

## ANALISIS PERBANDINGAN ALGORITMA SUPPORT VECTOR MACHINE DAN DECISION TREE UNTUK PREDIKSI STATUS KELULUSAN MAHASISWA BERDASARKAN IPK

Fito Mardianto<sup>1</sup>, Aziz Setyawan Hidayat<sup>2</sup>, Ade Priyatna<sup>3</sup>  
[fitomardianto15@gmail.com](mailto:fitomardianto15@gmail.com)<sup>1</sup>, [aziz.aiz@bsi.ac.id](mailto:aziz.aiz@bsi.ac.id)<sup>2</sup>, [ade.aeg@bsi.ac.id](mailto:ade.aeg@bsi.ac.id)<sup>3</sup>  
Universitas Bina sarana Informatika

### ABSTRAK

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja dua algoritma machine learning, yaitu Support Vector Machine (SVM) dan Decision Tree, dalam memprediksi status kelulusan mahasiswa berdasarkan Indeks Prestasi Kumulatif (IPK) serta pemenuhan persyaratan akademik seperti sertifikat TOEFL, BNSP, dan vendor. Dataset yang digunakan bersumber dari platform Kaggle dan dianalisis menggunakan perangkat lunak Orange. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, F1-score, AUC, dan MCC. Hasilnya menunjukkan bahwa Decision Tree memiliki akurasi 100% sementara SVM memiliki akurasi 99,7% dengan satu kesalahan klasifikasi. Meskipun Decision Tree lebih akurat dalam dataset ini, algoritma SVM menunjukkan kestabilan yang lebih baik terhadap data baru. Penelitian ini diharapkan menjadi acuan dalam pengembangan sistem prediktif akademik berbasis data.

**Kata Kunci:** Support Vector Machine, Decision Tree, Ipk, Prediksi Kelulusan, Orange Data Mining.

### ABSTRACT

*This study aims to analyze and compare the performance of two machine learning algorithms, Support Vector Machine (SVM) and Decision Tree, in predicting student graduation status based on GPA and academic requirements such as TOEFL, BNSP, and vendor certifications. The dataset used was obtained from Kaggle and analyzed using Orange software. Evaluation metrics include accuracy, precision, recall, F1-score, AUC, and MCC. The results showed that Decision Tree achieved 100% accuracy while SVM achieved 99.7% with only one misclassification. Although Decision Tree yielded perfect results, it is more prone to overfitting compared to the more generalizable SVM. This research is expected to serve as a reference for the development of data-driven academic predictive systems.*

**Keywords:** Support Vector Machine, Decision Tree, Gpa, Graduation Prediction, Orange Data Mining

### PENDAHULUAN

Prediksi kelulusan mahasiswa merupakan tantangan strategis dalam pengelolaan pendidikan tinggi di era digital. Setiap institusi pendidikan dituntut tidak hanya untuk meningkatkan mutu pembelajaran, tetapi juga memastikan mahasiswa menyelesaikan studi tepat waktu. Ketepatan waktu kelulusan menjadi salah satu indikator penting dalam penilaian akreditasi oleh lembaga seperti BAN-PT. Oleh karena itu, adanya sistem yang dapat memprediksi kelulusan mahasiswa secara akurat sejak dini sangat dibutuhkan.

Salah satu indikator utama dalam mengukur performa akademik mahasiswa adalah Indeks Prestasi Kumulatif (IPK). IPK mencerminkan konsistensi prestasi akademik mahasiswa selama masa studi. Namun, IPK tidak berdiri sendiri. Banyak faktor lain yang turut mempengaruhi kelulusan, seperti kepemilikan sertifikat akademik (TOEFL, BNSP, dan vendor), keaktifan dalam organisasi, kehadiran, hingga latar belakang sosial ekonomi. Oleh karena itu, perlu pendekatan yang komprehensif dan sistematis untuk menganalisis keterkaitan antar faktor-faktor tersebut.

Dalam beberapa tahun terakhir, perkembangan teknologi artificial intelligence (AI), khususnya machine learning, telah membuka peluang baru dalam analisis prediktif di bidang pendidikan. Metode machine learning mampu mengenali pola tersembunyi dalam data dan memberikan prediksi yang akurat berdasarkan data historis. Dua algoritma yang paling banyak digunakan dalam klasifikasi prediksi adalah Support Vector Machine (SVM) dan Decision Tree. Keduanya memiliki kelebihan masing-masing dalam hal akurasi, interpretabilitas, dan kecepatan pemrosesan.

Meskipun banyak penelitian telah dilakukan menggunakan salah satu dari kedua metode tersebut, masih sedikit yang secara langsung membandingkan performa SVM dan Decision Tree dalam konteks prediksi kelulusan mahasiswa berbasis IPK dan sertifikat. Penelitian ini hadir untuk mengisi celah tersebut, dengan membangun model klasifikasi berdasarkan data akademik dan mengevaluasi performanya secara menyeluruh menggunakan metrik evaluasi seperti akurasi, precision, recall, F1-score, AUC, dan MCC.

Dengan demikian, hasil penelitian ini diharapkan dapat menjadi referensi dalam pengembangan sistem akademik berbasis data untuk mendukung pengambilan keputusan strategis oleh institusi pendidikan tinggi. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi terhadap pengembangan literatur dalam bidang data mining pendidikan, khususnya yang berbasis pembelajaran mesin.

## **METODE PENELITIAN**

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi machine learning. Dataset yang digunakan berasal dari Kaggle dan mencakup data IPK mahasiswa dari semester 1 hingga 8, status kelulusan, serta kepemilikan sertifikat akademik (TOEFL, BNSP, dan vendor). Data diproses menggunakan software Orange Data Mining. Pra-pemrosesan meliputi normalisasi, pengelompokan status kelulusan menjadi dua kelas (LULUS dan TIDAK LULUS), dan penghapusan data yang tidak lengkap.

Proses modeling dilakukan dengan membagi data menjadi 80% data latih dan 20% data uji. Model dikembangkan menggunakan algoritma Support Vector Machine (SVM) dan Decision Tree. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, F1-score, AUC, dan MCC dengan 10-fold cross-validation.

### **1. Dataset yang digunakan**

berasal dari platform publik Kaggle, yang menyediakan data mahasiswa dengan sejumlah atribut sebagai berikut:

- Nilai IPK semester 1 hingga semester 8,
- Status kelulusan,
- Kepemilikan sertifikat TOEFL, BNSP, dan vendor.

Data ini dipilih karena memiliki karakteristik serupa dengan sistem pendidikan di Indonesia dan relevan dengan tujuan penelitian.

### **2. Pra-pemrosesan Data**

Langkah pertama dalam penelitian ini adalah melakukan proses pra-pemrosesan untuk memastikan kualitas data sebelum digunakan dalam pelatihan model. Tahapan ini meliputi:

- Normalisasi data IPK agar berada dalam rentang umum 0.00–4.00.
- Pengelompokan ulang status kelulusan menjadi dua kelas yaitu LULUS (status A, B, C) dan TIDAK LULUS (status D).
- Penghapusan data tidak lengkap yang memiliki missing values atau outlier yang tidak wajar.
- Encoding data kategorikal, terutama sertifikat, ke dalam bentuk numerik biner (0 = tidak memiliki, 1 = memiliki).

### 3. Algoritma yang Digunakan

Dua algoritma utama dalam penelitian ini adalah:

- Support Vector Machine (SVM)

SVM digunakan untuk membangun hyperplane yang secara optimal memisahkan dua kelas, yaitu mahasiswa yang lulus dan tidak lulus. Kernel default (linear) digunakan dalam eksperimen awal.

- Decision Tree

Algoritma ini membentuk pohon keputusan berdasarkan atribut input, seperti IPK dan kepemilikan sertifikat. Setiap simpul pohon mewakili keputusan logis berdasarkan nilai tertentu dari atribut.

## HASIL DAN PEMBAHASAN

Hasil evaluasi menunjukkan bahwa model Decision Tree mampu mengklasifikasikan status kelulusan dengan akurasi 100%. Seluruh data pada data uji berhasil diklasifikasikan tanpa kesalahan. Di sisi lain, model SVM menghasilkan akurasi sebesar 99,7%, dengan satu kesalahan klasifikasi pada data mahasiswa yang sebenarnya lulus namun diprediksi tidak lulus.

Performa model Decision Tree sangat tinggi namun berisiko mengalami overfitting karena terlalu sesuai dengan data pelatihan. Sementara itu, SVM menunjukkan kemampuan generalisasi yang lebih baik meskipun sedikit kurang akurat. Hal ini penting dalam konteks implementasi model prediktif pada data baru yang tidak termasuk dalam pelatihan awal.

### 1. Hasil Pengujian Algoritma Decision Tree

Model klasifikasi menggunakan algoritma Decision Tree menunjukkan performa yang sangat tinggi, dengan hasil sebagai berikut:

#### Metrik Evaluasi Nilai Decision Tree

#### Metrik Evaluasi Nilai Decision Tree

Akurasi (CA)	100%
Precision	1.000
Recall	1.000
F1-score	1.000
AUC	1.000
MCC	1.000

Seluruh data uji diklasifikasikan secara sempurna tanpa kesalahan. Keunggulan utama dari Decision Tree terletak pada interpretasi hasil yang mudah dipahami oleh pengguna non-teknis, seperti manajemen kampus atau dosen pembimbing. Visualisasi pohon keputusan juga membantu dalam memahami aturan klasifikasi yang terbentuk secara otomatis dari data.

Namun, hasil dengan nilai sempurna ini juga menunjukkan potensi overfitting. Model kemungkinan besar terlalu menyesuaikan diri terhadap data latih dan mungkin tidak memberikan hasil sebaik ini jika digunakan pada data baru dari luar dataset.

## 2. Hasil Pengujian Algoritma Support Vector Machine (SVM)

Model SVM menghasilkan performa yang hampir setara, dengan hasil evaluasi sebagai berikut:

Metrik Evaluasi	Nilai SVM
Akurasi (CA)	99.7%
Precision	0.997
Recall	0.997
F1-score	0.997
AUC	1.000
MCC	0.993

Terdapat satu kesalahan klasifikasi pada model ini, yaitu data mahasiswa yang sebenarnya lulus namun diprediksi tidak lulus. Meski begitu, model SVM menunjukkan kemampuan generalisasi yang lebih baik dibanding Decision Tree. Dengan margin optimal dan penggunaan hyperplane, SVM tetap stabil meskipun data memiliki dimensi yang kompleks.

## 3. Analisis Komparatif

Tabel berikut merangkum perbandingan kedua algoritma:

Aspek Evaluasi	Decision Tree	SVM
Akurasi	100%	99.7%
Kemampuan Umum	Kurang stabil	Stabil
Interpretasi	Sangat mudah	Kompleks
Overfitting	Tinggi	Rendah
Waktu Training	Cepat	Sedang

### Interpretabilitas:

Decision Tree unggul karena hasil klasifikasinya dapat divisualisasikan dan dijelaskan secara logis. Hal ini sangat berguna bagi pengguna non-teknis.

### Generalitas:

SVM memiliki margin klasifikasi yang optimal, menjadikannya lebih tahan terhadap perubahan data atau outlier dibanding Decision Tree yang kaku terhadap pola.

### Stabilitas Model:

Meskipun akurasi Decision Tree lebih tinggi, kestabilan SVM terhadap data baru menjadi nilai tambah dalam konteks implementasi jangka panjang.

#### **4. Implikasi Hasil**

Hasil penelitian menunjukkan bahwa kedua algoritma sangat layak digunakan untuk pengembangan sistem pendukung keputusan akademik. Pilihan algoritma dapat disesuaikan dengan kebutuhan:

- Jika diperlukan interpretasi yang jelas dan cepat, Decision Tree lebih disarankan.
- Jika dibutuhkan akurasi tinggi dan kestabilan terhadap variasi data, maka SVM menjadi pilihan utama.

Implementasi dari model ini dapat diintegrasikan dalam Sistem Informasi Akademik (SIKAD) untuk membantu pihak kampus mendeteksi lebih awal mahasiswa yang berpotensi mengalami keterlambatan kelulusan dan memberikan intervensi yang sesuai.

#### **KESIMPULAN**

Penelitian ini menunjukkan bahwa algoritma Support Vector Machine (SVM) dan Decision Tree dapat digunakan secara efektif untuk memprediksi status kelulusan mahasiswa berdasarkan data akademik seperti Indeks Prestasi Kumulatif (IPK) dan kepemilikan sertifikat pendukung (TOEFL, BNSP, dan vendor). Hasil eksperimen menunjukkan bahwa Decision Tree menghasilkan akurasi klasifikasi sempurna sebesar 100%, sedangkan SVM mencapai akurasi 99,7% dengan hanya satu kesalahan klasifikasi. Meskipun demikian, model SVM memiliki keunggulan dalam hal kemampuan generalisasi dan kestabilan terhadap data baru, sementara Decision Tree unggul dalam hal interpretabilitas dan transparansi model.

Perbedaan kinerja antara kedua algoritma mengindikasikan bahwa pemilihan metode klasifikasi sebaiknya disesuaikan dengan kebutuhan dan konteks implementasi. Untuk sistem yang membutuhkan penjelasan logika klasifikasi secara eksplisit kepada pengguna non-teknis, seperti dosen atau pengambil keputusan akademik, Decision Tree merupakan pilihan yang tepat. Sebaliknya, jika fokus utama adalah ketepatan klasifikasi dan generalisasi pada data skala besar yang dinamis, maka SVM menjadi alternatif unggulan.

Berdasarkan hasil dan pembahasan, penelitian ini merekomendasikan agar institusi pendidikan tinggi mulai mempertimbangkan integrasi model prediksi kelulusan ke dalam sistem informasi akademik. Dengan demikian, potensi mahasiswa yang mengalami hambatan kelulusan dapat terdeteksi lebih awal dan ditindaklanjuti melalui program pendampingan atau intervensi akademik yang tepat. Selain itu, pengembangan sistem prediktif ini dapat diperluas dengan memasukkan variabel lain seperti kehadiran, partisipasi organisasi, latar belakang sosial ekonomi, dan hasil asesmen kepribadian untuk meningkatkan akurasi dan cakupan analisis. Kombinasi atau ensemble dari berbagai algoritma juga dapat menjadi pendekatan selanjutnya guna memperoleh hasil prediksi yang lebih optimal dan adaptif terhadap perubahan data.

#### **Ucapan Terima Kasih**

Penulis mengucapkan terima kasih kepada dosen pembimbing Bapak Aziz Setyawan Hidayat, S.Kom., M.Kom dan Bapak Ade Priyatna, S.Kom., M.M., M.Kom atas bimbingan dan arahnya selama proses penyusunan penelitian ini. Penulis juga berterima kasih kepada Universitas Bina Sarana Informatika dan semua pihak yang telah memberikan dukungan moral dan teknis dalam penyelesaian karya ini.

#### **DAFTAR PUSTAKA**

- Aditya, M. F., et al. (2024). Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2. *MALCOM*, 4(3), 1104–1110.
- Haryatmi, E., & Hervianti, S. P. (2021). Penerapan SVM Untuk Model Prediksi Kelulusan Mahasiswa. *Jurnal RESTI*, 5(2), 386–392.
- Lestari, G. C., & Supatmi, S. (2023). Analisis Prediksi Kelulusan Course Pada E-Learning. *JTK3TI*,

9(2), 79–85.

Pranadjaya, E., et al. (2024). Perbandingan Algoritma Machine learning menggunakan Orange. *Jurnal Elektro*, 17(1), 41–47.

Septhya, D., et al. (2023). Algoritma Decision Tree dan SVM untuk Klasifikasi Penyakit Kanker Paru. *MALCOM*, 3(1), 15–19.