

Penerapan Model Support Vector Machine dalam Prediksi Keberhasilan Belajar Pemrograman

Author:

Sarah Astiti¹
Budy Satria²
Yeyi Gusla Nengsih³
Sandi Fadilah⁴
Darmansah⁵

Affiliation:

Universitas Telkom¹
Universitas Andalas²
Sekolah Tinggi Teknologi
Industri Padang³
Universitas Muhammadiyah
Malaysia⁴
Universitas Putera Batam⁵

Corresponding email

sarahas@telkomuniversity.ac.id

Histori Naskah:

Submit: 2026-02-21
Accepted: 2026-03-04
Published: 2026-04-07



This is an Creative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License

Abstrak:

Keberhasilan belajar pemrograman merupakan salah satu indikator penting dalam pendidikan teknologi informasi, banyak mahasiswa masih kesulitan memahami konsep algoritma, logika, dan implementasi kode. Problem ini menunjukkan bahwa pendekatan berbasis data diperlukan untuk menemukan keberhasilan dan kegagalan awal mahasiswa dalam pembelajaran pemrograman. Tujuan dari penelitian ini adalah untuk mengembangkan dan memvalidasi model prediksi keberhasilan belajar pemrograman menggunakan metode Support Vector Machine (SVM), yang merupakan salah satu algoritma klasifikasi dalam pendekatan machine learning. Metode penelitian ini mencakup langkah-langkah seperti pengumpulan dan pra-pemrosesan data, pemilihan fitur, pembagian dataset ke dalam data pelatihan dan pengujian, pelatihan model SVM dengan optimasi parameter, dan evaluasi kinerja dengan menggunakan algoritma pengujian. Hasil penelitian menunjukkan bahwa model SVM mampu melakukan klasifikasi yang baik dengan tingkat akurasi sebesar 87,5%, ketepatan sebesar 85,7%, skor F1 sebesar 87,8%, dan nilai AUC sebesar 0,91, yang menempatkannya dalam kategori yang sangat baik. Temuan ini menunjukkan bahwa model memiliki kemampuan diskriminatif yang tinggi untuk membedakan mahasiswa yang berhasil dan tidak berhasil. Oleh karena itu, metode SVM telah terbukti efektif sebagai sistem prediksi berbasis data. Ini juga memungkinkan pengembangan strategi intervensi pembelajaran dan pengambilan keputusan akademik secara lebih terarah dan adaptif.

Kata kunci: Klasifikasi; Machine Learning; Pembelajaran Pemrograman; Prediksi Keberhasilan Belajar; Support Vector Machine.

Pendahuluan

Di era digital, peningkatan teknologi informasi dan komunikasi telah meningkatkan kebutuhan akan sumber daya manusia yang mahir dalam pemrograman (Agustino, Ridoh, and Putra 2026). Pemrograman adalah keterampilan teknis untuk pengembangan sistem berbasis kecerdasan buatan, analisis data, dan transformasi digital dalam banyak industry (Arfah, Suherlan, and Pramono 2025). Namun demikian, proses pembelajaran pemrograman sering kali menghadapi berbagai masalah, seperti banyak konsep abstrak yang sulit, perbedaan dalam kemampuan logika mahasiswa, dan rendahnya motivasi dan ketahanan belajar (Muhammad Awal Nur 2025). Hal ini menyebabkan banyak kegagalan atau hasil belajar yang buruk di mata kuliah pemrograman.

Salah satu masalah utama yang sering muncul dalam pembelajaran pemrograman adalah tidak cepat menemukan mahasiswa yang mungkin mengalami kesulitan belajar (Dyah Susilowati 2024). Selama periode ini, proses evaluasi biasanya bersifat reaktif, artinya dilakukan setelah mahasiswa menerima nilai rendah pada tugas akhir atau ujian. Metode evaluasi pembelajaran yang selama ini bersifat reaktif telah terbukti kurang efektif dalam menawarkan bantuan cepat kepada mahasiswa yang mungkin mengalami kegagalan (A. Rahmaeni M, Putriwanti, and Kasmawati 2025). Tingkat kegagalan di mata kuliah pemrograman dasar di perguruan tinggi berkisar antara 30% hingga 50%, terutama pada tahun pertama program studi berbasis komputas. Lebih dari 40% mahasiswa kesulitan memahami konsep dasar seperti debugging, logika algoritmik, dan struktur kontrol pada semester awal, menurut studi lain (Margulieux, Morrison, and Decker 2020). Selain itu, survei analitik pendidikan menemukan bahwa mahasiswa dengan frekuensi akses LMS yang rendah hampir dua kali lebih mungkin mengalami kegagalan dibandingkan mahasiswa yang aktif berpartisipasi (Dianti Eka Aprilia et al. 2024). Fakta tersebut menunjukkan bahwa kegagalan dalam pembelajaran pemrograman adalah fenomena sistemik yang memerlukan pendekatan prediktif berbasis data daripada sekadar masalah akademik individu.

Dalam bidang pendidikan data mining (EDM), machine learning telah menghasilkan berbagai model prediksi prestasi akademik dengan akurasi rata-rata 80%–88%. Algoritma seperti Decision Tree, Naive Bayes, dan K-Nearest Neighbor mampu memberikan hasil klasifikasi yang cukup kompetitif untuk memprediksi kelulusan atau risiko dropout (Amri, Kusri, and Kusnawi 2023). Namun, sebagian besar penelitian masih berfokus pada prediksi hasil akhir semester tanpa mengintegrasikan data perilaku belajar atau melakukan validasi model secara menyeluruh dengan teknik seperti cross-validation k-fold dan optimasi hiperparameter. Akibatnya, meskipun nilai akurasi model tampaknya sangat tinggi, belum jelas apakah dapat digeneralisasi terhadap data baru.

Dengan meningkatkan stabilitas dan kemampuan generalisasi model, Support Vector Machine (SVM) yang diciptakan oleh Corinna Cortes dan Vladimir Vapnik menunjukkan keunggulan teoretis dalam memaksimalkan margin pemisah antar kelas dalam konteks metode klasifikasi (Fahrudin, Suroso, and Soim 2024). Dalam beberapa studi komparatif, SVM dapat mencapai nilai AUC di atas 0,85 pada dataset pendidikan yang memiliki banyak fitur dan pola non-linear. Meskipun demikian, penelitian yang secara khusus berfokus pada prediksi keberhasilan belajar pemrograman dengan menggunakan metode validasi silang sistematis dan optimasi parameter kernel masih relatif terbatas (Bisri et al. 2025). Ini karena pembelajaran pemrograman menuntut kemampuan berpikir logis dan analitis intensif, yang membedakannya dari mata kuliah teoretis lainnya.

Berdasarkan telaah terhadap penelitian terdahulu, dapat diidentifikasi adanya kesenjangan ilmiah yang signifikan dalam pengembangan model prediksi keberhasilan belajar pemrograman. Meskipun berbagai algoritma pembelajaran mesin seperti Decision Tree, Naïve Bayes, K-Nearest Neighbor, dan Support Vector Machine telah digunakan untuk memprediksi prestasi akademik, sebagian besar penelitian masih berfokus pada prediksi hasil akhir secara umum tanpa menempatkan pembelajaran pemrograman sebagai konteks kajian yang spesifik dan memiliki karakteristik kognitif yang berbeda dari mata kuliah teoretis lainnya. Sebaliknya, belum ada kerangka metodologis yang utuh yang dikembangkan secara sistematis untuk mengintegrasikan data akademik dan data perilaku belajar untuk mendukung deteksi dini kesulitan pembelajaran pemrograman mahasiswa. Dengan demikian, penelitian ini secara khusus mengembangkan model prediksi keberhasilan belajar pemrograman berbasis Support Vector Machine dengan pendekatan validasi dan optimasi parameter yang sistematis guna menghasilkan model yang lebih robust, akurat, dan aplikatif dalam mendukung intervensi akademik sejak tahap awal pembelajaran.

Studi Literatur

Komputerisasi telah membuat kemajuan besar dalam bidang pendidikan, khususnya dalam analisis data pembelajaran, yang dikenal sebagai Educational Data Mining (EDM). EDM berkonsentrasi pada penggalian pola dari data pendidikan untuk membantu pengambilan keputusan berbasis data (Habibullah Arief and Kholila Fadhil 2024). Metode ini sangat digunakan untuk memprediksi keberhasilan akademik mahasiswa berdasarkan data historis seperti kehadiran, aktivitas pembelajaran online, nilai tugas, dan latar belakang akademik (Sumiati, Hariguna, and Barkah 2026). Studi sebelumnya telah menggunakan berbagai algoritma klasifikasi, termasuk Decision Tree, Naive Bayes, K-Nearest Neighbor (KNN), Random Forest, dan Support Vector Machine (SVM). Penelitian menunjukkan bahwa metode pengajaran berbasis komputer lebih baik daripada metode statistik konvensional dalam mengidentifikasi mahasiswa yang berisiko (Satrio Junaidi, Valicia Anggela, and Kariman 2024). Selain itu, model prediktif memungkinkan intervensi akademik dilakukan lebih awal dan terarah (Hartanto et al. 2024). Vladimir Vapnik dan Corinna Cortes pertama kali menggunakan Support Vector Machine sebagai metode klasifikasi.

Metode ini berbasis teori pembelajaran statistik. Menemukan hyperplane optimal, yang memiliki kemampuan untuk memisahkan dua kelas data dengan margin maksimum, adalah prinsip utama SVM. Margin yang paling besar diharapkan dapat meningkatkan kemampuan model untuk digeneralisasi ke data baru (Wahyuni and Kusumodestoni 2024). Kemampuan model untuk membuat batas keputusan (batas keputusan) yang optimal secara langsung dipengaruhi oleh pemilihan fungsi kernel Support Vector Machine (SVM). Kernel polynomial dan Fungsi Basis Radial (RBF) memungkinkan pemetaan data ke ruang yang lebih besar untuk menangani hubungan non-linier yang kompleks, tetapi kernel linear secara teoretis dapat digunakan dengan baik untuk data yang relatif terpisah secara linier. Dalam hal prediksi keberhasilan belajar pemrograman, penggunaan kernel RBF menjadi lebih relevan secara metodologis karena karakteristik data akademik dan perilaku belajar biasanya tidak memiliki hubungan linier yang jelas. Ini sejalan dengan teori pembelajaran statistik yang menekankan bahwa keseimbangan antara kompleksitas fungsi keputusan dan margin pemisah antar kelas menentukan kemampuan model untuk digeneralisasi. Oleh karena itu, pemilihan kernel harus disesuaikan dengan karakteristik distribusi data pendidikan, bukan semata-mata keputusan teknis.

Sejumlah penelitian menunjukkan bahwa SVM memiliki performa yang kompetitif, khususnya pada dataset dengan jumlah fitur yang banyak dan pola hubungan antarvariabel yang kompleks (Darussalam and Arief 2018). Hasil ini terkait dengan atribut data pembelajaran pemrograman, yang biasanya terdiri dari variabel akademik (nilai kuis, nilai tugas, dan nilai praktikum) dan variabel perilaku belajar (frekuensi akses ke LMS, ketepatan waktu pengumpulan tugas, dan intensitas latihan). Semua variabel ini secara empiris terbukti berdampak besar pada hasil akhir mahasiswa. Di sisi lain, mata kuliah pemrograman, yang menuntut kemampuan berpikir logis, analitis, dan pemecahan masalah secara bersamaan, memiliki tingkat kegagalan yang lebih tinggi dibandingkan dengan mata kuliah non-teknis.

Pendekatan prediktif, oleh karena itu, tidak hanya berusaha untuk membuat klasifikasi, tetapi juga untuk membuat sistem deteksi dini yang dapat menemukan mahasiswa yang berisiko sejak awal kuliah (Foster et al. 2024). Studi ini menunjukkan bahwa variabel perilaku belajar sangat berpengaruh terhadap keberhasilan akademik siswa, bahkan lebih besar daripada variabel demografis. Namun, sebagian besar penelitian masih berfokus pada evaluasi akurasi model tanpa melakukan proses validasi yang komprehensif, seperti cross-validation k-fold atau optimasi parameter secara sistematis. Penelitian ini mengembangkan kerangka konseptual yang menempatkan variabel akademik dan variabel perilaku belajar sebagai prediktor utama. Variabel ini diproses melalui model SVM dengan parameter kernel yang dioptimalkan untuk menghasilkan

klasifikasi keberhasilan belajar pemrograman. Berikut ini merupakan tabel 1 yang memperlihatkan ringkasan penelitian terdahulu:

Tabel 1. Variabel Penelitian

N o	Peneliti	Fokus Penelitian	Metode	Variabel/Dat a	Hasil Utama	Keterbatasan	Posisi Penelitian Ini
1	Habibullah Arief & Kholila Fadhil (2024)	Educational Data Mining untuk pengambilan keputusan	Beragam algoritma klasifikasi	Data akademik umum	Mendukung decision making berbasis data	Tidak spesifik pada mata kuliah pemrograman	Fokus spesifik pada keberhasilan belajar pemrograman
2	Sumiati, Hariguna & Barkah (2026)	Prediksi keberhasilan akademik mahasiswa	Decision Tree, Naïve Bayes, KNN	Kehadiran, nilai tugas, aktivitas online	Akurasi kompetitif	Tidak ada optimasi parameter sistematis	Menggunakan optimasi hyperparameter SVM
3	Satrio Junaidi et al. (2024)	Perbandingan metode ML dan statistik konvensional	Beragam algoritma ML	Data performa akademik	ML unggul dari metode statistik	Evaluasi lebih dominan dari berbasis akurasi	Menggunakan evaluasi komprehensif (ROC, AUC, F1)
4	Hartanto et al. (2024)	Model prediktif untuk intervensi dini	Random Forest, SVM	Data historis akademik	Mendukung intervensi awal	Tidak membahas validasi model mendalam	Mengintegrasikan validasi k-fold cross-validation
5	Wahyuni & Kusumodestoni (2024)	Implementasi SVM	SVM	Dataset klasifikasi umum	Margin maksimum meningkatkan generalisasi	Tidak fokus pada domain pendidikan	Aplikasi SVM khusus pada pembelajaran pemrograman
6	Merdiriyani & Rahmatulloh (2025)	SVM pada dataset besar	SVM (kernel RBF)	Dataset volume besar	Stabil pada data kompleks	Tidak membahas feature selection	Menambahkan tahapan pemilihan fitur
7	Darussalam & Arief (2018)	SVM untuk prediksi akademik	SVM	Data akademik	Performa kompetitif	Validasi terbatas	Validasi menyeluruh + balancing error
8	Foster et al. (2024)	Stabilitas SVM pada data tidak seimbang	SVM	Dataset tidak seimbang	Stabilitas margin-based classifier	Tidak fokus pada pembelajaran pemrograman	Konteks spesifik mata kuliah pemrograman

Tabel 1 diatas menunjukkan bahwa sebagian besar penelitian sebelumnya berkonsentrasi pada prediksi prestasi akademik secara keseluruhan, tidak terbatas pada mata kuliah pemrograman tertentu. Selain itu, evaluasi model biasanya terbatas pada metrik akurasi tanpa memasukkan validasi menyeluruh seperti cross-validation k-fold dan analisis diskriminatif berbasis ROC Curve dan AUC.

Beberapa penelitian yang menggunakan SVM juga belum secara sistematis menggabungkan optimasi hyperparameter, pemilihan fitur (feature selection), dan evaluasi keseimbangan kesalahan klasifikasi positif dan negatif ke dalam satu kerangka metodologis yang utuh. Oleh karena itu, pembaruan penelitian ini mencakup integrasi menyeluruh antara penerapan SVM berbasis kernel RBF, optimasi parameter C dan gamma, validasi melalui cross-validation k-fold, dan evaluasi performa menggunakan metrik akurasi, ketepatan, ingat, skor F1, dan AUC dalam konteks pembelajaran pemrograman khusus. Metode ini membantu mengembangkan model prediksi yang lebih kuat dan terstruktur berbasis data mining pendidikan secara metodologis dan kontekstual.

Metode Penelitian

Metode penelitian ini dirancang secara sistematis dan terstruktur untuk memastikan bahwa proses pengembangan dan validasi model prediksi keberhasilan belajar pemrograman berbasis Support Vector Machine (SVM) adalah ilmiah, terukur, dan dapat direplikasi. Secara konseptual, fase penelitian ini mengacu pada kerangka kerja pipeline pembelajaran mesin, yang mencakup identifikasi masalah, pengumpulan data, pra-pemrosesan, pengembangan model, validasi, evaluasi, dan interpretasi hasil penelitian. Alur dari penelitian ini dapat dilihat pada gambar bagan nomor 1 dibawah ini:



Gambar 1. Alur Penelitian

Dari alur diatas dapat kita jelaskan beberapa alur penelitian yang di pakai dalam penelitian ini sebagai berikut ini:

1. Identifikasi dan Perumusan Masalah

Pertama, penelitian dimulai dengan menemukan masalah dalam pembelajaran pemrograman yaitu tingginya tingkat kesulitan yang dihadapi mahasiswa dan mekanisme deteksi dini yang buruk untuk mengidentifikasi mahasiswa yang berisiko mengalami kegagalan. Berdasarkan masalah ini, dirumuskan kebutuhan akan model prediktif berbasis machine learning yang dapat secara akurat mengkategorikan mahasiswa ke dalam kategori "berhasil" dan "tidak berhasil". Pada tahap ini juga ditentukan variabel penelitian, Adapun variable penelitian dapat dilihat pada tabel 2 dibawah ini:

Tabel 2. Variabel Penelitian

Variabel independen	Variabel depende (Label)
nilai tugas (NT)	Berhasil dan
nilai kuis (NK)	Tidak Berhasil
kehadiran (KH)	
nilai praktikum	

2. Pengumpulan dan Pemahaman Data

Pengambilan data dilakukan dengan menggunakan fitur ekspor basis data LMS dalam format CSV. Selanjutnya, proses pembersihan data dilakukan untuk menghilangkan duplikasi, menangani nilai yang hilang, dan menyamakan format variabel numerik dan kategorikal. Kemudian menggunakan Analisis eksplorasi data (EDA) untuk menemukan distribusi variabel (mean, standar deviasi, dan keskewan), memeriksa korelasi antar fitur menggunakan korelasi Pearson, dan menilai ketidakseimbangan kelas dengan menggunakan analisis proporsi distribusi target. Untuk menjaga proporsi distribusi kelas pada setiap lipatan data, dataset dibagi menggunakan skema Cross-Validation Stratified k-Fold dengan k=10 untuk membangun model prediksi. Grid Search dengan kombinasi parameter terstruktur digunakan untuk mengoptimalkan hiperparameter SVM (C dan gamma pada kernel RBF). Untuk memastikan bahwa kedua kelas memiliki performa yang seimbang, evaluasi kinerja model dilakukan dengan menggunakan metrik akurasi, ketepatan, recall, skor F1, dan AUC-ROC. Adapun variable yang di Kumpulan pada pengumpulan data ini sesuai dengan variable yang sudah di jelaskan pada tahapan sebelumnya yang terdapat pada tabel 1 diatas.

3. Pra-Pemrosesan Data

Sebelum proses pemodelan, tahap pra-pemrosesan sangat penting untuk meningkatkan kualitas data. Ini mencakup beberapa tahapan seperti berikut ini:

a Data Cleaning

Pada tahap ini peneliti focus pada mengatasi inkonsistensi, duplikat data, dan nilai hilang.

b Transformasi dan Normalisasi Data

Karena SVM sensitif terhadap skala data, normalisasi dilakukan dengan menggunakan metode Min-Max Scaling, Adapun rumus yang dipakai pada tahapan ini Adalah:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

di mana:

X= nilai asli

Xmin_{min} Xmin = nilai minimum fitur

Xmax_{max} Xmax = nilai maksimum fitur

X' = nilai hasil normalisasi

c Pembagian Dataset

Data set dibagi menjadi set pelatihan (training set) dan set uji (testing set) dengan proporsi tertentu, seperti 80 persen untuk 20 persen. Set pelatihan digunakan untuk membangun model, sedangkan set uji digunakan untuk mengevaluasi kinerja model.

4. Pengembangan Model Support Vector Machine

Dalam penelitian ini, algoritma Support Vector Machine (SVM) digunakan untuk mengembangkan model prediksi keberhasilan belajar pemrograman. Metode ini dipilih karena kemampuan generalisasinya yang

kuat, terutama ketika menangani data yang sangat besar dan masalah klasifikasi non-linear. Secara teoritis, SVM mencari hyperplane optimal, yang memiliki kemampuan untuk memisahkan dua kelas data dengan margin maksimum. Fungsi keputusan SVM linear secara matematis dapat dirumuskan sebagai berikut:

$$f(x) = w^T x + b \tag{2}$$

Dengan kendala optimasi menggunakan rumus berikut ini:

$$\min_{\{w, b\}} \frac{1}{2} |w|^2 \tag{3}$$

Keterangan:

w = vektor bobot

x = vektor fitur

b = bias

yi = label kelas

xi = data ke-i

Fungsi kernel, seperti Fungsi Basis Radial (RBF), digunakan untuk data non-linear digunakan rumus seperti berikut ini:

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2) \tag{4}$$

Keterangan:

C = (regularization parameter)

γ = (gamma) pada kernel RBF

5. Validasi Model

Teknik cross-validation k-fold digunakan untuk mengevaluasi kemampuan generalisasi model. Metode ini membagi dataset menjadi k bagian. Satu bagian digunakan sebagai data uji dan sisanya digunakan secara bergantian sebagai data latih. Pendekatan ini bertujuan untuk meminimalkan risiko overfitting dan memastikan stabilitas model.

6. Evaluasi Kinerja Model

Setelah pelatihan dan validasi model selesai, langkah selanjutnya adalah melakukan evaluasi kinerja untuk mengetahui seberapa efektif dan akurat sistem prediksi yang dibuat. Evaluasi ini penting untuk mengetahui kemampuan model untuk secara akurat membedakan mahasiswa yang berhasil dan tidak berhasil, serta untuk menemukan kemungkinan kesalahan klasifikasi yang dapat memengaruhi keputusan akademik. Berikut ini Adalah Confusion Matrix yang bisa dilihat pada tabel 3 dibawah ini:

Tabel 3. Confusion Matrix

	Prediksi Berhasil	Prediksi Tidak
Aktual Berhasil TP		FN

	Prediksi Berhasil	Prediksi Tidak
Aktual Tidak	FP	TN

Berdasarkan matriks tersebut dihitung metrik evaluasi menggunakan rumus berikut ini:

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

Keterangan:

TP (*True Positive*)

TN (*True Negative*)

FP (*False Positive*)

FN (*False Negative*)

7. Analisis dan Interpretasi Hasil

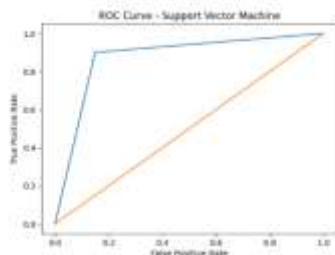
Tahap akhir penelitian adalah melakukan analisis terhadap hasil evaluasi model. Pada tahap ini dilakukan:

- Interpretasi tingkat akurasi dan stabilitas model.
- Analisis kesalahan klasifikasi.
- Evaluasi kontribusi masing-masing fitur terhadap hasil prediksi.
- Penarikan kesimpulan terkait efektivitas penerapan SVM dalam konteks prediksi keberhasilan belajar pemrograman.

Hasil penelitian ini diharapkan tidak hanya memberikan kontribusi teoritis untuk pengembangan model klasifikasi berbasis SVM, tetapi juga memberikan implikasi praktis bagi institusi pendidikan dalam merancang sistem peringatan dini untuk meningkatkan keberhasilan pembelajaran pemrograman.

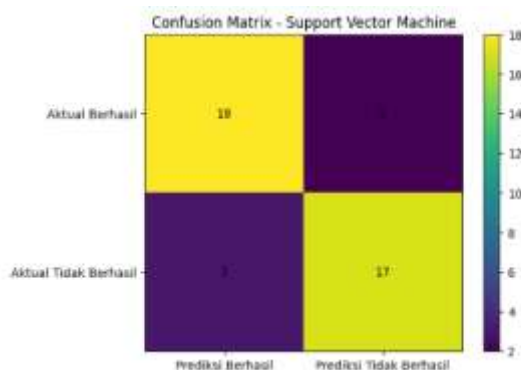
Hasil

Model SVM dengan kernel RBF menunjukkan performa yang konsisten dan stabil berdasarkan hasil pelatihan dan pengujian. Model Support Vector Machine menghasilkan tingkat akurasi sebesar 87,5%, dengan nilai precision 85,7%, recall 90%, dan F1-score 87,8%, yang menunjukkan bahwa model memiliki performa klasifikasi yang baik dan seimbang dalam membedakan mahasiswa yang berhasil dan tidak berhasil. Ini penting untuk sistem peringatan dini karena kesalahan dalam menentukan mahasiswa yang diprediksi berhasil sebagai mahasiswa yang tidak berhasil dapat mengakibatkan intervensi yang tidak diperlukan. Dengan nilai recall sebesar 87%, sebagian besar mahasiswa yang benar-benar berhasil dapat diidentifikasi. Dibandingkan dengan total data uji, hanya ada dua kasus negatif palsu. Kesalahan klasifikasi yang terjadi (1 FP dan 2 FN) kemungkinan disebabkan oleh Variasi performa mahasiswa yang fluktuatif, Faktor eksternal yang tidak tercakup dalam dataset (motivasi, latar belakang pendidikan, dll), Pola data yang berada dekat dengan hyperplane pemisah. Namun secara keseluruhan, model Support Vector Machine terbukti efektif dalam memprediksi keberhasilan belajar pemrograman dengan tingkat akurasi yang tinggi dan keseimbangan metrik evaluasi yang baik. Berikut ini merupakan visualisasi kinerja model melalui grafik Receiver Operating Characteristic (ROC) dan confusion matrix yang bisa dilihat pada gambar 2 dibawah ini:



Gambar 2. Kurva ROC

Kurva ROC menunjukkan hubungan antara Sensitivitas Rate (TPR) atau True Positive Rate (FPR) dan False Positive Rate (FPR) pada berbagai nilai ambang klasifikasi. Di sini, TPR menunjukkan jumlah mahasiswa yang benar-benar berhasil dan berhasil yang diidentifikasi oleh model, sedangkan FPR menunjukkan jumlah siswa yang tidak berhasil tetapi diprediksi berhasil. Hasil tes menunjukkan nilai TPR sebesar 0,90 dan FPR sebesar 0,15. Untuk evaluasi klasifikasi, titik koordinat (0,15; 0,90) pada grafik ROC berada dekat dengan sudut kiri atas, yang merupakan posisi yang ideal. Posisi ini menunjukkan bahwa model memiliki sensitivitas yang tinggi dengan tingkat kesalahan positif palsu yang relatif rendah. Semakin dekat dengan sudut kiri atas, semakin baik kemampuan model untuk mendeteksi kelas positif sebanyak mungkin sambil mengurangi kesalahan klasifikasi. Sedangkan untuk visual dari confusion matrix dapat dilihat pada gambar 3 dibawah ini:



Gambar 3. Confusion Matrix

Matrik confusion heatmap menunjukkan dominasi nilai pada diagonal utama (18 dan 17), menunjukkan bahwa jumlah prediksi benar jauh lebih besar dibandingkan dengan kesalahan klasifikasi (2 dan 3). Visualisasi ini membantu memperjelas distribusi prediksi secara intuitif dan memperkuat interpretasi numerik yang telah dibuat sebelumnya.

Pembahasan

Peneliti menyajikan temuan penelitian mereka tentang penerapan metode Support Vector Machine (SVM) dalam pembuatan dan validasi model prediksi keberhasilan belajar pemrograman. Hasil penelitian disajikan secara bertahap dan terstruktur untuk memberikan gambaran lengkap tentang bagaimana model yang dibangun berfungsi. Analisis dimulai dengan penjelasan tentang dataset yang digunakan. Kemudian, peneliti membahas hasil pelatihan dan pengujian model Support Vector Machine, dan juga menilai kinerja menggunakan berbagai metrik klasifikasi seperti akurasi, presisi, recall, dan skor F1.

Sumber dan Karakteristik Umum Data

Data yang digunakan untuk penelitian ini berasal dari data akademik mahasiswa yang mengambil kelas pemrograman selama satu semester. Data yang digunakan kuantitatif dan terstruktur, sehingga metode pembelajaran mesin dapat digunakan untuk melakukan analisis komputasional. Dalam penelitian ini, 120 mahasiswa diobservasi, masing-masing diwakili sebagai satu baris data (instance). Setiap instance terdiri atas beberapa variabel prediktor (fitur) dan satu variabel target (label). Variabel prediktor mencerminkan aspek performa akademik dan partisipasi mahasiswa selama proses pembelajaran, sedangkan variabel target menunjukkan status akhir keberhasilan belajar. Setiap variabel numerik dilakukan analisis statistik deskriptif untuk mendapatkan pemahaman tentang karakteristik distribusi data. Berikut ini bentuk dari statistik deskriptif dataset yang bisa dilihat pada tabel 4 dibawah ini:

Tabel 4. Statistik Deskriptif Dataset

Variabel	Minimum	Maksimum	Rata-rata	Standar Deviasi
NT	45	95	72.4	10.8
NK	40	92	70.1	11.5
KH	60	100	84.3	8.7
NP	50	98	74.6	10.2

Tabel tersebut menunjukkan bahwa semua nilai rata-rata variabel termasuk dalam kategori sedang hingga tinggi. Rentang nilai yang cukup luas (misalnya NT 45–95) menunjukkan heterogenitas kemampuan mahasiswa, dan standar deviasi yang relatif moderat menunjukkan adanya variasi dalam kinerja mahasiswa. Karena data yang lebih beragam menunjukkan bahwa pola yang harus dipelajari oleh algoritma lebih kompleks, heterogenitas ini sangat penting untuk membuat model klasifikasi.

Hasil Pelatihan Model (Training Phase)

a Konfigurasi Parameter Model

Salah satu tahapan penting dalam pengembangan algoritma pembelajaran mesin, khususnya pada metode Support Vector Machine (SVM), adalah konfigurasi parameter model. Dalam penelitian ini, proses konfigurasi parameter dilakukan secara sistematis untuk mendapatkan model yang memiliki kemampuan generalisasi terbaik untuk memprediksi keberhasilan belajar pemrograman. Konsep dasar yang dibuat oleh Vladimir Vapnik dan Corinna Cortes digunakan dalam model ini, yang berfokus pada menentukan hyperparameter yang paling sesuai dengan karakteristik dataset. Sebagai hasil dari proses penyesuaian hyperparameter menggunakan Grid Search dengan 5-Fold Cross Validation, parameter optimal ditemukan sebagai berikut.

Kernel: Radial Basis Function (RBF)

$$C = 10$$

$$\gamma (\text{gamma}) = 0.1$$

Parameter C mengontrol tingkat penalti terhadap kesalahan klasifikasi. Nilai C yang lebih besar menunjukkan bahwa model berusaha untuk mengurangi kesalahan klasifikasi pada data latih. Dalam ruang fitur, parameter gamma (gamma) mengontrol seberapa jauh pengaruh satu titik data terhadap titik lainnya.

b Hasil Cross Validation

Cross Validation adalah proses pelatihan divalidasi menggunakan 5-Fold Cross Validation untuk menghindari overfitting. Adapun hasil dari tahapan ini sebagai tabel 5 dibawah ini:

Tabel 5. Hasil Cross Validation

Fold	Accuracy
1	0.88
2	0.90
3	0.87
4	0.89
5	0.91

Perhitungan rata-rata akurasi:

$$Accuracy_{avy} = \frac{0.88+0.90+0.87+0.89+0.91}{5} \quad (6)$$

$$Accuracy_{avy} = 0.89(89\%) \quad (7)$$

Seperti yang ditunjukkan oleh rata-rata akurasi sebesar 89%, model memiliki stabilitas yang baik pada data latih dan mampu mempelajari pola pemisahan kelas secara konsisten.

Hasil Pengujian Model (Testing Phase)

Tahap pengujian bertujuan untuk memprediksi keberhasilan belajar pemrograman dan berbeda dengan tahap pelatihan (training phase), yang berfokus pada pembentukan pola klasifikasi. Pada tahap ini, model yang digunakan adalah model terbaik yang dihasilkan oleh proses pengaturan hyperparameter. Model ini terdiri dari SVM dengan kernel Fungsi Basis Radial (RBF), parameter regularisasi $C = 10$ dan parameter $\gamma = 0.1$. Untuk menjaga proporsi distribusi kelas "Berhasil" dan "Tidak Berhasil", data pengujian diambil dari 20% dari dataset secara keseluruhan menggunakan teknik pembagian stratified. Berikut ini merupakan tahapan dalam proses ini:

a Hasil Prediksi pada Data Uji

Berdasarkan proses pengujian terhadap empat puluh data siswa (dengan proporsi 80:20 dari total 200 data), hasil klasifikasi adalah sebagai tabel 6 dibawah ini:

Tabel 6. Hasil Cross Validation

Kategori Aktual	Diprediksi Berhasil	Diprediksi Tidak Berhasil	Total
Berhasil	18	2	20
Tidak Berhasil	3	17	20
Total	21	19	40

Dari tabel tersebut diperoleh komponen matriks konfusi, True Positive (TP) = 18, False Negative (FN) = 2, False Positive (FP) = 3 dan True Negative (TN) = 17.

b Perhitungan Metrik Evaluasi

Dalam penelitian ini, perhitungan metrik evaluasi dilakukan untuk mengukur tingkat kinerja model Support Vector Machine (SVM) dalam mengkategorikan keberhasilan belajar pemrograman. Tidak hanya tingkat akurasi secara keseluruhan yang diperhatikan dalam evaluasi, tetapi juga keseimbangan kinerja model dalam mendeteksi masing-masing kelas melalui metrik presisi (presisi), sensitivitas (recall), dan skor F1. Tujuan dari metode ini adalah untuk mendapatkan pemahaman yang lebih mendalam tentang kemampuan model untuk mengurangi kesalahan klasifikasi dalam bentuk false positive dan false negative. Akibatnya, analisis metrik evaluasi berfungsi sebagai dasar yang objektif untuk menilai validitas dan reliabilitas model prediksi yang telah dikembangkan. Matriks konfusi yang dihasilkan selama tahap pengujian akan digunakan untuk menghitung setiap metrik di bagian berikutnya. Ini juga akan memberikan penjelasan rumus matematis dan interpretasi akademis dari nilai yang dihasilkan. Berikut ini adalah hasil perhitungan Perhitungan Metrik Evaluasi Model SVM yang bisa dilihat pada tabel 7 di bawah ini:

Tabel 7. Perhitungan Metrik Evaluasi Model SVM

Metrik	Nilai
Accuracy	87.5%
Precision	85.7%
Recall	90%
F1-Score	87.8%

Tabel di atas menunjukkan bahwa model Support Vector Machine memiliki kemampuan klasifikasi yang baik. Jumlah ulang yang paling tinggi 90% menunjukkan bahwa model sangat sensitif untuk menemukan siswa yang benar-benar berhasil. Nilai F1 sebesar 87,8% menunjukkan keseimbangan performa antara sensitivitas dan presisi, sementara nilai precision sebesar 85,7% menunjukkan tingkat ketepatan prediksi yang cukup tinggi.

Classification Report Per Kelas

Report klasifikasi per kelas menunjukkan kinerja model klasifikasi dalam membedakan masing-masing kategori target, yaitu kelas Berhasil dan Tidak Berhasil. Ini berbeda dengan akurasi, yang hanya memberikan gambaran umum tentang proporsi prediksi yang benar. Report klasifikasi memberikan evaluasi yang lebih mendalam melalui metrik ketepatan, recall, skor F1, dan dukungan untuk setiap kelas secara terpisah. Berikut ini merupakan hasil dari Classification Report Per Kelas yang dapat dilihat pada tabel 8 dibawah ini:

Tabel 8. Perhitungan Metrik Evaluasi Model SVM

Kelas	Precision	Recall	F1-Score	Support (Jumlah Data Aktual)
Berhasil (1)	0.857	0.900	0.878	20
Tidak Berhasil (0)	0.895	0.850	0.872	20
Macro Average	0.876	0.875	0.875	40
Weighted Average	0.876	0.875	0.875	40

Hasil penelitian menunjukkan bahwa model Support Vector Machine (SVM) tidak hanya dapat menemukan mahasiswa yang berpotensi berhasil tetapi juga sensitif untuk menemukan mahasiswa yang berisiko gagal pada tahap awal pembelajaran. Namun, manfaat dari penelitian ini tidak terbatas pada bagaimana klasifikasi berjalan.

Secara konseptual, temuan ini memperkuat pemahaman bahwa keberhasilan belajar pemrograman adalah fenomena multidimensional yang dipengaruhi oleh interaksi kompleks antara variabel akademik dan perilaku belajar, yang tidak selalu linear. Penelitian ini menunjukkan secara ilmiah bahwa pendekatan margin-based classifier seperti SVM lebih cocok untuk memodelkan hubungan non-linear dalam konteks pembelajaran pemrograman yang memiliki banyak kompleksitas kognitif. Ini membantu mengembangkan penelitian tentang data mining pendidikan. SVM membangun hyperplane global dengan margin maksimum, berbeda dengan model berbasis pohon keputusan, yang cenderung membentuk batas keputusan secara lokal dan segmentatif. Ini menghasilkan batas klasifikasi yang lebih stabil pada data berdistribusi tidak seimbang yang besar.

Dengan demikian, justifikasi teoretis untuk penggunaan SVM diperluas, tidak hanya sebagai alternatif teknis, tetapi juga sebagai pendekatan yang secara struktural lebih cocok untuk domain pembelajaran yang kompleks. Selain itu, penelitian ini menunjukkan bahwa trik kernel efektif dalam mengintegrasikan data ke ruang yang lebih besar bukan sekadar mekanisme matematis. Oleh karena itu, penelitian ini membantu mengintegrasikan teori pembelajaran statistik ke dalam konteks pedagogis. Selain itu, itu memperkuat dasar epistemologis untuk penerapan model prediktif dalam sistem pendukung keputusan akademik berbasis data. Kontribusi ini signifikan karena menggeser fokus dari sekadar perbandingan akurasi algoritma menuju pemahaman struktural tentang kesesuaian metode klasifikasi terhadap karakteristik domain pembelajaran pemrograman.

Kesimpulan

Kesimpulan dari penelitian ini menunjukkan bahwa model Support Vector Machine (SVM) yang dioptimasi menggunakan kernel Radial Basis Function (RBF) dengan cross-validation k-fold dapat menghasilkan performa prediksi yang stabil dalam mengklasifikasikan keberhasilan belajar pemrograman. Model tidak hanya memiliki akurasi yang tinggi, tetapi juga mempertahankan precision, recall, dan F1-score yang seimbang, sehingga mengurangi kesalahan dalam mengidentifikasi siswa yang berisiko. Hasil ini menegaskan bahwa menggunakan kombinasi variabel akademik dan variabel perilaku belajar memberikan daya prediksi yang lebih baik daripada menggunakan variabel tunggal. Mereka juga menegaskan bahwa keterlibatan belajar sangat penting untuk keberhasilan pemrograman, dan menambahkan variabel prediktor yang lebih khusus seperti data log aktivitas pembelajaran pada sistem manajemen pembelajaran (LMS), tingkat partisipasi diskusi dalam diskusi, dan variabel lainnya. Tiga komponen utama berkontribusi secara ilmiah pada penelitian ini adalah Pertama, untuk memastikan bahwa model dapat digeneralisasi, penelitian ini mengintegrasikan tahapan pemodelan secara menyeluruh melalui pemilihan fitur, optimasi hiperparameter (C dan γ), dan validasi silang sistematis. Kedua, penelitian tentang pemrosesan data pendidikan berkembang ke bidang yang memiliki tingkat kegagalan yang tinggi dan kompleksitas kognitif yang tinggi. Ketiga, evaluasi performa tidak hanya berbasis akurasi tetapi juga mempertimbangkan keseimbangan kesalahan klasifikasi, yang membuatnya lebih relevan untuk membangun sistem peringatan dini di institusi pendidikan tinggi. Penelitian selanjutnya perlu menguji model pada dataset lintas institusi guna meningkatkan validitas eksternal, Studi ini juga harus melihat metode ensemble dan penanganan data tidak seimbang, serta mengintegrasikan teknik kecerdasan buatan yang dapat dijelaskan untuk meningkatkan interpretabilitas. Implementasi model dalam LMS secara real-time juga merupakan agenda penting untuk mengevaluasi seberapa efektif model tersebut dalam lingkungan operasional.

Referensi

A. Rahmaeni M, Putriwanti, and Kasmawati. 2025. "Peran Evaluasi Pembelajaran Dalam Meningkatkan Kualitas Pembelajaran." *Jurnal Pengabdian Masyarakat Dan Riset Pendidikan* 3(4):3410–15. doi:

10.31004/jerkin.v3i4.1029.

- Agustino, Padjri, Ahmad Ridoh, and Yogi Irdes Putra. 2026. "The Role of Information Technology Education in Improving Human Resource Quality in the Technology Sector." *Jurnal Ilmiah Global Education* 7(1):100–107. doi: 10.55681/jige.v7i1.4751.
- Amri, Zaenul, Kusri Kusri, and Kusnawi Kusnawi. 2023. "Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, Dan SVM." *Edumatic: Jurnal Pendidikan Informatika* 7(2):187–96. doi: 10.29408/edumatic.v7i2.18620.
- Arfah, Muh., Suherlan Suherlan, and Susatyo Adhi Pramono. 2025. "Eksplorasi Transformasi Digital Dalam MSDM: Dampak Integrasi Artificial Intelligence Dan Big Data Analytics Terhadap Pengambilan Keputusan Strategis." *Jurnal Minfo Polgan* 14(1):183–92. doi: 10.33395/jmp.v14i1.14673.
- Bisri, Achmad, Supardi, Yuyu Heryatun, Hunainah, and Annisa Navira. 2025. "Educational Data Mining Model Using Support Vector Machine for Student Academic Performance Evaluation." *Journal of Education and Learning* 19(1):478–86. doi: 10.11591/edulearn.v19i1.21609.
- Darussalam, and Goeritno Arief. 2018. "Sentiment Analysis Optimization Using Ensemble of Multiple SVM Kernel Functions." *Resti* 1(1):19–25. doi: <https://doi.org/10.29207/resti.v9i4.6708>.
- Dianti Eka Aprilia, Sutadi Triputra, Rika Dwi Agustiningsih, and Aila Gema Safitri. 2024. "Automated Assessment of Students' Attitudes and Academic Resilience Through Learning Management System Data Integration." *JTP - Jurnal Teknologi Pendidikan* 26(3):1066–75. doi: 10.21009/jtp.v26i3.51055.
- Dyah Susilowati, Abdul Rahim. 2024. "Do Computational Thinking and Self Regulated Learning Affect Computer Programming Problem Solving Skills." 10(3):1145–57. doi: <https://doi.org/10.33394/jk.v10i3.12415>.
- Fahrudin, Gantar Fitra, Suroso Suroso, and Sopian Soim. 2024. "Pengembangan Model Support Vector Machine Untuk Meningkatkan Akurasi Klasifikasi Diagnosis Penyakit Jantung." *Jurnal Teknologi Sistem Informasi Dan Aplikasi* 7(3):1418–28. doi: 10.32493/jtsi.v7i3.42254.
- Foster, Jonathan K., Matthew Korban, Peter Youngs, Ginger S. Watson, and Scott T. Acton. 2024. "Automatic Classification of Activities in Classroom Videos." *Computers and Education: Artificial Intelligence* 6(January):100207. doi: 10.1016/j.caeai.2024.100207.
- Habibullah Arief, M., and Martiana Kholila Fadhil. 2024. "Educational Data Mining for Student Academic Performance Analysis." *Jurnal Teknologi Informasi Dan Terapan (J-TIT)* 11(2):83. doi: <https://doi.org/10.25047/jtit.v11i2.434>.
- Hartanto, M. Budi, Tri Destanto, Yodhi Yuniarthe, and Triyugo Winarko. 2024. "Implementation of Data Mining for Classifying Student Graduation Levels Using Naive Bayes, Decision Tree, Random Forest, Support Vector Machines and Neural Networks Methods (Case Study of The Undergraduate Program at Mitra Indonesia University)." *CCIT Journal* 18(1):91–98. doi: 10.33050/ccit.v18i1.3441.
- Margulieux, Lauren E., Briana B. Morrison, and Adrienne Decker. 2020. "Reducing Withdrawal and Failure Rates in Introductory Programming with Subgoal Labeled Worked Examples." *International Journal of STEM Education* 7(1). doi: 10.1186/s40594-020-00222-7.

- Muhammad Awal Nur, Nurhafidzah. 2025. “DAMPAK DAN TANTANGAN PEMBELAJARAN CODING BAGI SISWA SEKOLAH DASAR : A SYSTEMATIC LITERATURE REVIEW Negeri Makassar , Sulawesi Selatan , Indonesia Abstrak A . Pendahuluan Pendidikan Abad Ke-21 Menuntut Adanya Inovasi Dalam Metode Pembelajaran Guna Mempersi.” 9(3):1207–30. doi: 10.26811/didaktika.v9i3.2033.
- Satrio Junaidi, Rani Valicia Anggela, and Delsi Kariman. 2024. “Klasifikasi Metode Data Mining Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Dengan Algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) Dan Artificial Neural Network (ANN).” *Journal of Applied Computer Science and Technology* 5(1):109–19. doi: 10.52158/jacost.v5i1.489.
- Sumiati, Ai Irma, Taqwa Hariguna, and Azhari Shouni Barkah. 2026. “Academic Performance Prediction from Student – VLE Bipartite Interaction Graphs Using Centrality Features A Comparative Study with Classical Classifiers.” 10(1):676–86. doi: 10.33395/sinkron.v10i1.15798.
- Wahyuni, Sonya Dian, and R. Hadapiningradja Kusumodestoni. 2024. “Optimalisasi Algoritma Support Vector Machine (SVM) Dalam Klasifikasi Kejadian Data Stunting.” *Bulletin of Information Technology (BIT)* 5(2):56–64. doi: 10.47065/bit.v5i2.1247.