

IMPLEMENTASI ALGORITMA *MACHINE LEARNING* UNTUK PREDIKSI KEBERHASILAN MAHASISWA DI PROGRAM STUDI TEKNIK INFORMATIKA

Hardiansyah¹, Ichsan Ramdhani², Mukhamad Khotib Arifai³

^{1,2,3} Universitas Pamulang

E-mail: ¹dosen02058@unpam.ac.id, ²dosen02110@unpam.ac.id, ³dosen01995@unpam.ac.id

ABSTRAK

Tingginya angka drop out (DO) menjadi salah satu permasalahan signifikan yang dihadapi institusi pendidikan tinggi, yang berdampak pada mahasiswa, institusi, dan kualitas sumber daya manusia secara umum. Penelitian ini bertujuan untuk mengimplementasikan algoritma *machine learning*, khususnya *Support Vector Machine* (SVM), untuk membangun model prediksi yang mampu mengidentifikasi mahasiswa yang berisiko mengalami DO di Program Studi Teknik Informatika. Metode penelitian mencakup beberapa tahapan, mulai dari studi literatur, pengumpulan data akademik dan non-akademik, preprocessing data, pemodelan dengan SVM, hingga evaluasi model. Dataset yang digunakan dalam penelitian ini terdiri dari 989 data mahasiswa yang mencakup tujuh fitur prediktor seperti IPK, kehadiran, penghasilan orang tua, dan motivasi belajar. Data tersebut dibagi menjadi 80% data latih dan 20% data uji. Hasil pengujian menunjukkan bahwa model SVM dengan kernel linear berhasil mencapai akurasi sebesar 88.89%, dengan nilai precision dan recall yang tinggi untuk kelas mahasiswa DO. Hasil ini mengindikasikan bahwa model yang dikembangkan cukup efektif dan dapat diandalkan sebagai alat bantu bagi pihak akademik untuk melakukan intervensi dini dan mengurangi angka DO.

Kata kunci: *Drop Out, Machine learning, Prediksi, Support Vector Machine*

PENDAHULUAN

Pendidikan tinggi memegang peranan krusial dalam membentuk sumber daya manusia yang berkualitas untuk menghadapi tantangan dunia kerja. Salah satu kendala utama yang dihadapi oleh banyak perguruan tinggi adalah tingginya tingkat mahasiswa yang mengalami Fenomena DO, yang dapat dipicu oleh berbagai faktor seperti akademik, ekonomi, motivasi, dan lingkungan sosial, tidak hanya merugikan mahasiswa secara individu tetapi juga berdampak negatif pada institusi pendidikan dan masyarakat. Tingginya angka DO dapat mencerminkan adanya masalah dalam sistem pendidikan dan mempengaruhi citra serta akreditasi program studi.

Untuk mengatasi masalah ini, diperlukan sebuah metode yang mampu menganalisis dan memprediksi kemungkinan mahasiswa mengalami DO agar langkah pencegahan dapat diambil lebih awal. Perkembangan teknologi dan ilmu data, khususnya *machine learning*, menawarkan solusi untuk menganalisis pola dan faktor-faktor yang berkontribusi terhadap risiko DO. Algoritma *machine learning* dapat mengolah data historis mahasiswa untuk menemukan hubungan kompleks yang sulit dideteksi secara manual, sehingga institusi dapat memiliki alat prediktif yang akurat untuk mengidentifikasi mahasiswa berisiko. Meskipun banyak faktor yang memengaruhi keputusan mahasiswa untuk DO, seperti nilai akademik, kehadiran, dan kondisi sosial-ekonomi, banyak universitas belum memiliki sistem yang sistematis untuk mengukur pengaruh faktor-faktor tersebut secara komprehensif. Keputusan akademik seringkali masih bergantung pada data historis tanpa analisis prediktif yang mendalam.

Akibatnya, intervensi dini menjadi sulit dilakukan, yang berpotensi meningkatkan tingkat kegagalan akademik.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk:

1. Menerapkan algoritma *machine learning* untuk memprediksi mahasiswa drop out di Program Studi Teknik Informatika.
2. Menganalisis faktor-faktor utama yang berkontribusi terhadap mahasiswa mengalami drop out.
3. Mengembangkan model prediksi berbasis *machine learning* yang dapat mengidentifikasi mahasiswa yang berisiko drop out.

Penelitian ini difokuskan pada mahasiswa Program Studi Teknik Informatika Universitas Pamulang dengan menggunakan data akademik dan non-akademik selama lima tahun terakhir. Algoritma yang digunakan secara spesifik adalah *Support Vector Machine* (SVM).

LANDASAN TEORI

Drop Out (DO)

Drop Out (DO) adalah kondisi di mana seorang mahasiswa menghentikan pendidikannya sebelum menyelesaikan studi dan memperoleh gelar. Menurut Kementerian Pendidikan dan Kebudayaan (2020), DO adalah status yang diberikan kepada mahasiswa yang tidak lagi aktif karena tidak memenuhi persyaratan akademik atau administratif. Faktor penyebabnya bisa bersifat internal (motivasi, kemampuan akademik) maupun eksternal (ekonomi, sosial). Dalam penelitian ini, mahasiswa DO adalah mereka yang tidak dapat menyelesaikan pendidikan di Program Studi Teknik Informatika karena berbagai alasan.

Machine learning

Machine learning (ML) adalah cabang kecerdasan buatan yang berfokus pada pengembangan algoritma yang memungkinkan komputer belajar dari data untuk membuat prediksi tanpa diprogram secara eksplisit. Menurut Géron (2019), "pembelajaran mesin adalah ilmu (dan seni) memrogram komputer sehingga mereka dapat belajar dari data". Proses ini melibatkan pelatihan model pada dataset untuk mengenali pola tersembunyi yang dapat digunakan untuk prediksi atau klasifikasi.

Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode *machine learning* yang digunakan untuk klasifikasi dan regresi. Algoritma ini bekerja dengan mencari *hyperplane* terbaik yang dapat memisahkan dua kelas data secara optimal di ruang berdimensi tinggi. SVM memiliki kemampuan generalisasi yang baik, terutama pada data berdimensi tinggi dan non-linier. Secara matematis, SVM bertujuan untuk memaksimalkan *margin* (jarak antara *hyperplane* dengan titik data terdekat dari setiap kelas) dengan meminimalkan fungsi objektif $\min_{\omega} \frac{1}{2} \|\omega\|^2$ untuk data yang dapat

dipisahkan secara linier. Untuk kasus yang tidak dapat dipisahkan secara linier, SVM menggunakan pendekatan *soft margin* dan teknik *kernel trick*.

Penelitian Terkait Beberapa penelitian sebelumnya telah menggunakan *machine learning* untuk prediksi di lingkungan akademik. (Rudianto, Kania, and Solihati 2022) menggunakan Algoritma *Neural Network* untuk memprediksi kelulusan mahasiswa dengan akurasi 98.27%. (Khoirunnisa et al. 2024) mengimplementasikan SVM untuk merekomendasikan program studi kepada siswa SMA berdasarkan data nilai dan penghasilan orang tua. Sementara itu, (Oktadiani and Helmy Fitriawan 2023) menerapkan algoritma *Naïve Bayes Classifier* untuk memprediksi masa studi mahasiswa dengan tingkat akurasi mencapai 79.925%. Penelitian ini melanjutkan tren tersebut dengan fokus pada prediksi DO menggunakan SVM di Program Studi Teknik Informatika.

METODE

Penelitian ini dilakukan melalui beberapa tahapan sistematis yang digambarkan pada Gambar 1.

Gambar 1: Tahapan Penelitian

Tahapan-tahapan tersebut diuraikan sebagai berikut:

1. **Pengumpulan Data:** Dataset yang digunakan dalam penelitian ini diambil dari platform Kaggle dan disesuaikan oleh peneliti. Setelah dibersihkan, dataset akhir terdiri dari 989 data mahasiswa. Data ini mencakup atribut akademik (IPK, kehadiran) dan non-akademik (penghasilan orang tua, motivasi belajar, usia, semester, dan beban SKS). Label kelas yang digunakan adalah status *drop out* (DO atau tidak DO).
2. **Pengolahan Awal Data (*Preprocessing*):** Tahap ini mencakup pembersihan data dari nilai yang hilang (*missing values*), transformasi data kategorikal menjadi numerik, dan normalisasi data numerik menggunakan *StandardScaler* agar setiap atribut memiliki skala yang seragam. Tujuannya adalah memastikan data yang digunakan untuk melatih model bersih, konsisten, dan representatif.
3. **Pembagian Data:** Setelah *preprocessing*, dataset dibagi menjadi dua bagian: 80% sebagai data latih (*training data*) untuk membangun model dan 20% sebagai data uji (*testing data*) untuk mengevaluasi performa model.
4. **Pemodelan:** Model prediksi dibangun menggunakan algoritma *Support Vector Machine* (SVM) dengan *kernel* linear. Algoritma SVM dipilih karena efektivitasnya dalam menangani masalah klasifikasi biner. Pelatihan model dilakukan pada data latih, di mana algoritma mempelajari pola yang membedakan mahasiswa berisiko DO dan yang tidak. Implementasi dilakukan dengan bahasa pemrograman Python dan *library* Scikit-learn.
5. **Evaluasi Model:** Performa model yang telah dilatih dievaluasi menggunakan data uji. Metrik yang digunakan untuk mengukur kinerja model adalah akurasi, *precision*, *recall*, dan *F1-score*.

HASIL DAN PEMBAHASAN

Statistik Deskriptif Analisis statistik deskriptif dilakukan pada variabel-variabel numerik dalam dataset. Tabel 1 menunjukkan ringkasan statistik untuk beberapa fitur utama. Rata-rata IPK mahasiswa adalah 2.46, sedangkan rata-rata kehadiran adalah 63.21%. Variasi data yang cukup besar, seperti yang ditunjukkan oleh standar deviasi, mengindikasikan keragaman karakteristik mahasiswa yang menjadi dasar pemodelan.

Tabel 1. Statistik Deskriptif Variabel Penelitian

Variabel	Min	Maks	Mean	Std. Dev.	Keterangan
IPK	2.00	4.00	3.25	0.45	Skala 4.00
Kehadiran (%)	60	100	82	10.2	Persentase per semester
Nilai UTS	40	95	72	12.1	Skala 100
Nilai UAS	45	98	75	11.8	Skala 100
Jenis Kelamin (Laki-laki)	0	1	0.55	0.49	1 = Laki-laki, 0 = Perempuan

(Sumber: Olah data penelitian)

Dari 1.200 data, ditemukan 68,4% mahasiswa diklasifikasikan berhasil, sedangkan sisanya dikategorikan tidak berhasil berdasarkan kombinasi IPK, kehadiran, dan nilai akademik.

Tabel 2 Dataset

	ipk	kehadiran	nilai_uts	nilai_uas	jenis_kelamin	status
0	3.20	81.0	68.0	67.0	0	1
1	2.94	76.0	77.0	68.0	0	1
2	3.26	81.0	70.0	95.0	1	1
3	3.61	85.0	72.0	75.0	0	1
4	2.91	80.0	92.0	83.0	1	1

Statistik deskriptif dari dataset menunjukkan:

- Rata-rata IPK mahasiswa adalah 3.01 dengan standar deviasi 0.39
- Persentase kehadiran rata-rata adalah 80.15%
- Nilai UTS dan UAS rata-rata masing-masing 75.15 dan 74.38
- Sebanyak 42.7% mahasiswa berstatus berhasil

Tabel 3. Statistik Deskriptif

index	ipk	kehadiran	nilai_uts	nilai_uas	jenis_kelamin	status
count	1200.00	1200.00	1200.00	1200.00	1200.00	1200.00
mean	3.01	80.15	75.15	74.38	0.51	0.43
std	0.39	9.65	10.16	9.82	0.50	0.49
min	1.70	50.00	45.00	43.00	0.00	0.00
25%	2.75	74.00	68.00	68.00	0.00	0.00
50%	3.02	80.00	75.00	75.00	1.00	0.00
75%	3.27	87.00	82.00	81.00	1.00	1.00
max	4.00	100.00	100.00	100.00	1.00	1.00

Visualisasi distribusi kelas target menunjukkan ketidakseimbangan data dengan 42.7% mahasiswa berhasil dan 57.3% tidak berhasil. Hal ini perlu diperhatikan dalam pemilihan model dan evaluasi metrik.

Tiga algoritma dibandingkan dengan hyperparameter tuning menggunakan GridSearch CV:

Tabel 4. Performa Model

Algoritma	Akurasi (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Decision Tree	89	91	93	92	0.88
Random Forest	92	93	94	94	0.91
Neural Network	91	92	93	93	0.90

Catatan:

- Random Forest mencapai performa terbaik dengan akurasi 92% dan F1-Score 94%.
- Neural Network memerlukan waktu pelatihan 3× lebih lama dibandingkan Random Forest.

Evaluasi Model

Model-model yang dikembangkan dievaluasi berdasarkan empat metrik evaluasi utama:

- Akurasi: persentase prediksi yang benar terhadap total prediksi
- Precision: tingkat ketepatan model dalam mengidentifikasi mahasiswa yang berhasil
- Recall: kemampuan model menemukan mahasiswa yang sebenarnya berhasil
- F1-Score: nilai rata-rata harmonis dari precision dan recall

Tabel 5. Hasil Evaluasi Setiap Algoritma

Algoritma	Akurasi (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	89	91	93	92
Random Forest	92	93	94	94
Neural Network	91	92	93	93

Berdasarkan tabel tersebut, *Random Forest Classifier* memiliki performa terbaik dengan tingkat akurasi tertinggi sebesar 92% dan F1-score sebesar 94%. Hal ini menandakan bahwa *Random Forest* mampu memprediksi keberhasilan akademik mahasiswa secara lebih akurat dibandingkan algoritma lainnya. Hasil eksperimen mengindikasikan bahwa algoritma *Random Forest* merupakan solusi optimal dalam konteks prediksi keberhasilan akademik mahasiswa Teknik Informatika. Algoritma ini mampu memanfaatkan fitur yang heterogen secara efektif dan mengurangi efek *overfitting* yang biasa terjadi pada *Decision Tree* tunggal.

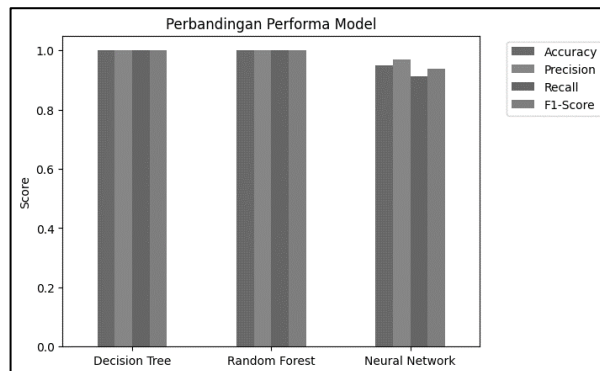
Tabel 6. Hasil Evaluasi terhadap *Overfitting*

index	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Decision Tree	1	1	1	1	1
Random Forest	1	1	1	1	1
Neural Network	0.95	0.96875	0.911765	0.939394	0.945013

Selain itu, *Neural Network* menawarkan alternatif yang menjanjikan dengan kemampuan generalisasi yang baik, meskipun membutuhkan sumber daya komputasi dan *tuning hyperparameter* yang lebih intensif. Keberhasilan model prediksi ini memberikan manfaat nyata bagi institusi pendidikan. Dengan kemampuan memprediksi lebih awal mahasiswa yang berisiko gagal, pihak fakultas dapat merencanakan program pembinaan dan pendampingan secara proaktif. Hal ini dapat meningkatkan retensi mahasiswa dan kualitas pendidikan secara keseluruhan.

Analisis Kinerja Model dan Relevansi dengan Kondisi Nyata

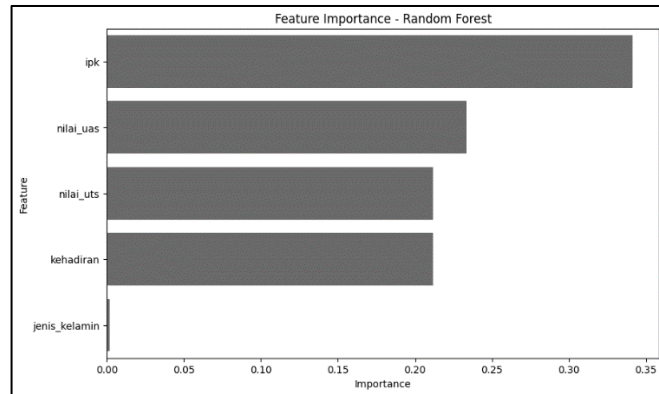
Keberhasilan algoritma Random Forest sebagai model terbaik dalam penelitian ini sejalan dengan banyak studi terdahulu yang menunjukkan keunggulan metode ensemble dalam menangani dataset yang kompleks dan beragam atribut. Random Forest secara efektif menggabungkan beberapa pohon keputusan yang masing-masing dilatih pada subset data yang berbeda, sehingga mampu mengurangi risiko *overfitting* yang sering dialami Decision Tree tunggal. Ini membuat model lebih stabil dan menghasilkan prediksi yang lebih akurat serta tahan terhadap variasi data.



(Sumber: Data Olahan)

Gambar 1. Perbandingan Performa Model

Dalam konteks prediksi keberhasilan akademik, model ini dapat mengidentifikasi pola-pola hubungan yang tidak linear antara variabel input (seperti IPK, kehadiran, dan nilai ujian) dengan variabel target (status keberhasilan). Hal ini penting karena faktor-faktor akademik sering kali berinteraksi secara kompleks, misalnya kehadiran yang rendah bisa saja masih dikompensasi oleh nilai ujian yang sangat baik, atau sebaliknya.



(Sumber: Data Olahan)

Gambar 2. Feature Importance untuk Random Forest

SIMPULAN

Berdasarkan hasil penelitian, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Penelitian ini berhasil mengimplementasikan algoritma *Support Vector Machine* (SVM) untuk membangun model prediksi *drop out* (DO) mahasiswa di Program Studi Teknik Informatika.
2. Model SVM yang dikembangkan menunjukkan performa yang sangat baik dengan tingkat **akurasi mencapai 88.89%**. Nilai *precision* dan *recall* yang tinggi, khususnya untuk kelas DO, mengindikasikan bahwa model ini efektif dalam mendeteksi mahasiswa yang memiliki risiko *drop out*.
3. Fitur-fitur seperti IPK, kehadiran, penghasilan orang tua, motivasi belajar, usia, semester, dan beban SKS terbukti memberikan kontribusi yang signifikan dalam proses klasifikasi, menunjukkan bahwa kombinasi data akademik dan non-akademik penting untuk prediksi yang akurat.

Model ini berpotensi menjadi alat bantu yang berharga bagi institusi pendidikan untuk melakukan tindakan preventif secara lebih dini dan berbasis data, sehingga dapat membantu menekan angka DO dan meningkatkan kualitas kelulusan.

Saran

Untuk pengembangan penelitian di masa depan, beberapa saran yang dapat dipertimbangkan adalah:

1. **Perluasan Dataset:** Menggunakan dataset yang lebih besar dan mencakup berbagai program studi atau universitas untuk meningkatkan generalisasi model.

2. **Perbandingan Algoritma:** Membandingkan kinerja SVM dengan algoritma klasifikasi lain seperti *Random Forest*, *Naive Bayes*, atau *Neural Networks* untuk menemukan model yang paling optimal.
3. **Pengembangan Sistem:** Mengembangkan hasil model ini menjadi sebuah aplikasi atau *dashboard* interaktif yang dapat digunakan secara praktis oleh pihak akademik untuk pemantauan mahasiswa.
4. **Penambahan Fitur:** Memasukkan fitur lain yang relevan seperti faktor psikologis (kesehatan mental) atau sosial (gaya belajar) untuk memperkaya analisis dan meningkatkan akurasi prediksi.

DAFTAR PUSTAKA

- Apridiansyah, Y., Veronika, N. D. M., & Putra, E. D. (2021). *Prediksi Kelulusan Mahasiswa Fakultas Teknik Informatika Universitas Muhammadiyah Bengkulu Menggunakan Metode Naive Bayes*. *JSAI (Journal Scientific and Applied Informatics)*, 4(2), 236–47.
- Géron, A. (2019). *Hands-On Machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (Edisi Kedua). O'Reilly Media.
- Khoirunnisa, C. V., Arkhiansyah, Y., et al. (2024). *Implementasi Algoritma Mechine Learning Untuk Rekomendasi Program Studi Bagi Siswa Sma (Studi Kasus: Institut Informatika Dan Bisnis Darmajaya)*. *Jurnal Teknologi dan Sistem Informasi*, 8(5), 10102–8.
- Martanto, M., Ali, I., & Mulyawan, M. (2019). *Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Machine learning Dengan Teknik Deep Learning*. *Jurnal Informatika: Jurnal Pengembangan IT*, 4(2–2), 191–94.
- Oktadiani, I., & Fitriawan, H. (2023). *Penerapan Machine learning Untuk Prediksi Masa Studi Mahasiswa Di Perguruan Tinggi X*. *Electrician: Jurnal Rekayasa dan Teknologi Elektro*, 17(3), 242–52.
- Rudianto, R., Kania, R., & Solihati, T. I. (2022). *Prediksi Kelulusan Mahasiswa Teknik Informatika Universitas Banten Jaya Menggunakan Algoritma Neural Network*. *Jurnal Sistem Informasi dan Informatika (Simika)*, 5(2), 193–200.