diiringi dengan nilai presisi, recall, dan F1-score yang stabil di angka 95%. Di sisi lain, algoritma Naive Bayes pada data pengujian hanya mampu menghasilkan tingkat akurasi sebesar 90.82%, serta nilai presisi, recall, dan F1-score yang tertahan di angka 91%. Selain menunjukkan performa akurasi yang superior, kedua model ini juga terbukti memiliki stabilitas dan kemampuan generalisasi yang sangat baik terhadap data baru. Hal ini tecermin dari selisih performa yang sangat tipis dan minim antara data pelatihan (training) dan data pengujian (testing). Pada model SVM, penurunan akurasi hanya berkurang sebesar 2.04%, sedangkan pada model Naive Bayes penurunannya bahkan berada di bawah 1%, yaitu hanya sebesar 0.85%. Oleh karena itu, dengan mempertimbangkan aspek konsistensi, keandalan, dan tingginya nilai keseluruhan metrik performa pada data uji, algoritma Support Vector Machine merupakan pilihan model yang paling optimal, efektif, serta sangat direkomendasikan untuk digunakan dalam menangani kasus klasifikasi pada dataset ini.

III. PEMBAHASAN

Berdasarkan hasil pengujian, algoritma *Support Vector Machine* (SVM) menunjukkan performa yang sangat baik pada proses klasifikasi. Nilai *accuracy* yang diperoleh menunjukkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar oleh model. Tingginya nilai *precision* mengindikasikan bahwa prediksi yang dihasilkan oleh SVM memiliki tingkat kesalahan yang relatif rendah, sehingga sebagian besar data yang diprediksi sebagai suatu kelas benar-benar merupakan anggota kelas tersebut. Selain itu, nilai *recall* yang tinggi menunjukkan bahwa model mampu mengenali sebagian besar data aktual pada masing-masing kelas sehingga kemungkinan terjadinya kesalahan berupa false negative dapat diminimalkan. Kombinasi antara *precision* dan *recall* tersebut menghasilkan nilai F1-score yang tinggi, yang

menunjukkan bahwa model memiliki keseimbangan yang baik dalam melakukan proses klasifikasi. Performa SVM yang relatif tinggi tidak terlepas dari karakteristik algoritma tersebut yang bekerja dengan membangun hyperplane optimal untuk memisahkan setiap kelas pada ruang berdimensi tinggi. SVM memiliki kemampuan untuk memaksimalkan margin antar kelas sehingga batas keputusan (*decision boundary*) yang dihasilkan menjadi lebih optimal dibandingkan metode klasifikasi sederhana. Selain itu, penggunaan kernel pada SVM memungkinkan model memetakan data yang tidak dapat dipisahkan secara linear ke ruang berdimensi lebih tinggi sehingga pola-pola kompleks dalam data dapat dikenali dengan lebih baik. Karakteristik tersebut menjadikan SVM sangat efektif ketika digunakan pada data akademik yang memiliki hubungan antar atribut yang cukup kompleks.

SVM mampu menghasilkan tingkat generalisasi yang baik terhadap data pengujian. Hal ini terlihat dari tidak adanya perbedaan yang terlalu signifikan antara hasil evaluasi pada data pelatihan dan data pengujian. Kondisi tersebut menunjukkan bahwa model tidak mengalami overfitting secara berlebihan sehingga masih mampu mempertahankan kemampuan prediksinya ketika dihadapkan pada data yang belum pernah dipelajari sebelumnya. Dengan demikian, model SVM dapat dikatakan memiliki kemampuan prediksi yang stabil dan layak digunakan sebagai model klasifikasi pada kasus prediksi kelulusan mahasiswa.

Di sisi lain, algoritma Naïve Bayes juga mampu memberikan hasil klasifikasi yang cukup baik meskipun secara keseluruhan performanya masih berada di bawah algoritma SVM. Berdasarkan hasil evaluasi, nilai *accuracy, precision, recall*, maupun F1-score yang diperoleh Naïve Bayes menunjukkan bahwa algoritma ini tetap mampu mengidentifikasi pola data dengan tingkat ketepatan yang cukup tinggi. Hal tersebut menunjukkan bahwa pendekatan probabilistik yang digunakan Naïve Bayes masih relevan untuk diterapkan pada permasalahan klasifikasi kelulusan mahasiswa. Kinerja Naïve Bayes dipengaruhi oleh prinsip dasar *Teorema Bayes* yang mengasumsikan bahwa setiap atribut bersifat independen terhadap atribut lainnya. Dalam praktiknya, asumsi independensi tersebut sering kali sulit dipenuhi pada data akademik. Variabel seperti Indeks Prestasi Kumulatif (IPK), jumlah SKS yang telah ditempuh, jumlah mata kuliah yang diulang, maupun lama masa studi umumnya memiliki hubungan yang saling memengaruhi. Ketika hubungan antar atribut cukup kuat, asumsi independensi pada Naïve Bayes menjadi kurang sesuai sehingga probabilitas posterior yang dihasilkan tidak selalu merepresentasikan kondisi sebenarnya. Akibatnya, kemampuan model dalam membedakan kelas menjadi sedikit menurun dibandingkan algoritma yang tidak bergantung pada asumsi tersebut.

Meskipun demikian, Naïve Bayes memiliki beberapa keunggulan yang patut diperhatikan. Algoritma ini memiliki proses pelatihan yang relatif cepat, kebutuhan komputasi yang rendah, serta implementasi yang sederhana. Selain itu, Naïve Bayes mampu bekerja secara efisien pada dataset berukuran besar dan tetap memberikan hasil klasifikasi yang kompetitif meskipun jumlah data pelatihan tidak terlalu banyak. Oleh karena itu, algoritma ini masih menjadi salah satu alternatif yang layak digunakan ketika efisiensi komputasi menjadi pertimbangan utama. Jika dibandingkan secara langsung, hasil penelitian menunjukkan bahwa *Support Vector Machine* memiliki performa yang lebih unggul dibandingkan Naïve Bayes pada seluruh metrik evaluasi yang digunakan. Perbedaan tersebut menunjukkan bahwa kemampuan SVM dalam membentuk batas pemisah antar kelas lebih efektif dibandingkan pendekatan probabilistik yang diterapkan oleh Naïve Bayes. Selain itu, SVM mampu mengakomodasi hubungan yang kompleks antar atribut sehingga menghasilkan tingkat klasifikasi yang lebih akurat. Sebaliknya, performa Naïve Bayes cenderung dipengaruhi oleh asumsi independensi antar atribut sehingga ketika asumsi tersebut tidak terpenuhi, kemampuan klasifikasinya mengalami penurunan. Hasil tersebut menunjukkan bahwa karakteristik dataset sangat berpengaruh terhadap performa algoritma klasifikasi. Dataset prediksi kelulusan mahasiswa mengandung berbagai atribut akademik yang saling berkorelasi sehingga algoritma yang

mampu memodelkan hubungan antar fitur secara lebih fleksibel akan menghasilkan performa yang lebih baik. Kondisi ini menjelaskan mengapa SVM memperoleh nilai accuracy, precision, recall, dan F1-score yang lebih tinggi dibandingkan Naïve Bayes. Secara keseluruhan, hasil penelitian menunjukkan bahwa kedua algoritma mampu melakukan klasifikasi terhadap data kelulusan mahasiswa dengan baik. Namun demikian, berdasarkan seluruh metrik evaluasi yang diperoleh, *Support Vector Machine* merupakan metode yang memberikan performa terbaik dalam penelitian ini. Dengan tingkat akurasi yang lebih tinggi serta keseimbangan *precision, recall,* dan F1-score yang lebih baik, SVM dinilai lebih efektif dalam mengidentifikasi mahasiswa yang berpotensi lulus tepat waktu maupun tidak tepat waktu. Sementara itu, Naïve Bayes tetap memberikan performa yang cukup kompetitif dan memiliki keunggulan pada aspek efisiensi komputasi, sehingga masih dapat dipertimbangkan sebagai metode alternatif pada kondisi tertentu. Temuan ini mengindikasikan bahwa pemilihan algoritma klasifikasi hendaknya disesuaikan dengan karakteristik data yang digunakan agar model prediksi yang dihasilkan memiliki tingkat akurasi dan reliabilitas yang optimal.

IV. KESIMPULAN

Berdasarkan hasil implementasi dan evaluasi model klasifikasi yang telah dilakukan, dapat disimpulkan bahwa algoritma *Support Vector Machine* (SVM) dan *Naïve Bayes* (NB) mampu digunakan untuk melakukan prediksi ketepatan waktu kelulusan mahasiswa berdasarkan data akademik yang tersedia. Kedua algoritma berhasil membangun model klasifikasi yang mampu mengidentifikasi pola hubungan antara atribut-atribut akademik dengan status kelulusan mahasiswa, sehingga dapat dimanfaatkan sebagai dasar dalam mendukung proses pengambilan keputusan di lingkungan perguruan tinggi. Hasil pengujian menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memberikan performa klasifikasi yang lebih baik dibandingkan Naïve Bayes, yang ditunjukkan oleh nilai accuracy, precision, recall, dan F1-score yang lebih tinggi pada proses evaluasi model. Kinerja tersebut mengindikasikan bahwa SVM memiliki kemampuan yang lebih optimal dalam membentuk batas pemisah antar kelas sehingga mampu menghasilkan prediksi yang lebih akurat dan memiliki tingkat generalisasi yang baik terhadap data pengujian. Keunggulan tersebut menunjukkan bahwa SVM lebih efektif dalam menangani karakteristik data akademik yang memiliki hubungan antaratribut yang kompleks.

Sementara itu, algoritma Naïve Bayes juga menghasilkan performa klasifikasi yang cukup baik meskipun masih berada di bawah SVM. Pendekatan probabilistik yang diterapkan oleh Naïve Bayes mampu menghasilkan prediksi yang kompetitif dengan proses komputasi yang relatif sederhana dan efisien. Namun, asumsi independensi antaratribut yang menjadi dasar metode ini menyebabkan kemampuan klasifikasinya cenderung menurun ketika diterapkan pada data akademik yang memiliki keterkaitan antarvariabel.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa *Support Vector Machine* merupakan metode yang paling optimal untuk prediksi kelulusan mahasiswa pada dataset yang digunakan dalam penelitian ini, sedangkan Naïve Bayes tetap menjadi alternatif yang layak apabila efisiensi komputasi dan kemudahan implementasi menjadi pertimbangan utama. Temuan ini diharapkan dapat menjadi referensi dalam pengembangan sistem prediksi kelulusan mahasiswa berbasis data mining serta mendukung pengambilan keputusan akademik yang lebih cepat, objektif, dan berbasis data.

REFERENCES

[1] A. A. Zulfa, T. Ibrahim, and O. Arifudin, "PERAN SISTEM INFORMASI AKADEMIK BERBASIS WEB DALAM UPAYA MENINGKATKAN EFEKTIVITAS DAN EFISIENSI PENGELOLAAN AKADEMIK DI PERGURUAN TINGGI," *Jurnal Tahsinia*, vol. 6, no. 1, pp. 115–134, Jan. 2025, doi: 10.57171/JT.V6I1.615.

[2] A. Ardhi Baskara, N. Maharani Piranti, M. Fahrury Romdendine Manajemen Teknologi Keimigrasian, P. Imigrasi Jalan Raya Gandul No, and K. Depok, "FRAMEWORK DATA MINING: SEBUAH SURVEI," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 3, pp. 4886–4895, May 2025, doi: 10.36040/JATI.V9I3.13803.

[3] I. Kamil, I. A. Irawan, A. Ratnasari, E. Wahyuningsih, and M. Ariani, "Implementasi Strategi Lulus Tepat Waktu Sarjana Strata Satu (S1) dalam Era Digital," *Jurnal Pengabdian Kepada Masyarakat*, vol. 5, no. 01, pp. 6–11, Mar. 2024, doi: 10.54147/JPKM.V5I01.1029.

[4] nurul khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, "PREDIKSI KELULUSAN MAHASISWA DENGAN METODE NAIVE BAYES," *Technologia : Jurnal Ilmiah*, vol. 13, no. 3, pp. 207–214, Aug. 2022, doi: 10.31602/TJI.V13I3.7312.

[5] Y. E. Yuspita, R. Okra, and M. Rezeki, "PENERAPAN ALGORITMA KLASIFIKASI UNTUK PREDIKSI TINGKAT KELULUSAN MAHASISWA MENGGUNAKAN RAPPIDMINER," *Djtechno: Jurnal Teknologi Informasi*, vol. 6, no. 1, pp. 376–388, Apr. 2025, doi: 10.46576/djtechno.

[6] A. M. Shimaoka, R. C. Ferreira, and A. Goldman, "The evolution of CRISP-DM for Data Science: Methods, Processes and Frameworks," *SBC Computing Reviews*, vol. 4, no. 1, pp. 28–43, Oct. 2024, doi: 10.5753/REVIEWS.2024.3757.

[7] G. Salazar-Salazar, M. Mora, H. A. Duran-Limon, and F. J. Á. Rodríguez, "A Selective Comparative Review of CRISP-DM and TDSP Development Methodologies for Big Data Analytics Systems," pp. 161–185, 2024, doi: 10.1007/978-3-031-40956-1_6.

[8] M. Gheisari *et al.*, "Data Mining Techniques for Web Mining: A Survey," *Artificial Intelligence and Applications*, vol. 1, no. 1, pp. 3–10, Jan. 2023, doi: 10.47852/BONVIEWAIA2202290.

[9] J. Elektronika and D. Komputer, "Mengoptimalkan Proses Pembersihan Data dalam Analisis Big Data Menggunakan Pipeline Berbasis AI," *Elkom: Jurnal Elektronika dan Komputer*, vol. 17, no. 2, pp. 657–666, Dec. 2024, doi: 10.51903/ELKOM.V17I2.2311.

[10] V. R. Joseph, "Optimal ratio for data splitting," *Stat. Anal. Data Min.*, vol. 15, no. 4, pp. 531–538, Aug. 2022, doi: 10.1002/SAM.11583;SUBPAGE:STRING:FULL.

[11] H. Bichri, A. Chergui, and M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 2, pp. 331–339, Feb. 2024, doi: 10.14569/IJACSA.2024.0150235.

[12] N. Lundén, E. T. Bekar, A. Skoogh, and J. Bokrantz, "Domain Knowledge in CRISP-DM: An Application Case in Manufacturing," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 7603–7608, Jan. 2023, doi: 10.1016/J.IFACOL.2023.10.1156.

[13] A. Roy and S. Chakraborty, "Support vector machine in structural reliability analysis: A review," *Reliab. Eng. Syst. Saf.*, vol. 233, p. 109126, May 2023, doi: 10.1016/J.RESS.2023.109126.

[14] A. Çınar, "Multi-Class Classification with the Gaussian Naive Bayes Algorithm," *Journal of Data Applications*, vol. 0, no. 2, pp. 1–13, Jun. 2024, doi: 10.26650/JODA.1389471.
[15] R. R. Fiska, "Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera)," *Sains dan Teknologi Informasi*, vol. 3, no. 1, pp. 15–23, Jun. 2017, doi: 10.33372/STN.V3I1.200.
[16] M. Buyukkececi and M. C. Okur, "A Comprehensive Review of Feature Selection and Feature Selection Stability in Machine Learning," *Gazi University Journal of Science*, vol. 36, no. 4, pp. 1506–1520, Dec. 2023, doi: 10.35378/GUJS.993763.
[17] "Towards Finding a Minimal Set of Features for Predicting Students' Performance Using Educational Data Mining." Accessed: Jul. 10, 2026. [Online]. Available: https://www.mecs-press.org/ijmecs/ijmecs-v15-n3/v15n3-4.html?utm_source=chatgpt.com

BIOGRAPHY

Suwitno, Saat ini sebagai dosen tetap pada Program Studi Sistem Informasi di Universitas Buddhi Dharma**.**

Benny Daniawan, Dosen tetap Program Studi Sistem Infirmasi di Fakultas Sains dan Teknologi Universitas Buddhi Dharma.

Andri Wijaya, Dosen tetap Program Studi Sistem Infirmasi di Fakultas Sains dan Teknologi Universitas Buddhi Dharma.

Leona Fandini, Mahasiswa Program Studi Sistem Infirmasi, dengan penjurusan E-business di Fakultas Sains dan Teknologi Universitas Buddhi Dharma.

Gesima Chintia Angel Sirait, Mahasiswa Program Studi Sistem Infirmasi, dengan penjurusan E-business di Fakultas Sains dan Teknologi Universitas Buddhi Dharma.