

**ANALISIS PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU  
MENGUNAKAN METODE ALGORITMA MACHINE LEARNING DAN  
FEATURE SELECTION**

**Royan Habibie Sukarna<sup>1</sup>, Holilah<sup>2</sup>, Fitri Damyati<sup>3</sup>, Mohamad Hilman<sup>4</sup>**

Program Studi Informatika, Fakultas Teknik, Universitas Sultan Ageng Tirtayasa<sup>1,2,3,4</sup>

Jl. Jenderal Sudirman Km 3, Kotabumi, Kec. Purwakarta, Kota Cilegon, Banten 42435

e-mail: \*<sup>1</sup>[royan@untirta.ac.id](mailto:royan@untirta.ac.id), <sup>2</sup>[holilah@untirta.ac.id](mailto:holilah@untirta.ac.id), <sup>3</sup>[fitridamyati@untirta.ac.id](mailto:fitridamyati@untirta.ac.id),

<sup>4</sup>[mohamad.hilman@untirta.ac.id](mailto:mohamad.hilman@untirta.ac.id)

**Abstract**

*The primary goal is to improve prediction accuracy by selecting the most relevant aspects. Random Forest, Support Vector Machine (SVM), Logistic Regression, and K-Nearest Neighbors (KNN) are the four machine learning algorithms we utilized. Each algorithm was run using a dataset with complete features and a dataset tested using the Gain Ratio Attribute Evaluator. The KNN algorithm with feature selection achieved the highest accuracy of 74.44% and an Area Under the Curve (AUC) of 0.8212. This model showed a balanced accuracy improvement compared to its performance using the dataset with complete features, which had an accuracy of 72.83% and an AUC of 0.8071. Similarly, the Random Forest model with feature selection showed an accuracy of 72.00% and an AUC of 0.7741, compared to an accuracy of 70.52% and an AUC of 0.7672 with all features. The SVM model with feature selection also improved, reaching an accuracy of 72.28% and an AUC of 0.7812, compared to an accuracy of 69.80% and an AUC of 0.774 with all features. Logistic Regression showed minimal change, with an accuracy of 69.14% and an AUC of 0.7644 after feature selection, compared to an accuracy of 69.25% and an AUC of 0.7645 with all features. This study finds that the prediction of on-time graduation is better predicted with feature selection. The KNN and SVM models excel in feature selection due to their highest discrimination and prediction accuracy capabilities. These results indicate that more effective and accurate models can be created using feature selection.*

**Keywords:** Machine Learning, Feature Selection, On-Time Graduation Prediction, Random Forest, SVM, Logistic Regression

**PENDAHULUAN**

*Machine learning (ML)* adalah cabang dari kecerdasan buatan (AI) yang memungkinkan komputer untuk belajar dari data tanpa diprogram secara eksplisit. Dengan memanfaatkan algoritma yang dilatih pada dataset, model ML dapat membuat prediksi dan klasifikasi berdasarkan data baru. Manfaat utama dari machine learning termasuk kemampuan untuk menganalisis jumlah data yang besar dan kompleks secara efisien, yang pada gilirannya dapat meningkatkan pengambilan keputusan dan prediksi di berbagai bidang (Du et al., 2020; Namoun & Alshantqi, 2021).

Penerapan machine learning telah meluas ke berbagai sektor seperti kesehatan, keuangan, pemasaran, dan keamanan siber. Di sektor kesehatan, ML digunakan untuk menganalisis data genetik dan membantu dalam diagnosis penyakit. Di bidang keuangan, ML membantu dalam mendeteksi penipuan dan mengelola risiko. Dalam pemasaran, ML digunakan untuk mempersonalisasi pengalaman pelanggan dan mengoptimalkan strategi kampanye (Aditya

Nirwana et al., 2023; Awalina & Rahayu, 2023; Waheed et al., 2020). Selain itu, di sektor keamanan siber, ML berperan dalam mendeteksi dan mencegah serangan siber melalui analisis pola data (Situmorang, 2023; Uzhah et al., 2024).

Dalam dunia pendidikan, penerapan machine learning dikenal sebagai Educational Data Mining (EDM). EDM memungkinkan analisis data akademik untuk mengidentifikasi pola dan tren yang dapat membantu dalam meningkatkan kinerja pendidikan. Salah satu manfaat utama EDM adalah kemampuannya untuk memprediksi kinerja akademik siswa, mengidentifikasi siswa yang berisiko, dan memberikan intervensi yang tepat waktu (Tekin, 2014; Thakar et al., 2015; Xing et al., 2015). Teknik-teknik ini telah digunakan untuk mengatasi masalah seperti rendahnya tingkat kelulusan tepat waktu di banyak institusi pendidikan.

Studi ini mengkaji prediksi kelulusan mahasiswa tepat waktu dengan menggunakan berbagai algoritma machine learning dan teknik seleksi fitur. Tujuan utama adalah meningkatkan akurasi prediksi dengan memilih aspek yang paling relevan. Algoritma yang digunakan dalam penelitian ini meliputi *Random Forest*, *Support Vector Machine (SVM)*, *Logistic Regression*, dan *K-Nearest Neighbors (KNN)*. Setiap algoritma diuji menggunakan dataset lengkap dan dataset yang telah mengalami seleksi fitur menggunakan *Gain Ratio Attribute Evaluator* (Du et al., 2020; Sukarna & Ansori, 2022).

Hasil penelitian menunjukkan bahwa seleksi fitur secara signifikan mempengaruhi kinerja algoritma. Algoritma KNN dengan feature selection mencapai akurasi tertinggi sebesar 74,44% dan *Area Under the Curve (AUC)* sebesar 0,8212. Model ini menunjukkan peningkatan akurasi dibandingkan dengan kinerjanya menggunakan dataset lengkap, yang memiliki akurasi 72,83% dan AUC 0,8071. Demikian pula, model Random Forest dengan seleksi fitur menunjukkan akurasi sebesar 72,00% dan AUC 0,7741, dibandingkan dengan akurasi 70,52% dan AUC 0,7672 dengan semua fitur. Model SVM dengan seleksi fitur juga meningkat, mencapai akurasi 72,28% dan AUC 0,7812, dibandingkan dengan akurasi 69,80% dan AUC 0,774 dengan semua fitur. Logistic Regression menunjukkan perubahan minimal, dengan akurasi 69,14% dan AUC 0,7644 setelah seleksi fitur, dibandingkan dengan akurasi 69,25% dan AUC 0,7645 dengan semua fitur.

Penelitian ini menemukan bahwa prediksi kelulusan mahasiswa tepat waktu lebih baik diprediksi dengan feature selection. Model KNN dan SVM unggul dalam feature selection karena mereka memiliki kemampuan diskriminasi dan akurasi prediksi yang paling tinggi. Hasil ini menunjukkan bahwa model yang lebih efektif dan akurat dapat dibuat dengan menggunakan feature selection.

## METODE PENELITIAN

Penelitian ini dibagi menjadi beberapa tahap, diantaranya:

### 1. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari basis data akademik mahasiswa, termasuk data penerimaan mahasiswa baru dan profil akademik mahasiswa. Data ini mencakup berbagai atribut yang relevan untuk analisis prediksi kelulusan tepat waktu, seperti IPK, jalur masuk, kode program studi, usia, dan sumber dana.

### 2. Pra-pemrosesan Data

Data mentah yang diperoleh dari basis data akademik dibersihkan dan diintegrasikan untuk memastikan kelengkapan dan konsistensi data. Langkah-langkah pra-pemrosesan data meliputi:

- 2.1. Menghapus data duplikat dan data yang tidak lengkap.
- 2.2. Mengisi nilai yang hilang dengan metode imputasi.
- 2.3. Mengonversi atribut kategorikal menjadi bentuk numerik menggunakan teknik encoding.

### 3. Seleksi Fitur

Untuk meningkatkan kinerja model (Sukarna & Ansori, 2022), seleksi fitur dilakukan menggunakan *Gain Ratio Attribute Evaluator*. Metode ini membantu mengidentifikasi fitur yang paling relevan dan mengurangi fitur yang tidak signifikan atau redundan.

## Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Algoritma Machine Learning Dan Feature Selection

Proses seleksi fitur dilakukan dengan bantuan alat WEKA, dan fitur-fitur yang dipilih meliputi IPK, jalur masuk, dan kode program studi.

### 4. Penerapan Algoritma Machine Learning

Empat algoritma *machine learning* digunakan dalam penelitian ini:

4.1. *Random Forest*

4.2. *Support Vector Machine (SVM)*

4.3. *Logistic Regression*

4.4. *K-Nearest Neighbors (KNN)*

Setiap algoritma dijalankan pada dataset lengkap dan dataset yang telah mengalami seleksi fitur. Proses pelatihan model dilakukan menggunakan *cross-validation* dengan 10-fold untuk menghindari *overfitting* dan memastikan generalisasi model.

### 5. Evaluasi Model

Kinerja model dievaluasi menggunakan metrik berikut:

5.1. Akurasi: Persentase prediksi yang benar.

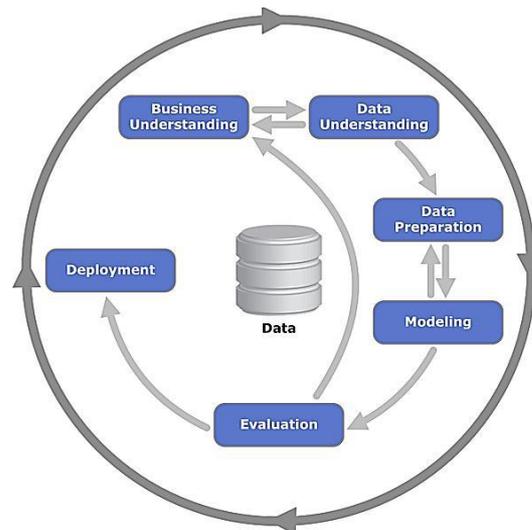
5.2. *Area Under the Curve (AUC)*: Mengukur kemampuan model untuk membedakan antara kelas positif dan negatif.

5.3. *Confusion Matrix*: Untuk menghitung nilai *Sensitivity*, *Specificity*, *Precision*, dan *Recall*.

Kurva ROC digunakan untuk memvisualisasikan kinerja model dan menghitung nilai AUC untuk setiap model.

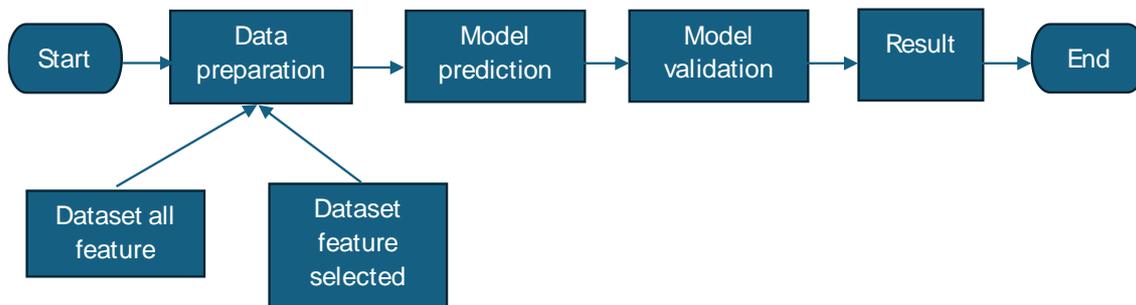
## HASIL DAN PEMBAHASAN

Penelitian ini dikembangkan berdasarkan model CRISP-DM (*Cross Standard Industries Process for Data Mining*) yang memiliki 6 tahapan, yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, Deployment. Berikut ini adalah gambaran model penelitian :



Gambar 1. Model CRISP-DM

Kerangka penelitian yang dilakukan kali ini digambar dalam diagram berikut ini:



Gambar 2. Kerangka Penelitian

Dibawah ini adalah deskripsi dari kerangka penelitian yang dilakukan:

1. Data preparation: tahapan ini mempersiapkan data, mulai dari data *collecting*, data *cleaning*, dan data *transforming*. Kemudian dipilih atribut yang memiliki relevansi terhadap penelitian. Lalu data tersebut akan menjadi dataset *all feature* (dataset dengan semua atribut yang memiliki relevansi). Dataset ini kemudian dinilai bobotnya menggunakan *Gain Ratio Attribute Evaluator* untuk dipilih atribut yang memiliki bobot tinggi atau pengaruh terhadap target kelas, dataset ini kemudian menjadi dataset *feature selection*.
2. Model prediction: tahapan ini akan membuat model untuk prediksi kelulusan mahasiswa berdasarkan dataset yang telah disiapkan. Pembuatan model menggunakan algoritma *Random Forest*, *SVM*, *Logistic Regression*, dan *KNN*. Setiap model menggunakan dataset *all feature* dan dataset *feature selection*.
3. Model validation: tahapan ini akan mengevaluasi kinerja model yang dihasilkan dari tahap sebelumnya. Tahapan ini akan menghasilkan *confusion matrix*, akurasi, AUC, dan kurva ROC.
4. Result: tahapan ini akan menyajikan hasil dari setiap model yang telah dibuat.

Atribut yang memiliki relevansi terhadap penelitian adalah: jumlah\_saudara, sumber\_dana, jalur\_masuk, kode\_prodi, ip\_rata, usia, status\_lulus. Atribut ip\_rata adalah nilai rata-rata indeks prestasi semester 1 sampai semester 6. Berikut adalah deskripsi dari masing-masing atribut:

Tabel 1. Atribut Dataset All Feature

No	Atribut	Nilai	Nilai baru
1	jumlah_saudara	0 – 2	SEDIKIT
		3 – 5	SEDANG
		> 5	BANYAK
2	sumber_dana	Lain-lain	Lain-lain
		Orang Tua Kandung	Orang Tua Kandung
		Orang Tua Angkat	Orang Tua Angkat
		Orang Tua Asuh	Orang Tua Asuh
		Saudara Kandung	Saudara Kandung
		Saudara Bukan Kandung	Saudara Bukan Kandung
3	Jalur_masuk	SMMPN	SMMPN
		SNMPN	SNMPN
		SBMPN	SBMPN
		UMBPT	UMBPT
		UMM	UMM
		JPL	JPL
		BTT	BTT
		ADIK	ADIK

**Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Algoritma  
Machine Learning Dan Feature Selection**

4	Kode_prodi	1111	1111
		2221	2221
		2222	2222
		2223	2223
		2224	2224
		2225	2225
		2227	2227
		2228	2228
		2280	2280
		2281	2281
		2282	2282
		2283	2283
		2284	2284
		2285	2285
		2286	2286
		2287	2287
		2288	2288
		2289	2289
		2290	2290
		2291	2291
		3331	3331
		3332	3332
		3333	3333
		3334	3334
		3335	3335
		3336	3336
		4441	4441
		4442	4442
		4443	4443
		4444	4444
		5503	5503
		5504	5504
		5551	5551
5552	5552		
5553	5553		
5554	5554		
6661	6661		
6662	6662		
6670	6670		
7777	7777		
7779	7779		
8881	8881		
8882	8882		
5	Ip_rata	$\geq 3.7$	A
		$\geq 3.3$	A-
		$\geq 3.0$	B+
		$\geq 2.7$	B
		$\geq 2.3$	B-
		$\geq 2.0$	C+
		$\geq 1.5$	C
		$\geq 1.0$	D
	$< 1.0$	E	
6	Usia	16 – 17	REMAJA1
		18 – 20	REMAJA2
		21 – 25	REMAJA3

7	Status_lulus	TEPAT TERLAMBAT	TEPAT TERLAMBAT
---	--------------	--------------------	--------------------

Seluruh atribut tadi dinilai bobotnya untuk mencari nilai yang paling berpengaruh terhadap kelulusan dalam hal ini adalah target kelas status\_lulus. Berdasarkan penilaian, atribut yang dipilih adalah ip\_rata, jalur\_masuk, dan kode\_prodi. Berikut ini adalah hasil evaluasi menggunakan bantuan Tools WEKA dengan Gain Ratio Atribut Evaluator dengan method Ranking:

```

=== Attribute Selection on all input data ===

Search Method:
    Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 7 status_lulus):
    Gain Ratio feature evaluator

Ranked attributes:
0.084297  5 ip_rata
0.034798  3 Jalur Masuk
0.01841   4 Kode Prodi
0.002213  2 Sumber Dana
0.001231  6 usia
0.00028   1 i»Jumlah Saudara

Selected attributes: 5,3,4,2,6,1 : 6
    
```

Gambar 3. Gain Ratio Attribute Evaluator

Tabel 2. Atribut Dataset Feature Selection

No	Atribut	Bobot
1	Ip_rata	0.084297
2	Jalur_masuk	0.034798
3	Kode_prodi	0.01841

Untuk membuat prediksi kelulusan mahasiswa tepat waktu menggunakan empat algoritma machine learning: Random Forest, Support Vector Machine (SVM), Logistic Regression, dan K-Nearest Neighbors (KNN). Bahasa Pemrograman yang digunakan dalam penelitian menggunakan Bahasa R dengan IDE R Studio. Setiap algoritma diuji menggunakan dataset lengkap dan dataset yang telah mengalami seleksi fitur menggunakan Gain Ratio Attribute Evaluator. Berikut adalah ringkasan hasil evaluasi masing-masing model:

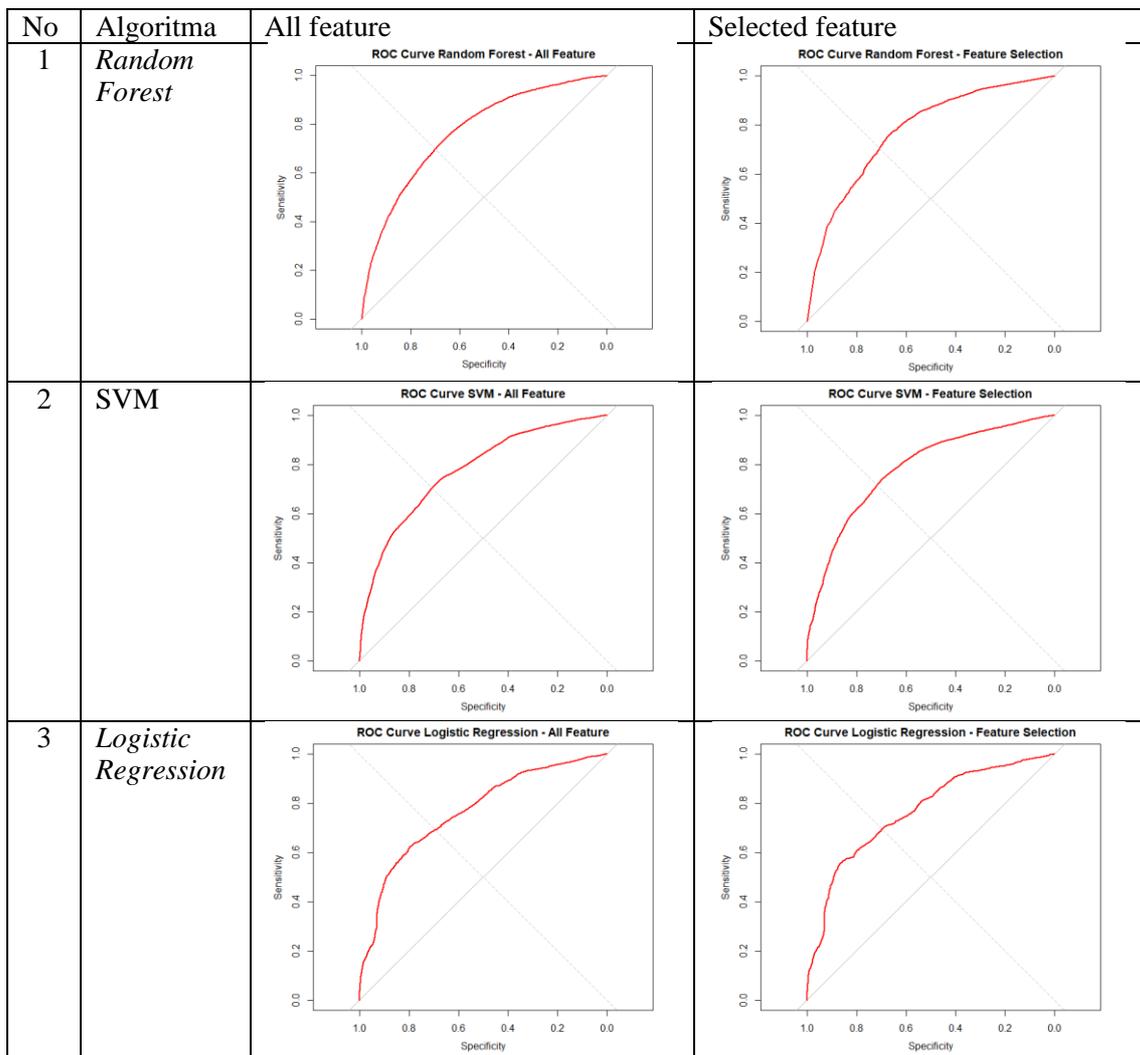
Algoritma	Akurasi	AUC	Sensitivity	Specificity	Kappa
Random Forest (all features)	70.52%	0.7672	65.22%	74.48%	0.3974
Random Forest (selected features)	72.00%	0.7741	65.73%	76.69%	0.4259
SVM (all features)	69.80%	0.774	54.82%	81.01%	0.3675
SVM (selected features)	72.28%	0.7812	64.34%	78.21%	0.4291
Logistic Regression (all features)	69.25%	0.7645	66.19%	71.53%	0.3753
Logistic Regression (selected features)	69.14%	0.7644	66.40%	71.20%	0.3737
KNN (all features)	72.83%	0.8071	68.16%	76.33%	0.4451

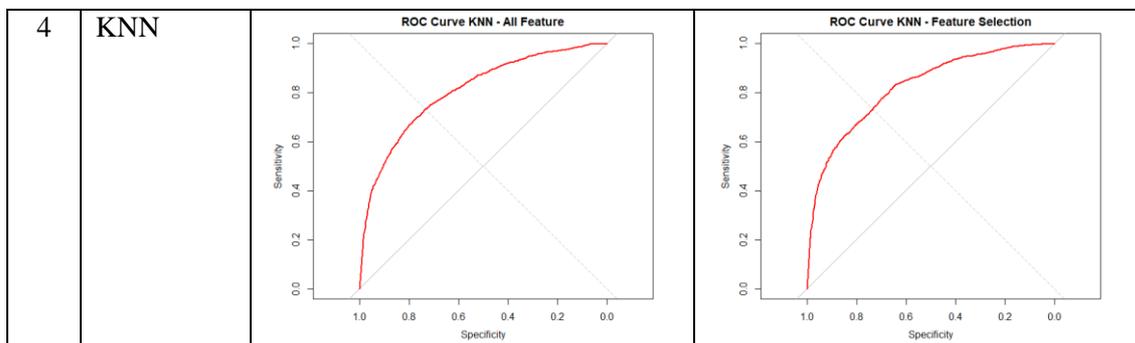
## Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Algoritma Machine Learning Dan Feature Selection

<b><i>KNN (selected features)</i></b>	74.44%	0.8212	66.62%	80.30%	0.4734
---------------------------------------	--------	--------	--------	--------	--------

Tabel 2. Evaluasi masing masing model

Untuk mengevaluasi kinerja model klasifikasi, alat grafis yang digunakan adalah kurva karakteristik operasi penerima (ROC). *True Positive Rate (Sensitivity)* di sumbu Y dan *False Positive Rate (1—Specificity)* di sumbu X ditunjukkan pada kurva ini pada berbagai threshold klasifikasi. *Sensitivity (True Positive Rate)* merupakan kemampuan model untuk mengidentifikasi kelas yang positif—dalam hal ini, siswa yang lulus tepat waktu. Nilai-nilai ini ditampilkan pada sumbu Y. sedangkan *Specificity (True Negative Rate)*: Kemampuan model untuk mengidentifikasi kelas negatif, yaitu siswa yang tidak lulus tepat waktu. Pada sumbu X, kurva ROC menunjukkan  $(1 - Specificity)$ . Berikut ini adalah gambar kurva ROC untuk tiap model:





Tabel 3. Kurva ROC

Hasil penelitian ini menunjukkan bahwa seleksi fitur dapat secara signifikan meningkatkan kinerja prediksi dari berbagai algoritma machine learning yang digunakan. Berikut adalah pembahasan lebih lanjut mengenai temuan ini.

1. **Peningkatan Akurasi dan AUC**

- 1.1. **KNN** menunjukkan peningkatan yang paling signifikan setelah seleksi fitur, dengan akurasi meningkat dari 72.83% menjadi 74.44% dan AUC dari 0.8071 menjadi 0.8212. Hal ini menunjukkan bahwa fitur-fitur yang dipilih memiliki kontribusi yang lebih relevan terhadap prediksi kelulusan tepat waktu dibandingkan dengan menggunakan semua fitur.
- 1.2. **SVM** juga menunjukkan peningkatan yang signifikan dalam hal akurasi dan AUC setelah seleksi fitur. Akurasi meningkat dari 69.80% menjadi 72.28%, dan AUC dari 0.774 menjadi 0.7812, yang menunjukkan bahwa model menjadi lebih mampu membedakan antara mahasiswa yang lulus tepat waktu dan yang terlambat.
- 1.3. **Random Forest** dan **Logistic Regression** juga mengalami peningkatan, meskipun tidak sebesar KNN dan SVM. Hal ini menunjukkan bahwa beberapa fitur yang tidak relevan dapat mengurangi kinerja model, dan pemilihan fitur yang tepat dapat membantu meningkatkan akurasi prediksi.

2. **Sensitivity dan Specificity**

- 2.1. **Sensitivity** (kemampuan untuk mengidentifikasi mahasiswa yang lulus tepat waktu) dan **Specificity** (kemampuan untuk mengidentifikasi mahasiswa yang tidak lulus tepat waktu) juga meningkat pada sebagian besar model setelah seleksi fitur. KNN dan SVM menunjukkan peningkatan yang signifikan dalam specificity setelah seleksi fitur, yang berarti bahwa model menjadi lebih baik dalam mengidentifikasi mahasiswa yang tidak lulus tepat waktu.
- 2.2. **Random Forest** menunjukkan peningkatan yang seimbang dalam sensitivity dan specificity, menunjukkan bahwa model ini menjadi lebih baik dalam mengidentifikasi kedua kelas setelah seleksi fitur.

3. **Kappa**

Nilai **Kappa** untuk semua model meningkat setelah seleksi fitur, menunjukkan bahwa ada peningkatan dalam tingkat kesepakatan antara prediksi model dan hasil aktual. Peningkatan ini menunjukkan bahwa model menjadi lebih akurat dan konsisten setelah fitur yang tidak relevan dihapus.

Berdasarkan data yang diperoleh, penggunaan algoritma KNN dan SVM menunjukkan peningkatan dengan menggunakan *feature selection*. Namun penggunaan algoritma KNN menunjukkan model dengan kinerja yang sangat baik. Nilai AUC (*Area Under the Curve*) yang sebesar sebesar 0.8212 dan Kurva ROC yang mendekati nilai 1, yang menunjukkan bahwa model KNN memiliki kemampuan yang sangat baik dalam membedakan antara kelas positif dan negatif.

**KESIMPULAN**

Penelitian ini menemukan bahwa seleksi fitur meningkatkan kinerja prediksi model *machine learning* untuk kelulusan mahasiswa tepat waktu. Model KNN dan SVM menunjukkan

## **Analisis Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Algoritma Machine Learning Dan Feature Selection**

peningkatan paling signifikan, dengan akurasi dan AUC yang lebih tinggi setelah seleksi fitur. Hasil ini menunjukkan pentingnya pemilihan fitur yang relevan untuk meningkatkan akurasi dan efektivitas model prediksi. Implementasi seleksi fitur dalam EDM dapat membantu institusi pendidikan dalam mengambil keputusan yang lebih tepat dan meningkatkan tingkat kelulusan mahasiswa tepat waktu.

### **SARAN**

Berdasarkan hasil penelitian yang telah dilakukan, berikut adalah beberapa saran yang dapat dipertimbangkan untuk implementasi lebih lanjut:

#### **1. Peningkatan Penggunaan Seleksi Fitur**

Seleksi fitur telah terbukti meningkatkan kinerja model prediksi kelulusan tepat waktu. Oleh karena itu, sangat disarankan untuk selalu mengimplementasikan teknik seleksi fitur seperti Gain Ratio Attribute Evaluator dalam proses data mining di bidang pendidikan. Penggunaan seleksi fitur dapat membantu mengidentifikasi atribut yang paling relevan dan mengurangi kompleksitas model, sehingga meningkatkan akurasi dan efisiensi prediksi.

#### **2. Pemilihan Algoritma yang Tepat**

Berdasarkan hasil penelitian, algoritma KNN dan SVM menunjukkan peningkatan yang signifikan setelah seleksi fitur. Institusi pendidikan harus mempertimbangkan penggunaan algoritma ini dalam sistem prediksi mereka. Selain itu, penting untuk melakukan evaluasi kinerja berbagai algoritma secara berkala untuk memastikan model yang digunakan tetap optimal.

#### **3. Pengembangan Sistem Prediksi yang Terintegrasi**

Untuk memaksimalkan manfaat dari machine learning, institusi pendidikan harus mengembangkan sistem prediksi yang terintegrasi dengan database akademik mereka. Sistem ini harus mampu mengolah data secara real-time dan memberikan rekomendasi atau intervensi yang tepat bagi mahasiswa yang berisiko tidak lulus tepat waktu. Implementasi sistem prediksi yang terintegrasi dapat membantu dalam pengambilan keputusan yang lebih cepat dan tepat.

#### **4. Pelatihan dan Pengembangan Staf**

Penerapan machine learning dalam pendidikan membutuhkan keahlian khusus. Oleh karena itu, penting untuk memberikan pelatihan kepada staf pengajar dan administrasi tentang penggunaan dan interpretasi model prediksi. Pelatihan ini dapat mencakup dasar-dasar machine learning, cara kerja algoritma yang digunakan, dan interpretasi hasil prediksi untuk pengambilan keputusan yang lebih baik.

#### **5. Penelitian Lanjutan**

Penelitian ini menunjukkan hasil yang menjanjikan, namun masih ada ruang untuk perbaikan dan pengembangan lebih lanjut. Penelitian lanjutan dapat fokus pada eksplorasi teknik seleksi fitur lainnya, penggunaan data tambahan seperti data non-akademik, dan pengembangan model prediksi yang lebih kompleks dan akurat. Selain itu, studi longitudinal yang melibatkan analisis data dalam jangka waktu yang lebih panjang dapat memberikan wawasan lebih mendalam tentang faktor-faktor yang mempengaruhi kelulusan tepat waktu.

#### **6. Kolaborasi Antar Institusi**

Kolaborasi antar institusi pendidikan dapat memberikan manfaat yang besar dalam pengembangan dan penerapan model prediksi. Berbagi data, metode, dan hasil penelitian dapat mempercepat kemajuan dalam bidang ini dan memastikan bahwa praktik terbaik dapat diterapkan secara luas. Institusi dapat membentuk konsorsium atau forum untuk berbagi pengetahuan dan pengalaman terkait EDM.

#### **7. Kebijakan dan Regulasi**

Penerapan *machine learning* dalam pendidikan harus disertai dengan kebijakan dan regulasi yang jelas untuk melindungi privasi dan keamanan data mahasiswa. Institusi pendidikan harus memastikan bahwa data yang digunakan dalam model prediksi diproses secara etis dan sesuai dengan peraturan yang berlaku. Selain itu, transparansi

dalam penggunaan model prediksi harus dijaga untuk memastikan kepercayaan dari semua pemangku kepentingan.

## UCAPAN TERIMA KASIH

Pertama penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) Universitas Sultan Ageng Tirtayasa yang telah mendanai kegiatan ini melalui Skema Penelitian Dosen Pemula (PDP) tahun 2024, kedua ditunjukkan kepada UPA TIK yang telah membantu dalam menyediakan data dan rekan rekan yang telah membantu support dan sarannya sehingga penelitian ini bisa kami selesaikan

## DAFTAR PUSTAKA

- Aditya Nirwana, Sudarmiati, & Melany. (2023). Implementation of Artificial Intelligence in Digital Marketing Development: a Thematic Review and Practical Exploration. *Jurnal Manajemen Bisnis, Akuntansi Dan Keuangan*, 2(1), 85–112. <https://doi.org/10.55927/jambak.v2i1.4034>
- Awalina, E. F. L., & Rahayu, W. I. (2023). Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail. *Jurnal Teknologi Dan Informasi*, 13(2), 122–137. <https://doi.org/10.34010/jati.v13i2.10090>
- Du, X., Yang, J., Hung, J. L., & Shelton, B. (2020). Educational data mining: a systematic review of research and emerging trends. *Information Discovery and Delivery*, 48(4), 225–236. <https://doi.org/10.1108/IDD-09-2019-0070>
- Namoun, A., & Alshantqiti, A. (2021). Predicting student performance using data mining and learning analytics techniques: A systematic literature review. *Applied Sciences (Switzerland)*, 11(1), 1–28. <https://doi.org/10.3390/app11010237>
- Situmorang, S. (2023). Analisis Kinerja Algoritma Machine Learning Dalam Deteksi Anomali Jaringan (LAZY LEARNING). *Jurnal Matematika Dan Ilmu Pengetahuan Alam*, 1(4), 259–269. <https://doi.org/10.59581/konstanta.v1i4.1722>
- Sukarna, R. H., & Ansori, Y. (2022). Implementasi Data Mining Menggunakan Metode Naive Bayes Dengan Feature Selection Untuk Prediksi Kelulusan Mahasiswa Tepat Waktu. *Jurnal Ilmiah Sains Dan Teknologi*, 6(1), 50–61. <https://doi.org/10.47080/saintek.v6i1.1467>
- Tekin, A. (2014). Early Prediction of Students' Grade Point Averages at Graduation: A Data Mining Approach. *Eurasian Journal of Educational Research*, 14(54), 207–226. <https://doi.org/10.14689/ejer.2014.54.12>
- Thakar, P., Mehta, A., & Manisha. (2015). Performance analysis and prediction in educational data mining: A research travelogue. *International Journal of Computer Applications*, 110(15), 60–68.
- Uzlah, L. I., Saputra, R. A., & Isnawaty. (2024). Deteksi Serangan Siber Pada Jaringan Komputer Menggunakan Metode Random Forest. *Jurnal Mahasiswa Teknik Informatika*, 8(3), 2787–2793. <https://bit.ly/CyberSecurityAttacks>.
- Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., Alelyani, S., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*, 104. <https://doi.org/10.1016/j.chb.2019.106189>
- Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, 47, 168–181. <https://doi.org/10.1016/j.chb.2014.09.034>