

## Penerapan Metode *Support Vector Machine* Untuk Memprediksi Kelulusan Tepat Waktu

DOI: <http://dx.doi.org/10.35889/progresif.v21i2.2619>

Creative Commons License 4.0 (CC BY –NC)



**Avif Setyawan<sup>1\*</sup>, Rahmat Haryadi Kiswanto<sup>2</sup>, Heru Sutejo<sup>3</sup>**

Teknik Informatika, Universitas Sepuluh Nopember Papua, Jayapura, Indonesia

\*e-mail *Corresponding Author*: [avifsetyawan3@gmail.com](mailto:avifsetyawan3@gmail.com)

### Abstract

*Timely graduation is a key indicator of student academic success in higher education. This study develops a predictive model for on-time student graduation at Universitas Sepuluh Nopember Papua using the Support Vector Machine (SVM) method with a linear kernel. The model examines the influence of academic performance and student status as Indigenous Papuans (OAP) or non-OAP in predicting graduation probability. Model evaluation was conducted using Confusion Matrix, ROC Curve, and Cross-Validation, demonstrating that the model achieved high accuracy of 92% in the initial testing phase, increasing to 97% after cross-validation. The evaluation also showed a Precision of 90%, Recall of 100%, and F1-Score of 95%, confirming the model's effectiveness in distinguishing students at risk of delayed graduation. With its high predictive accuracy, this model can serve as a data-driven academic decision-making tool to identify at-risk students and implement more targeted academic interventions to improve timely graduation rates.*

**Keywords:** *graduation prediction; model evaluation; machine learning; Support Vector Machine.*

### Abstrak

Kelulusan tepat waktu menjadi indikator utama keberhasilan akademik mahasiswa di perguruan tinggi. Penelitian ini mengembangkan model prediksi kelulusan tepat waktu mahasiswa Universitas Sepuluh Nopember Papua menggunakan metode *Support Vector Machine* (SVM) dengan kernel linier. Model ini menganalisis pengaruh kinerja akademik dan status mahasiswa sebagai Orang Asli Papua (OAP) atau non-OAP dalam menentukan probabilitas kelulusan tepat waktu. Evaluasi model dilakukan menggunakan *Confusion Matrix*, *ROC Curve*, dan *Cross-Validation*, yang menunjukkan bahwa model memiliki akurasi tinggi sebesar 92% pada tahap pengujian awal dan meningkat menjadi 97% setelah validasi silang. Hasil pengujian juga menunjukkan nilai *Precision* 90%, *Recall* 100%, dan *F1-Score* 95%, yang menegaskan efektivitas model dalam membedakan mahasiswa yang berisiko mengalami keterlambatan kelulusan. Dengan tingkat akurasi yang tinggi, model ini dapat digunakan sebagai alat bantu akademik berbasis data untuk mengidentifikasi mahasiswa berisiko dan menerapkan intervensi akademik yang lebih tepat sasaran guna meningkatkan tingkat kelulusan tepat waktu.

**Kata kunci:** *Prediksi kelulusan; Evaluasi model; Machine learning; Support Vector Machine.*

### 1. Pendahuluan

Kelulusan tepat waktu merupakan salah satu indikator keberhasilan pendidikan tinggi yang sangat penting[1] dan menjadi perhatian utama di Universitas Sepuluh Nopember. Perguruan tinggi harus mengantisipasi efek negatif yang akan terjadi pada kelulusan mahasiswa. Penggunaan teknologi sistem informasi yang canggih sangat penting untuk mendukung upaya ini[2]. Perkembangan ini memungkinkan perguruan tinggi untuk menganalisis data secara lebih efisien dan akurat, yang memungkinkan untuk mendapatkan pemahaman yang lebih mendalam tentang faktor-faktor yang mempengaruhi kelulusan mahasiswa[3]. Jika terjadi penurunan tingkat kelulusan maka akan menjadi suatu permasalahan yang akan mempengaruhi akreditasi di sebuah perguruan tinggi[4]. Oleh karena itu, institusi dapat menerapkan pendekatan yang

lebih tepat sasaran untuk meningkatkan tingkat kelulusan dan mempertahankan standar akreditasi yang tinggi[5].

Permasalahan kelulusan tepat waktu dan tidak tepat waktu di Universitas Sepuluh Nopember (USN) menunjukkan kompleksitas interaksi antara berbagai faktor dalam konteks pendidikan yang lebih luas. Data empiris menunjukkan bahwa *persentase* kelulusan tepat waktu dari tahun 2018 hingga 2020 berada jauh di bawah standar yang ditetapkan oleh Badan Akreditasi Nasional Perguruan Tinggi (BAN-PT). Pada tahun 2018, hanya 9% dari 90 mahasiswa yang berhasil lulus tepat waktu, sedangkan 91% lainnya mengalami keterlambatan. Kondisi ini berlanjut pada tahun 2019, di mana dari 91 mahasiswa, hanya 13% yang lulus tepat waktu, dan meningkat menjadi 19% dari 78 mahasiswa pada tahun 2020. Rendahnya tingkat kelulusan tepat waktu ini dapat berdampak negatif terhadap berbagai aspek, seperti efisiensi sumber daya universitas, peringkat akreditasi, reputasi akademik, serta prospek karir mahasiswa. Untuk mengatasi permasalahan ini, diperlukan pendekatan yang lebih sistematis dalam mengidentifikasi faktor-faktor penyebab keterlambatan serta solusi yang dapat diterapkan untuk meningkatkan tingkat kelulusan tepat waktu[6]. Salah satu faktor utama yang berkontribusi terhadap kelulusan tepat waktu adalah Indeks Prestasi Semester (IPS), yang mencerminkan kinerja akademik mahasiswa. Selain itu, status sebagai putra daerah atau non-putra daerah juga dapat mempengaruhi kelulusan, karena perbedaan dalam tingkat adaptasi dan dukungan akademik yang diterima mahasiswa[7]. Oleh karena itu, memahami pengaruh variabel IPS dan status keaslian daerah terhadap kelulusan tepat waktu sangat penting bagi lembaga pendidikan tinggi dalam upaya mereka untuk memberikan dukungan terbaik kepada mahasiswa dan meningkatkan kinerja program akademik mereka. Untuk menjalankan penelitian ini, dibutuhkan beberapa kriteria sebagai penentu data kelulusan mahasiswa, termasuk data detail mahasiswa seperti NPM, IPS dari semester 1 hingga 4 [8][9]. Analisis terhadap faktor-faktor ini diharapkan dapat membantu universitas dalam mengambil langkah intervensi yang lebih tepat sasaran guna meningkatkan tingkat kelulusan tepat waktu.

Dalam penelitian ini, algoritma SVM dengan *kernel linier* akan digunakan secara khusus untuk memprediksi kemungkinan mahasiswa lulus tepat waktu atau tidak, berdasarkan variabel IPS dan status sebagai putra daerah atau non-putra daerah, serta faktor-faktor lain yang relevan. Penggunaan metode SVM dipilih karena efektivitasnya dalam menangani dataset dengan jumlah variabel yang relatif kecil, memungkinkan analisis yang lebih fokus dan mengurangi kompleksitas komputasi tanpa mengorbankan akurasi[10]. Alasan utama pemilihan algoritma *Support Vector Machine* (SVM) dalam penelitian ini adalah karena kemampuannya yang tinggi dalam menangani data berdimensi kecil dengan pemisahan kelas yang linier secara efektif. SVM terbukti unggul dalam klasifikasi biner dan dapat memberikan akurasi prediksi yang tinggi, sebagaimana dibuktikan dalam penelitian oleh Haryatmi et al. [11] yang mencatat akurasi sebesar 94,4% dalam prediksi kelulusan tepat waktu menggunakan SVM. Selain itu, penelitian oleh Junaidi et al. [12] menunjukkan bahwa SVM memberikan hasil terbaik dibandingkan metode lain seperti Naïve Bayes, Random Forest, dan ANN, dengan nilai akurasi mencapai 94%. Kinerja SVM yang konsisten juga ditunjukkan oleh Qisthiano [13], dengan akurasi 85% dalam klasifikasi kelulusan mahasiswa.

Penelitian ini bertujuan untuk mengembangkan model prediksi kelulusan tepat waktu mahasiswa di Universitas Sepuluh Nopember menggunakan metode *Support Vector Machine* (SVM) dengan *kernel linier*. Fokus utama penelitian ini adalah untuk menganalisis keterkaitan antara Indeks Prestasi Semester (IPS) dan status mahasiswa sebagai putra daerah atau non-putra daerah terhadap kelulusan tepat waktu. Selain itu, penelitian ini bertujuan untuk mengukur tingkat keakuratan metode SVM dalam memprediksi kelulusan serta mengevaluasi efektivitasnya dibandingkan dengan metode konvensional yang selama ini digunakan. Dengan adanya model prediksi ini, diharapkan universitas dapat lebih proaktif dalam mengidentifikasi mahasiswa yang berisiko mengalami keterlambatan kelulusan, sehingga strategi intervensi akademik dapat diterapkan lebih tepat sasaran. Model ini juga diharapkan dapat membantu universitas dalam menyusun kebijakan berbasis data untuk meningkatkan efisiensi akademik dan mempertahankan standar akreditasi. Untuk menjaga fokus penelitian, studi ini tidak mencakup mahasiswa yang drop out, serta akan membatasi variabel hanya pada faktor yang relevan, guna menghasilkan model prediksi yang lebih akurat dan dapat diandalkan.

## 2. Tinjauan Pustaka

Penelitian terdahulu yang sudah dibuat untuk prediksi kelulusan tepat waktu telah banyak dilakukan dengan berbagai metode data mining. Dari salah satu metode yang dipakai adalah metode *Support Vector Machine*, merupakan metode pengelompokan yang memiliki keakuratan tinggi dalam proses peramalan dan pengelompokan karakteristik.

Penelitian yang dibuat Emy Haryatmi dan teman-teman membahas penerapan algoritma *Support Vector Machine* (SVM) untuk model prediksi kelulusan mahasiswa tepat waktu di Fakultas Teknik Universitas Swasta di Indonesia[11]. Penelitian menggunakan metode CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang mencakup pemahaman bisnis, pemahaman data, pengolahan data, pemodelan, dan evaluasi. Data penelitian melibatkan atribut seperti usia, jenis kelamin, IPS, IPS, SKS, dan lama masa studi, dengan total data sebanyak 2.192 mahasiswa dari tahun 2009 hingga 2016. Setelah melalui proses pembersihan dan transformasi, 2.181 data digunakan untuk analisis. Data dibagi menjadi tiga kelompok berdasarkan proporsi data pelatihan (70%, 80%, dan 90%) dan data pengujian. Hasil pengujian menunjukkan bahwa algoritma SVM dengan 90% data pelatihan dan 10% data pengujian memberikan akurasi tertinggi sebesar 94,4%, diikuti oleh 80% (92,5%) dan 70% (91,4%). SVM terbukti memiliki keunggulan dibandingkan algoritma lain seperti *Naïve Bayes*, *Decision Tree*, dan *Neural Network*, terutama dalam hal akurasi dan kemampuan menangani data kompleks. Penelitian ini menyimpulkan bahwa model SVM efektif untuk memprediksi kelulusan mahasiswa tepat waktu dan dapat digunakan sebagai alat bantu dalam pengambilan keputusan di institusi pendidikan tinggi.

Berikut penelitian yang telah dibuat oleh Wijiyanto dan teman-teman membahas kontribusi data keluarga dalam memprediksi kelulusan mahasiswa tepat waktu menggunakan algoritma *Support Vector Machine* (SVM) dengan kernel linear[14]. Data penelitian berasal dari 365 mahasiswa lulusan tahun 2023 di Fakultas Ilmu Komputer Universitas Duta Bangsa Surakarta, dengan pembagian 50% untuk data pelatihan dan 50% untuk data pengujian. Atribut yang digunakan mencakup asal kota, pendidikan dan pekerjaan orang tua, penghasilan orang tua, IPS, serta usia mahasiswa. Setelah melalui proses preprocessing, data dimasukkan ke dalam model SVM untuk pelatihan dan pengujian. Hasil evaluasi menggunakan *Confusion Matrix* menunjukkan nilai akurasi 0,86, *Precision* 0,86, *Recall* 0,99, dan *F1-Score* 0,92. Temuan menunjukkan bahwa fitur keluarga, seperti pekerjaan dan pendidikan orang tua, memiliki kontribusi signifikan terhadap kelulusan mahasiswa tepat waktu. Penelitian ini menyarankan implementasi model SVM untuk membantu pihak universitas mengantisipasi risiko ketidaktepatan waktu kelulusan mahasiswa, serta merekomendasikan penelitian lanjutan dengan penanganan data yang tidak seimbang untuk meningkatkan hasil prediksi.

Penelitian yang dilakukan oleh M Riski Qisthiano dengan judul "Klasifikasi Terhadap Prediksi Kelulusan Mahasiswa dengan Menggunakan Metode *Support Vector Machine* (SVM)" [13] dilakukan dengan menggunakan data dari beberapa perguruan tinggi di Kota Palembang. Data yang digunakan mencakup jurusan, perguruan tinggi setiap mahasiswa, jenis kelas, nilai semester, tahun lulus, dan angkatan. Metode SVM digunakan untuk melakukan klasifikasi apakah mahasiswa tersebut lulus tepat waktu atau tidak. Alat bantu yang digunakan dalam penelitian ini adalah Rapidminer, dan penelitian ini menggunakan teknik 5 kali *K-Fold Validation* untuk membagi data menjadi data training dan testing. Hasil dari penelitian ini menunjukkan bahwa model SVM yang digunakan memiliki akurasi sebesar 85.06%, yang menunjukkan efektivitas SVM dalam klasifikasi data kelulusan mahasiswa. Penelitian ini memberikan wawasan yang berguna bagi universitas untuk meningkatkan tingkat kelulusan tepat waktu mahasiswa dengan mengimplementasikan intervensi yang tepat berdasarkan prediksi yang dihasilkan oleh model SVM.

Penelitian yang dilakukan oleh Satrio Junaidi dan juga kawan-kawan yang berjudul "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)" membahas tentang klasifikasi metode data mining untuk prediksi kelulusan tepat waktu mahasiswa menggunakan empat algoritma utama: *Naïve Bayes*, *Random Forest*, *Support Vector Machine* (SVM), dan *Artificial Neural Network* (ANN). Penelitian ini menggunakan atribut seperti jenis kelamin, penghasilan orang tua, lama bimbingan, status mahasiswa bekerja atau tidak, nilai dari semester 1 sampai semester 8, dan IPS. Data diolah menggunakan Python 3 di Jupyter Notebook pada Anaconda, dengan dataset dibagi menjadi 70% untuk data latihan dan 30% untuk data pengujian. Hasil menunjukkan bahwa algoritma

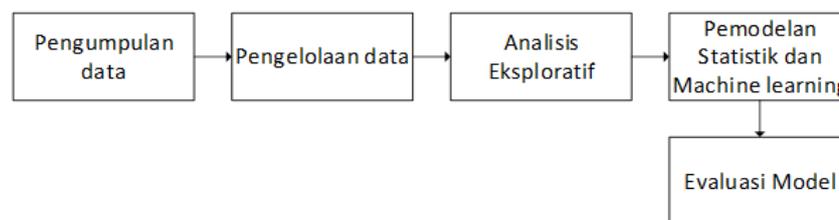
SVM memberikan akurasi terbaik dengan nilai 0.94, yang menunjukkan efektivitasnya dalam memprediksi kelulusan tepat waktu mahasiswa. Penelitian ini memberikan wawasan baru dan berguna bagi perguruan tinggi untuk mengantisipasi dan mengelola kelulusan mahasiswa yang tidak tepat waktu, sehingga dapat meningkatkan kualitas pendidikan dan akreditasi institusi[12].

Penelitian yang dibuat oleh Andi Iwan Nurhidayat dan kawan-kawan ini membahas prediksi kinerja akademik mahasiswa menggunakan metode *machine learning*, khususnya *Support Vector Machine* (SVM) dengan algoritma Sequential Minimal Optimization (SMO), di Jurusan Teknik Informatika Universitas Negeri Surabaya[15]. Penelitian bertujuan untuk membuat model prediksi kinerja akademik mahasiswa yang mencakup kelulusan tepat waktu dan IPS sebagai parameter utama. SMO digunakan untuk meningkatkan efisiensi pemrosesan data dan mengatasi keterbatasan SVM pada optimasi kuadratik, dengan hasil penelitian menunjukkan akurasi sebesar 93,94%, presisi 94,7%, dan *Recall* 94,7%. Penelitian ini menggunakan dataset dari mahasiswa lulusan tahun 2015 hingga 2017 sebanyak 330 data, yang dibagi menjadi data pelatihan dan pengujian dengan teknik percentage split 80%-20%. Proses analisis melibatkan preprocessing data, transformasi data kategori ke numerik, dan seleksi atribut untuk memastikan validitas model. Penggunaan kernel pada SVM seperti Polynomial Kernel, Radial Basis Function, dan Linear Kernel mempermudah pemisahan data non-linear secara optimal. Selain itu, penelitian ini membandingkan metode SMO dengan algoritma regresi logistik untuk evaluasi performa. Evaluasi dilakukan menggunakan *Confusion Matrix* dengan menghitung akurasi, presisi, *Recall*, dan F-measure. Meskipun hasil penelitian menunjukkan akurasi yang tinggi, peneliti merekomendasikan penggunaan dataset yang lebih besar dan penambahan variabel baru seperti asal sekolah atau lokasi geografis untuk meningkatkan detail hasil. Penelitian ini mendukung pengembangan sistem evaluasi akademik berbasis data mining yang dapat membantu pengelolaan program studi di perguruan tinggi.

Berdasarkan tinjauan pustaka, terdapat beberapa gap yang menjadi dasar penelitian ini. Pertama, penelitian sebelumnya cenderung memisahkan faktor akademik dan sosial dalam prediksi kelulusan. Misalnya, Wijiyanto *et al* meneliti faktor keluarga[14], sedangkan Haryatmi *et al*. fokus pada IPS dan SKS[11]. Penelitian ini menggabungkan keduanya, dengan menyoroti IPS dan status putra daerah/non-putra daerah, yang belum banyak dikaji. Kedua, penelitian terdahulu menggunakan dataset dari berbagai universitas dengan rentang waktu yang berbeda, seperti M Riski Qisthiano *et al*[13] di Kota Palembang dan Haryatmi *et al*. [11] pada data 2009–2016, yang dapat mempengaruhi akurasi model. Penelitian ini menggunakan dataset terbaru (2021–2024) dari Universitas Sepuluh Nopember Papua, agar hasilnya lebih relevan dengan kondisi akademik saat ini. Ketiga, meskipun SVM sering dibandingkan dengan metode lain seperti *Naïve Bayes* dan *Decision Tree*, penelitian yang secara spesifik menyoroti SVM dengan kernel linier masih terbatas. Junaidi *et al*. membahas SVM, tetapi tidak menyoroti keunggulan kernel linier dalam klasifikasi akademik. Oleh karena itu, penelitian ini berfokus pada kernel linier, karena hubungan antara IPS dan status mahasiswa diperkirakan bersifat linier. Keempat, banyak penelitian membuktikan efektivitas SVM, tetapi belum banyak yang membahas penerapan hasil prediksi dalam kebijakan akademik universitas[11][12]. Oleh karena itu, penelitian ini tidak hanya meningkatkan akurasi prediksi, tetapi juga memberikan rekomendasi berbasis data agar hasil model dapat digunakan untuk intervensi dini bagi mahasiswa yang berisiko terlambat lulus. Dengan mengatasi gap ini, penelitian ini bertujuan untuk mengembangkan model prediksi kelulusan tepat waktu menggunakan SVM dengan kernel linier, yang lebih terintegrasi, relevan, dan dapat diterapkan dalam kebijakan akademik universitas.

### 3. Metodologi

Menggunakan metode kuantitatif dalam penelitian prediksi kelulusan mahasiswa di Universitas Sepuluh Nopember sangat efektif karena pendekatan ini memungkinkan analisis data yang objektif dan sistematis, menawarkan presisi dan konsistensi yang tinggi. Dengan teknik statistik dan model pembelajaran *Support Vector Machine* (SVM), dapat mengidentifikasi faktor-faktor kunci yang mempengaruhi kelulusan tepat waktu dan mengembangkan model prediktif yang akurat. Pendekatan ini juga memfasilitasi penanganan data dalam jumlah besar, yang ideal untuk institusi pendidikan dengan banyak data mahasiswa, memungkinkan evaluasi yang efisien dan implementasi strategi intervensi yang berbasis data untuk meningkatkan tingkat kelulusan[14]. Adapun gambar tahapan dan langkah langkah nya dapat di lihat pada gambar 1.



Gambar 1. Tahapan penelitian

- 1) **Pengumpulan Data**  
Data dikumpulkan secara sistematis, mencakup nama mahasiswa, Indeks Prestasi Semester (IPS) dari semester 1 hingga 4, serta status sebagai OAP (Orang Asli Papua) atau non-OAP. Sumber data utama berasal dari sistem informasi akademik universitas guna memastikan keakuratan dan kelengkapan data.
- 2) **Pengelolaan data**  
Setelah data diperoleh, langkah berikutnya adalah mengelola data sesuai kebutuhan penelitian. Proses ini mencakup pembersihan data untuk menghilangkan anomali, pengorganisasian data berdasarkan kategori yang relevan, serta transformasi data sesuai format yang dibutuhkan untuk analisis lebih lanjut. Pengelolaan data ini bertujuan untuk menyediakan dataset yang terstruktur dan siap untuk tahap analisis serta pemodelan prediktif.
- 3) **Analisis Eksploratif**  
Dilakukan analisis statistik deskriptif untuk memahami distribusi data, tren akademik mahasiswa, serta mendeteksi pola atau *outlier* yang dapat mempengaruhi hasil prediksi. Visualisasi data digunakan untuk memperoleh wawasan lebih lanjut mengenai hubungan antar variabel.
- 4) **Pemodelan Statistik dan *Machine learning***  
Pada tahap ini, diterapkan model statistik atau algoritma *machine learning*, seperti *Support Vector Machine* (SVM), untuk mengembangkan prediksi kelulusan mahasiswa. Proses ini mencakup pemilihan fitur yang relevan, pelatihan model, serta validasi model guna memastikan performa yang optimal dalam prediksi kelulusan.
- 5) **Evaluasi Model**  
Kinerja model diukur menggunakan metrik evaluasi, seperti akurasi, *Area Under Curve - Receiver Operating Characteristic* (AUC-ROC), serta *Confusion Matrix*. Dari *Confusion Matrix*, dihasilkan metrik tambahan seperti *Accuracy* (Akurasi), *Precision* (Presisi), *Recall* (*Sensitivity/True Positive Rate*), dan *F1-Score*, yang digunakan untuk menilai tingkat keakuratan serta keandalan model prediktif. Evaluasi ini sangat penting untuk memastikan bahwa model mampu memprediksi kelulusan mahasiswa dengan tingkat akurasi yang tinggi.

### 3.2. *Support Vector Machine*

*Support Vector Machine* (SVM) dapat dijelaskan sebagai metode untuk mencari *hyperplane* terbaik yang berfungsi sebagai pemisah antara dua kelas pada ruang input. Dalam klasifikasi, SVM berusaha menemukan *hyperplane* yang dapat memisahkan kedua kelompok *pattern*, di mana *pattern* yang tergabung pada kelas negatif disimbolkan dengan kotak, dan kelas positif disimbolkan dengan lingkaran[12]. *Hyperplane* pemisah terbaik dapat ditemukan dengan mengukur margin, yaitu jarak antara *hyperplane* dengan data terdekat dari masing-masing kelas, dan mencari titik maksimal dari margin tersebut. Data yang paling dekat dengan *hyperplane* tersebut disebut sebagai *support vector*[16]. SVM merupakan metode klasifikasi *supervised* yang melibatkan dua versi:

- a. **Linear:** Menggunakan *hyperplane* garis lurus untuk memisahkan kelas.
- b. **Non-linear:** Menggunakan fungsi kernel untuk memetakan data ke dimensi lebih tinggi, memungkinkan pemisahan non-linear. Kernel seperti RBF, polinomial, dan MLP sering digunakan untuk mengatasi kompleksitas data dunia nyata yang umumnya non-linear.

Hampir semua dataset pada kehidupan nyata bersifat non-linear, sehingga untuk mengklasifikasikan data tersebut, SVM menggunakan pendekatan kernel. Kernel adalah fungsi yang memetakan fitur data dari dimensi awal ke dimensi yang lebih tinggi, berbeda dengan metode klasifikasi umum yang mengurangi dimensi untuk menyederhanakan komputasi dan meningkatkan akurasi prediksi[17]. Metode SVM juga banyak digunakan dalam klasifikasi untuk

analisis sentimen dan dapat mempelajari pengklasifikasi seperti *radial basis functional* (RBF), polinomial, dan *multi-layer perceptron* (MLP)[13]. Adapun berikut rumus SVM.

$$y = \text{sign}\left(\sum_{i=1}^N \alpha_i y_i K(x_i, x) + b\right) \tag{1}$$

Keterangan:

$\alpha_i$ : Koefisien Lagrange yang dioptimalkan selama pelatihan.

$y_i$ : Label kelas (+1 atau -1).

$x_i$ : Data pelatihan.

$K(x_i, x)$ : Fungsi kernel yang menghitung kesamaan antara data pelatihan  $x_i$  dan data baru  $x$

$b$ : Bias, ditentukan dari solusi optimasi.

### 3.3. Kernel SVM

Rumus Dasar Kernel pada SVM digunakan untuk menangani data non-linear dengan memetakan data dari ruang asli ke ruang berdimensi lebih tinggi. Ide utamanya adalah menggunakan fungsi kernel  $K(x_i, x)$  untuk menghitung produk dalam ruang dimensi tinggi tanpa secara eksplisit melakukan transformasi data, sehingga menghemat komputasi. Fungsi kernel  $K(x_i, x)$  didefinisikan sebagai:

$$K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j) \tag{2}$$

### 3.4. Confusion Matrix

*Confusion Matrix* adalah alat evaluasi dalam pembelajaran mesin yang menampilkan jumlah prediksi benar dan salah dalam empat kategori: *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN). Matriks ini digunakan untuk menghitung metrik evaluasi seperti *akurasi*, *presisi*, *recall*, dan *F1-Score*, yang penting dalam menilai performa model klasifikasi[18]. Dalam analisis klasifikasi, TP menunjukkan data positif yang diklasifikasikan benar, TN adalah data negatif yang diklasifikasikan benar, sementara FP (galat Tipe I) terjadi saat data negatif diprediksi sebagai positif, dan FN (galat Tipe II) terjadi saat data positif diprediksi sebagai negatif. Dengan *Confusion Matrix*, distribusi kesalahan model dapat dianalisis guna mengoptimalkan strategi klasifikasi dan meningkatkan akurasi prediksi[19]. Pentingnya *Confusion Matrix* dalam berbagai bidang analisis data, terutama dalam klasifikasi berbasis *machine learning*. Misalnya, dalam kategorisasi akademis untuk memprediksi keberhasilan siswa, *Confusion Matrix* digunakan untuk menilai keakuratan prediksi sistem, apakah siswa akan lulus tepat waktu berdasarkan kriteria yang telah ditentukan sebelumnya. Dengan memahami distribusi kesalahan yang terkait dengan *Confusion Matrix*, pengoptimalan dapat dilakukan pada model klasifikasi yang digunakan, ini akan menghasilkan hasil yang lebih akurat dan praktis daripada prediksi yang dibuat oleh *Confusion Matrix*[20]. Dalam evaluasi model klasifikasi, *Confusion Matrix* menghasilkan metrik utama seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, yang mengukur keberhasilan prediksi model. *Accuracy* menunjukkan persentase prediksi yang benar dari keseluruhan data yang diuji dan dirumuskan sebagai:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Di samping itu, *Precision* digunakan untuk mengukur ketepatan model dalam mengklasifikasikan sampel positif, yaitu seberapa banyak dari prediksi positif yang benar-benar positif. *Precision* dirumuskan sebagai:

$$\text{precision} = \frac{TP}{TP + FP} \tag{4}$$

Sementara itu, *Recall* (juga dikenal sebagai *Sensitivity* atau *True Positive Rate*) mengukur sejauh mana model mampu mendeteksi kelas positif dari seluruh data yang sebenarnya positif. Rumusnya adalah:

$$\text{recall} = \frac{TP}{TP + FN} \tag{5}$$

Terakhir, *F1-Score* merupakan ukuran keseimbangan antara *Precision* dan *Recall*, yang dirumuskan sebagai rata-rata harmonis dari kedua metrik tersebut:

$$F1\ score = \frac{2 \times recall \times precision}{recall + precision} \quad (6)$$

### 3.5. Area Under the Curve (AUC)

AUC adalah metrik evaluasi yang mengukur kemampuan model klasifikasi dalam membedakan antara kelas positif dan negatif. Secara grafis, AUC merepresentasikan luas area di bawah kurva ROC (*Receiver Operating Characteristic*), yang memplot hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR) pada berbagai threshold. Semakin besar nilai AUC, semakin baik model dalam membedakan kedua kelas tersebut. Nilai AUC berkisar antara 0 hingga 1, di mana nilai mendekati 1 menunjukkan kinerja model yang sangat baik, sedangkan nilai mendekati 0,5 mengindikasikan kinerja yang setara dengan tebakan acak[21].

### 3.6. Cross-Validation (Evaluasi Stabilitas Model)

*Cross-Validation* adalah teknik yang digunakan untuk mengevaluasi kinerja model dengan membagi dataset menjadi beberapa subset atau "*fold*". Salah satu metode yang umum digunakan adalah *K-Fold Cross-Validation*, di mana data dibagi menjadi *k* subset. Model dilatih menggunakan *k-1* subset dan diuji pada subset yang tersisa. Proses ini diulang sebanyak *k* kali, dengan setiap subset digunakan sekali sebagai data uji. Teknik ini membantu dalam mengurangi bias evaluasi dan memastikan bahwa model memiliki kinerja yang konsisten di berbagai subset data. Selain itu, *Cross-Validation* membantu dalam mendeteksi masalah *overfitting* dan memastikan bahwa model dapat digeneralisasi dengan baik ke data yang tidak terlihat sebelumnya[22].

### 3.7. Data yang di kelola

Adapun berikut data yang akan digunakan sebagai dataset dapat di lihat pada tabel 1

Tabel 1. Data 2018-2022

NPM	Angkatan	IPS 1	IPS 2	IPS 3	IPS 4	OAP/Non-OAP
201811001	2018	3,41	3,09	3,78	3,19	Non-OAP
201811002	2018	3,09	3,13	3,35	0,95	OAP
201811003	2018	3,32	3,35	3,85	2,98	Non-OAP
201811004	2018	3,27	3,70	3,65	2,90	Non-OAP
201811005	2018	3,45	3,57	3,93	2,88	Non-OAP
...	...	...	...	...	...	...
...	...	...	...	...	...	...
202211062	2022	1,55	0,52	0,00	0,00	Non-OAP
202211063	2022	0,35	0,92	2,68	0,00	OAP
202211064	2022	0,52	0,16	1,28	0,00	Non-OAP
202211065	2022	2,40	1,60	0,00	0,00	Non-OAP
202211066	2022	1,96	1,60	0,42	0,29	OAP

Keterangan:

NPM = Nomor Pokok Mahasiswa

IPS = Indeks Prestasi Semester

IPK = Indeks Prestasi Kumulatif

OAP/Non-OAP = Status mahasiswa sebagai Orang Asli Papua (OAP) atau Non-OAP.

## 4. Hasil dan Pembahasan

### 4.1. hasil

Pada penelitian ini, model *Support Vector Machine* (SVM) dengan *kernel linear* digunakan untuk memprediksi kelulusan mahasiswa berdasarkan data Indeks Prestasi Semester (IPS) dari angkatan 2018 hingga 2022. Proses evaluasi dilakukan dalam dua tahap utama: Tahap 1 (Penguji Model pada Angkatan 2018-2020) dan Tahap 2 (Prediksi Kelulusan pada Angkatan 2021-2022).

1) Tahap 1 (Penguujian Model pada Angkatan 2018-2020)

Adapun tahapan 1 yang menggunakan data dari 2018-2020 untuk penguujian model dapat dilihat pada gambar 2. Dataset dibagi menjadi 80% data latih dan 20% data uji menggunakan metode train-test split.

	npm	angkatan	IPS 1	IPS 2	IPS 3	IPS 4	oap/non-oap
0	201811001	2018	3.409091	3.086957	3.782609	3.190476	NON-OAP
1	201811002	2018	3.090909	3.130435	3.347826	0.952381	OAP
2	201811003	2018	3.318182	3.347826	3.847826	2.976190	NON-OAP
3	201811004	2018	3.272727	3.695652	3.652174	2.904762	NON-OAP
4	201811005	2018	3.454545	3.565217	3.934783	2.880952	NON-OAP
285	202211062	2022	1.550000	0.523810	0.000000	0.000000	NON-OAP
286	202211063	2022	0.350000	0.916667	2.681818	0.000000	OAP
287	202211064	2022	0.523810	0.157895	1.277778	0.000000	NON-OAP
288	202211065	2022	2.400000	1.600000	0.000000	0.000000	NON-OAP
289	202211066	2022	1.960000	1.600000	0.421053	0.285714	OAP

Gambar 2. Penguujian Model pada Angkatan 2018-2020

Setelah model dilatih menggunakan data latih, dilakukan evaluasi menggunakan data uji. Berikut hasil evaluasi model pada tahap ini:

a) Evaluasi model

Evaluasi dilakukan menggunakan Accuracy, Precision, Recall, dan F1-Score untuk melihat seberapa baik model dalam memprediksi kelulusan mahasiswa. Hasil evaluasi ditunjukkan pada Tabel berikut:

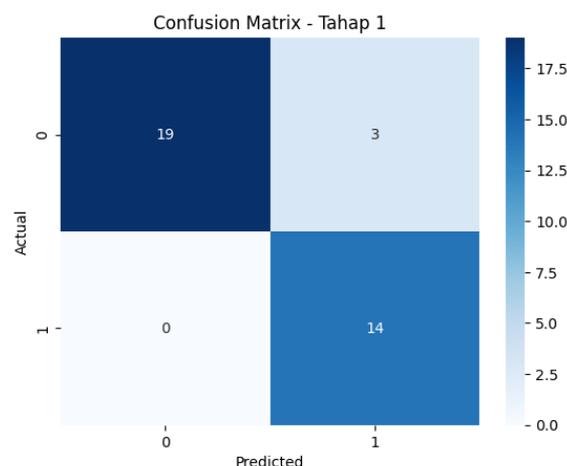
Tabel 2. Evaluasi model tahapan 1

Metrik	Nilai
Accuracy	0.92
Precision	0.86
Recall	1.00
F1-Score	0.92
Mean CV Accuracy	0.98

Dari hasil di atas, dapat dilihat bahwa model memiliki akurasi yang cukup tinggi sebesar 92%, menunjukkan bahwa model dapat mengklasifikasikan mayoritas mahasiswa dengan benar. Recall yang mencapai 100% menunjukkan bahwa model tidak melewatkan mahasiswa yang benar-benar lulus (*False Negative* = 0). Namun, nilai Precision = 86% menunjukkan bahwa masih terdapat False Positive, yaitu mahasiswa yang diprediksi lulus tetapi sebenarnya tidak lulus.

b) Evaluasi Menggunakan Confusion Matrix

Adapun berikut evaluasi model menggunakan Confusion Matrix dapat di lihat pada gambar 3



Gambar 3. Confusion Matrix tahapan 1

Berdasarkan tabel di atas:

*True Positive* (TP) = 19, menunjukkan bahwa sebanyak 19 mahasiswa lulus terprediksi dengan benar sebagai lulus.

*False Positive* (FP) = 3, menunjukkan bahwa terdapat 3 mahasiswa yang diprediksi lulus padahal sebenarnya tidak lulus.

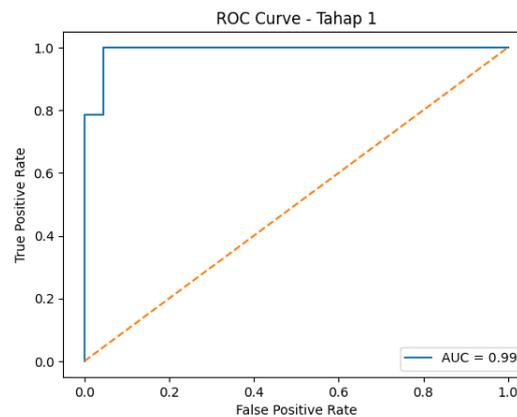
*True Negative* (TN) = 14, menunjukkan bahwa 14 mahasiswa tidak lulus terprediksi dengan benar sebagai tidak lulus.

*False Negative* (FN) = 0, menunjukkan bahwa tidak ada mahasiswa yang seharusnya lulus tetapi diprediksi tidak lulus, yang berarti Recall mencapai 100%.

#### c) Evaluasi Menggunakan ROC Curve dan AUC Score

Selain evaluasi metrik dasar, model diuji menggunakan ROC Curve dan AUC (*Area Under the Curve*) untuk melihat kemampuannya dalam membedakan mahasiswa yang lulus dan tidak lulus. ROC Curve merupakan kurva yang menunjukkan hubungan antara *True Positive Rate* (TPR) dan *False Positive Rate* (FPR). Semakin tinggi AUC, semakin baik model dalam membedakan antara mahasiswa yang lulus dan tidak lulus.

Pada tahap ini, nilai AUC yang diperoleh sangat tinggi, yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik. Ini berarti bahwa model dapat secara efektif memprediksi mahasiswa yang akan lulus dan yang tidak lulus dengan tingkat kesalahan yang minim. Adapun berikut visualisasi dapat dilihat pada gambar 4



Gambar 4. ROC Curve

## 2) Tahap 2 Prediksi Model pada Angkatan 2021-2022

Setelah model diuji pada tahap pertama dan menunjukkan performa yang baik, model yang telah dilatih kemudian digunakan untuk memprediksi kelulusan mahasiswa angkatan 2021-2022. Pada tahap ini, data mahasiswa tahun 2021-2022 digunakan sebagai data prediksi, tanpa mengetahui hasil kelulusan aktualnya.

#### a) Hasil Prediksi Kelulusan Mahasiswa Angkatan 2021-2022

Berdasarkan hasil prediksi, mahasiswa dikelompokkan ke dalam dua kategori, yaitu Lulus dan Tidak Lulus, berdasarkan hasil klasifikasi dari model. Distribusi hasil prediksi dapat divisualisasikan dalam bentuk grafik, yang menunjukkan proporsi mahasiswa yang diprediksi lulus dan tidak lulus. Selain itu, model diuji kembali menggunakan *Cross-Validation*, untuk memastikan bahwa model tetap stabil dan dapat digeneralisasi terhadap data mahasiswa terbaru. Hasil evaluasi tetap menunjukkan *Mean CV Accuracy* yang tinggi, yang mengindikasikan bahwa model tetap dapat memprediksi dengan baik meskipun diterapkan pada data mahasiswa angkatan baru. Adapun berikut hasil evaluasi model.

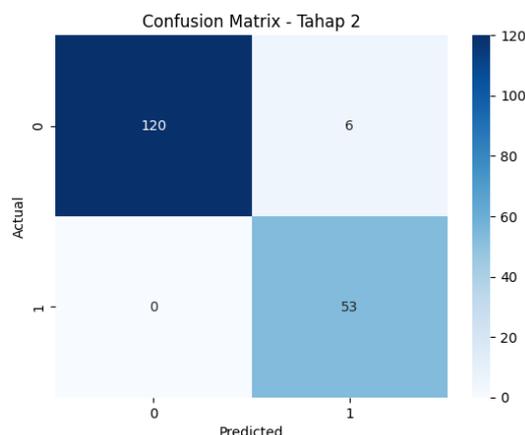
Tabel 3. Evaluasi model tahapan 2

Metrik	Nilai
<i>Accuracy</i>	0.97
<i>Precision</i>	0.90
<i>Recall</i>	1.00
<i>F1-Score</i>	0.95
<i>Mean CV Accuracy</i>	0.98

Hasil *Cross-Validation* pada Tahap 2 menunjukkan bahwa model memiliki akurasi yang stabil di berbagai subset data, dengan skor [1.00, 0.9722, 0.9167, 0.9722, 0.9714] dan Mean CV Accuracy sebesar 0.97 (97%). Nilai ini menegaskan bahwa model tidak mengalami *overfitting* dan tetap akurat serta andal dalam memprediksi kelulusan mahasiswa angkatan 2021-2022.

b) Evaluasi Menggunakan *Confusion Matrix*

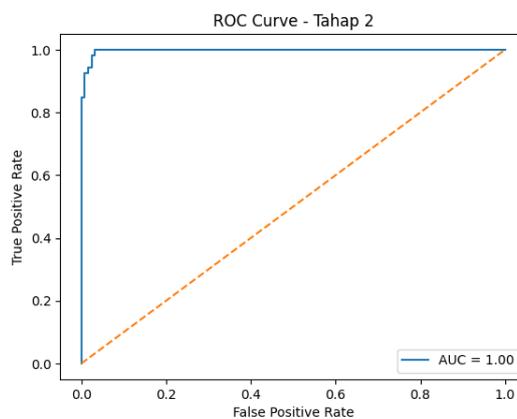
Adapun hasil evaluasi yang divisualisasikan menggunakan *Confusion Matrix* dapat di lihat pada gambar 5.



Gambar 5. *Confusion Matrix* tahapan 2

c) Evaluasi Menggunakan *ROC Curve* dan *AUC Score*

Evaluasi lebih lanjut dilakukan menggunakan *ROC Curve* dan *AUC Score*. Hasil menunjukkan bahwa model masih memiliki AUC yang tinggi, yang berarti model tetap memiliki kemampuan diskriminatif yang baik terhadap mahasiswa angkatan 2021-2022.



Gambar 6. *ROC Curve*

d) Visualisasi OAP dan Non OAP

Grafik menunjukkan prediksi kelulusan tepat waktu berdasarkan kategori OAP dan Non-OAP. Sebagian besar mahasiswa, baik OAP maupun Non-OAP, diprediksi tidak lulus tepat waktu, dengan proporsi ketidakkelulusan lebih tinggi pada Non-OAP.

**4.2. Pembahasan**

Hasil penelitian ini menunjukkan bahwa algoritma *Support Vector Machine* (SVM) dengan kernel linier mampu memprediksi kelulusan tepat waktu dengan akurasi tinggi, sejalan dengan penelitian Haryatmi et al. [11], Junaidi et al. [12], dan Qisthiano [13] yang juga mencatat performa unggul SVM dalam klasifikasi akademik. Keunggulan studi ini terletak pada integrasi dua variabel penting—IPS dan status OAP/non-OAP—yang belum banyak dikombinasikan dalam penelitian terdahulu, seperti pada Wijiyanto et al. [14] yang hanya menyoroti aspek keluarga, atau Haryatmi et al. [11] yang fokus pada data akademik. Selain itu, penekanan pada

penggunaan kernel linier memberikan kejelasan bahwa pendekatan ini efektif untuk data dengan relasi linier antar variabel [15]. Temuan ini memperkaya literatur dengan model yang lebih komprehensif dan relevan untuk kebijakan akademik berbasis data.

## 5. Simpulan

Penelitian ini menunjukkan bahwa *Support Vector Machine* (SVM) dengan kernel linear dapat memprediksi kelulusan tepat waktu mahasiswa dengan akurasi tinggi dan stabilitas yang baik. Pada Tahap 1, model memiliki akurasi 92% dengan *Recall* sempurna (1.00), meskipun masih terdapat *False Positive*. Pada Tahap 2, model tetap stabil dengan *Mean Cross-Validation Accuracy* 97%, membuktikan bahwa model tidak mengalami *overfitting* dan dapat melakukan prediksi pada data baru dengan baik. Evaluasi menggunakan *Confusion Matrix* dan *ROC Curve* juga menegaskan bahwa model memiliki kemampuan klasifikasi yang sangat baik. Hasil penelitian ini dapat digunakan sebagai alat bantu akademik untuk mengidentifikasi mahasiswa yang berisiko tidak lulus tepat waktu, sehingga intervensi akademik dapat dilakukan lebih awal. Namun, penelitian ini masih terbatas pada variabel IPS, sehingga penelitian selanjutnya disarankan untuk menambahkan variabel baru dan membandingkan model dengan algoritma lain untuk meningkatkan akurasi prediksi.

## Daftar Referensi

- [1] D. Marcelina, F. Aziz, and Y. Manoppo, "Sistem Pendukung Keputusan Memprediksi Kelulusan Mahasiswa Informatika Menggunakan Metode Simple Additive Weighting," *J. Inf. dan Teknol.*, vol. 5, no. 4, pp. 73–77, 2023, doi: 10.60083/jidt.v5i4.420.
- [2] F. Irahma Rachman, S. Mujadilah, T. Wahyuni, and L. Anas, "Prediksi Tingkat Kelulusan Menggunakan K-Means Pada Program Studi Informatika Unismuh Makassar," *J. Fasikom*, vol. 13, no. 3, pp. 504–510, 2023, doi: 10.37859/jf.v13i3.6061.
- [3] S. Andriani, A. Nazir, R. M. Candra, F. Syafria, and I. Afrianty, "Implementasi Algoritma K-Nearest Neighbor Untuk Menentukan Klasifikasi Kelulusan Mahasiswa Teknik Informatika," *J. Comput. Syst. Informatics*, vol. 4, no. 4, pp. 922–930, 2023, doi: 10.47065/josyc.v4i4.3914.
- [4] A. Fatkhudin, M. Y. Febrianto, F. A. Artanto, M. W. N. Hadinata, and R. Fahlevi, "Algoritma Decision Tree C.45 Dalam Analisa Kelulusan Mahasiswa Program Studi Manajemen Informatika Ump," *J. Ilm. Ilmu Komput.*, vol. 8, no. 2, pp. 83–86, 2022, doi: 10.35329/jiik.v8i2.240.
- [5] A. Rahmayanti, L. Rusdiana, and S. Suratno, "Perbandingan Metode Algoritma C4.5 Dan Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa," *Walisongo J. Inf. Technol.*, vol. 4, no. 1, pp. 11–22, 2022, doi: 10.21580/wjit.2022.4.1.9654.
- [6] A. Khaerunnisa, "Analisis Tingkat Kelulusan Mahasiswa di Unisba dengan menggunakan Algoritma K-Means Clustering," *J. Ris. Mat.*, pp. 67–76, 2022, doi: 10.29313/jrm.v2i1.1018.
- [7] R. Rudianto, R. Kania, and T. I. Solihati, "Prediksi Kelulusan Mahasiswa Teknik Informatika Universitas Banten Jaya Menggunakan Algoritma Neural Network," *J. Sist. Inf. dan Inform.*, vol. 5, no. 2, pp. 193–200, 2022, doi: 10.47080/simika.v5i2.2123.
- [8] K. Suhada, A. Elanda, and A. Aziz, "Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan Algoritma C4.5 (Studi Kasus: STMIK Rosma Karawang)," *Dirgamaya J. Manaj. dan Sist. Inf.*, vol. 1, no. 2, pp. 14–27, 2021, doi: 10.35969/dirgamaya.v1i2.182.
- [9] A. Gadi Ana Amas, G. Kopong Pati, F. Ema Ose Sanga, T. Informatika, and S. Stella Maris Sumba, "JESCE (Journal of Electrical and System Control Engineering) Penerapan K-Optimal Pada Algoritma KNN Untuk Prediksi Kelulusan Tepat Waktu Mahasiswa Program Studi Teknik Informatika Application of K-Optimal in the KNN Algorithm to Predict Timely Graduation," *Jesce*, vol. 7, no. 2, pp. 92–97, 2024, doi: 10.31289/jesce.v6i2.10536.
- [10] N. M. A. Mahar, Vihi Atina, and Nugroho Arif Sudibyo, "Pemodelan Prediksi Kelulusan Mahasiswa Dengan Metode Naïve Bayes Di Uniba," *J. Manaj. Inform. dan Sist. Inf.*, vol. 6, no. 2, pp. 148–158, 2023, doi: 10.36595/misi.v6i2.875.
- [11] E. Haryatmi and S. Pramita Hervianti, "Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 386–392, 2021, doi: 10.29207/resti.v5i2.3007.
- [12] Satrio Junaidi, R. Valicia Anggela, and D. Kariman, "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)," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 109–119, 2024, doi: 10.52158/jacost.v5i1.489.
- [13] M Riski Qisthiano, "Klasifikasi Terhadap Prediksi Kelulusan Mahasiswa Dengan Menggunakan Metode Support Vector Machine (Svm)," *Semin. Nas. Teknol. dan Multidisiplin Ilmu*, vol. 2, no. 2, pp. 203–207, 2022, doi: 10.51903/semnastekmu.v2i1.170.
- [14] W. Wijiyanto and S. Sopingi, "Kontribusi Keluarga Dalam Prediksi Mahasiswa Lulus Tepat Waktu Menggunakan Model Support Vector Machine," *DutaCom*, vol. 17, no. 1, pp. 25–36, 2023, doi: 10.47701/dutacom.v17i1.3784.
- [15] A. Nurhidayat, A. Asmunin, and D. F. Suyatno, "Prediksi Kinerja Akademik Mahasiswa Menggunakan Machine Learning dengan Sequential Minimal Optimization untuk Pengelola Program Studi," *J. Inf. Eng. Educ. Technol.*, vol. 5, no. 2, pp. 84–91, 2021, doi: 10.26740/jieet.v5n2.p84-91.
- [16] A. Putri *et al.*, "Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
- [17] B. I. Nugroho, N. A. Santoso, and A. A. Murtopo, "Prediksi Kemampuan Akademik Mahasiswa dengan Metode Support Vector Machine," *Remik*, vol. 7, no. 1, pp. 177–188, 2023, doi: 10.33395/remik.v7i1.12010.
- [18] M. F. Nasrullah, R. R. Saedudin, and F. Hamami, "Perbandingan Akurasi Algoritma C4.5 Dan K-Nearest Neighbors Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Indonesia," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 2, pp. 628–638, May 2024, doi: 10.29100/JUPI.V9I2.4655.
- [19] W. A. Naseer, S. Sarwido, and B. B. Wahono, "Gradient Boosting Optimization With Pruning Technique For Prediction Of Bmt Al-Hikmah Permata Customer Data," *J. Inform. Teknol. dan Sains*, vol. 6, no. 3, pp. 719–727, Aug. 2024, doi: 10.51401/JINTEKS.V6I3.4702.
- [20] R. Indransyah, Y. H. Chrisnanto, P. N. Sabrina, and S. Kom, "Klasifikasi Sentimen Pergelaran Motogp Di Indonesia Menggunakan Algoritma Correlated Naïve Bayes Clasifier," *INFOTECH J.*, vol. 8, no. 2, pp. 60–66, Oct. 2022, doi: 10.31949/INFOTECH.V8I2.3103.
- [21] Y. A. Singgalen, "Analisis Performa Algoritma NBC, DT, SVM dalam Klasifikasi Data Ulasan Pengunjung Candi Borobudur Berbasis CRISP-DM," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1634–1646-1634–1646, Dec. 2022, doi: 10.47065/BITS.V4I3.2766.
- [22] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *J. Algoritm.*, vol. 21, no. 1, May 2024, doi: 10.33364/algoritma/v.21-1.1618.