

# PERBANDINGAN METODE KLASIFIKASI RANDOM FOREST DAN SUPPORT VECTOR MACHINE DALAM MEMPREDIKSI CAPAIAN STUDI MAHASISWA

Enggar Novianto\*<sup>1)</sup>, Suhirman Suhirman<sup>2)</sup>, Damar Prasetyo<sup>3)</sup>

1. Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Indonesia
2. Magister Teknologi Informasi, Universitas Teknologi Yogyakarta, Indonesia
3. Sistem Informasi, Universitas Teknologi Yogyakarta, Indonesia

## Article Info

**Kata Kunci:** Klasifikasi; Mahasiswa; *Random Forest*; *Support Vector Machine*

**Keywords:** *Classification*; *Random Forest*; *Support Vector Machine*; *Student*

## Article history:

Received 29 September 2024

Revised 13 Oktober 2024

Accepted 4 November 2024

Available online 4 December 2024

## DOI :

<https://doi.org/10.29100/jipi.v9i4.5423>

\* Corresponding author.

Enggar Novianto

E-mail address:

[6220211003.enggar@student.uty.ac.id](mailto:6220211003.enggar@student.uty.ac.id)

## ABSTRAK

Keberhasilan Universitas, swasta dan negeri bergantung pada mahasiswa dan untuk mengurangi tingkat kegagalan akademik, diperlukan sistem yang dapat memprediksi mahasiswa berdasarkan data akademik serta membuat penilaian untuk memprediksi seberapa baik capaian studi mahasiswa. Data program studi dapat diolah dengan cepat dan akurat, dan data mining adalah proses penambangan data untuk membuat prediksi capaian studi berdasarkan data tentang mahasiswa. Salah satu keunggulan dari penelitian ini adalah penggunaan prosedur untuk mengoptimalkan model RF dan SVM serta pembentukan fitur yang berdampak pada akurasi prediksi capaian studi mahasiswa Fakultas Hukum Universitas Sebelas Maret dalam program studi S1 Ilmu Hukum. Penelitian ini menggunakan metode seleksi fitur Forward Selection. Pemodelan menggunakan RF sebelum dilakukan seleksi fitur mendapatkan hasil akurasi sebesar 97,67%, sedangkan pemodelan menggunakan SVM mendapatkan hasil akurasi sebesar 91,47% dengan menggunakan data mahasiswa angkatan tahun 2021 sejumlah 433 data dengan pembagian 70% data latih dan 30% data uji. Penggunaan seleksi fitur menggunakan metode Forward Selection tidak dapat meningkatkan hasil akurasi pada algoritma RF serta menghasilkan empat atribut yang berpengaruh pada klasifikasi prediksi capaian studi mahasiswa. Pada pemodelan SVM, seleksi fitur dapat meningkatkan nilai akurasi sebesar 6,2%, sehingga hasil akurasi SVM setelah dilakukan seleksi fitur adalah sebesar 97,67% dengan menghasilkan satu atribut yang berpengaruh pada klasifikasi prediksi capaian studi mahasiswa. Perbandingan metode klasifikasi RF dan SVM setelah dilakukan seleksi fitur mendapatkan akurasi yang sama yaitu 97,67%, oleh karena itu, hasil penelitian ini termasuk dalam kategori model yang cukup. Hasil penelitian dapat menjadi acuan bagi pengelola program studi dalam memberikan perlakuan khusus kepada mahasiswa yang diprediksi tidak tercapai pembelajarannya.

## ABSTRACT

The success of universities, private and public depends on students and to reduce the rate of academic failure, a system is needed that can predict students based on academic data and make assessments to predict how well students will achieve in their studies. Study program data can be processed quickly and accurately, and data mining is a data mining process to make predictions about study outcomes based on data about students. The novelty of this research is that it uses a process to optimize the RF and SVM models and produces attributes that influence accuracy in predicting study outcomes for undergraduate students in the Legal Sciences study program, Faculty of Law, Sebelas Maret University by selecting features using Forward Selection. Modeling using RF before feature selection got an accuracy result of 97.67%, while modeling using SVM got an accuracy result of 91.47% using 433 student data from the class of 2021 with a division of 70% training data and 30% test data. The use of feature selection using the Forward Selection method cannot improve the accuracy results of the RF algorithm and produces four attributes that influence the classification of student study achievement predictions. In SVM modeling, feature selection can increase the accuracy value

by 6.2%, so that the SVM accuracy result after feature selection is 97.67% by producing one attribute that influences the prediction classification of student study achievement. Comparison of the RF and SVM classification methods after feature selection obtained the same accuracy, namely 97.67%, therefore, the results of this study are included in the adequate model category. The results of the research can be a reference for study program managers in providing special treatment to students whose learning is predicted to fail.

## I. PENDAHULUAN

**P**ERGURUAN Tinggi Negeri (PTN) dan Perguruan Tinggi Swasta (PTS) bersaing ketat untuk menghasilkan lulusan berkualitas tinggi seiring dengan perkembangan dunia pendidikan di Indonesia [1]. Rata-rata hasil belajar mahasiswa pada tahun akademik pertama menimbulkan banyak masalah yang akan berdampak pada keberhasilan program studi perguruan tinggi. Indeks Prestasi Kumulatif (IPK) digunakan untuk menilai kinerja akademik mahasiswa [2]. Kualitas pembelajaran dapat didefinisikan sebagai ukuran pencapaian tujuan pembelajaran utama. Rencana awal perkuliahan termasuk seni penyampaian materi sampai dengan peningkatan pengetahuan, keterampilan, dan sikap mahasiswa. Salah satu faktor yang memengaruhi keberhasilan mahasiswa di sekolah adalah motivasi mereka untuk belajar. Selama menjadi mahasiswa, seseorang harus belajar secara mandiri dan aktif mencari informasi, bukan hanya belajar dari dosen. Menjadi mahasiswa mandiri akan sangat sulit jika mahasiswa tidak termotivasi untuk mencapai tujuan akademik mereka [3][4]. Perguruan Tinggi sendiri menghasilkan ilmu pengetahuan yang kreatif, yang diperlukan untuk mendorong para ahli menghasilkan pengetahuan relevan dengan situasi berbeda, dan transisi serta perkembangan pengetahuan di berbagai Negara [5].

Keberhasilan universitas, baik swasta maupun negeri, bergantung pada mahasiswa. Untuk mengurangi tingkat kegagalan akademik, diperlukan sistem yang memprediksi mahasiswa berdasarkan data akademik dan memungkinkan mahasiswa membuat penilaian untuk memprediksi seberapa baik mahasiswa dalam menyelesaikan mata kuliah [6]. Kinerja mahasiswa di institusi pendidikan tinggi sangat penting, disebabkan fakta bahwa salah satu karakteristik universitas yang berkualitas tinggi adalah catatan akademik yang luar biasa. Menurut penelitian sebelumnya, ada banyak definisi kinerja mahasiswa. Universitas memiliki banyak variabel, salah satunya adalah keberhasilan mahasiswa yang telah menyelesaikan pelajaran dalam waktu yang telah ditentukan [7].

Data program studi dapat diolah dengan cepat dan akurat, dan data mining adalah proses penambangan data untuk membuat prediksi capaian studi berdasarkan data tentang mahasiswa. Data tentang mahasiswa dapat memberikan informasi yang berguna bagi program studi jika di manfaatkan dengan maksimal [8][9]. Berdasarkan Permendikbud No. 49/2014 tentang Standar Nasional Pendidikan Tinggi, beban belajar mahasiswa diukur dalam Satuan Kredit Semester (SKS). SKS adalah satuan waktu kegiatan pembelajaran yang berlangsung selama minimal 16 (enam belas) minggu, termasuk ujian tengah dan akhir semester. Satu tahun pendidikan terdiri dari dua semester, dan perguruan tinggi dapat mengadakan semester tambahan di antaranya. Pembelajaran terjadi melalui penggunaan sumber belajar dalam lingkungan belajar tertentu dan interaksi dosen dengan mahasiswa. Sangat penting untuk menggunakan metode pembelajaran yang sesuai dengan karakteristik mata kuliah untuk mencapai kemampuan tertentu yang ditetapkan dalam mata kuliah [10]. Dalam buku kurikulum pendidikan tinggi disebutkan bahwa salah satu proses pembelajaran yang baik adalah pemahaman yang jelas tentang capaian pembelajaran. Capaian pembelajaran adalah hasil akhir dari pembelajaran yang disimpan dan memberikan perubahan, seperti perubahan cara berpikir dan berperilaku [11]. Jumlah mahasiswa yang terdapat di program studi S1 Ilmu Hukum Fakultas Hukum hingga tahun 2024 ini sebanyak kurang lebih 2052 mahasiswa aktif. Banyaknya jumlah mahasiswa membuat pimpinan program studi kesulitan dalam memprediksi capaian studi mahasiswa saat ini. Pencapaian studi mahasiswa merupakan aspek penting yang di miliki oleh program studi untuk mempertahankan dan meningkatkan kualitas pendidikan. Semakin bertambah banyaknya mahasiswa dalam setiap tahun pelajaran baru memerlukan teknologi informasi yang bisa digunakan untuk memprediksi capaian studi mahasiswa dan dapat memberikan rekomendasi serta keputusan yang dapat di ambil oleh pimpinan program studi.

Dalam penerapannya pada penambangan data terdapat beberapa metode algoritma klasifikasi diantaranya yang dapat diimplementasi adalah algoritma *Random Forest* (RF) dan *Support Vector Machine* (SVM). Metode *random forest* adalah model klasifikasi yang menggunakan seleksi data dan variabel acak untuk menghasilkan beberapa pohon keputusan. Sekumpulan pohon acak yang dihasilkan dari metode ini adalah kelas yang paling banyak dibuat oleh pohon acak yang ada [12]. Selama proses klasifikasi, individu-individu dipilih berdasarkan suara terbanyak dari kumpulan populasi pohon. *Random forest* adalah pengembangan dari metode CART yang menggunakan metode *aggregating bootstrap* (bagging) dan random feature selection. *Random forest* memiliki beberapa kelebihan, seperti kemampuan untuk meningkatkan akurasi dalam kasus data yang hilang, kemampuan untuk

mencegah kesalahan, kemampuan untuk menyimpan data dengan efisien, dan kemampuan untuk menahan keluaran, serta penyimpanan data yang efisien. Selain itu, lereng acak memiliki pilihan fitur yang memungkinkan untuk mengambil fitur terbaik, yang dapat meningkatkan kinerja terhadap model klasifikasi [13]. Sedangkan metode *Support Vector Machine* adalah teknik perhitungan data dengan menggunakan pembelajaran dengan yang diharapkan dan statistik mengembangkan kemampuan untuk memprediksi, dengan bantuan kernel SVM menghubungkan input pelatihan ke dimensi ruang yang lebih luas dan mengidentifikasi *hyperplane* sebagai ruang pemisah [14]. Metode ini sangat akurat karena kemampuan untuk menangani model nonlinier yang kompleks. Prediksi dan klasifikasi dapat dilakukan dengan SVM [15]. SVM memiliki kelebihan diantaranya adalah memiliki kemampuan generalisasi yang baik, mampu menghasilkan model klasifikasi yang baik meskipun dilatih dengan himpunan data yang relative kecil sedangkan kelemahan SVM adalah sulit diaplikasikan untuk himpunan data dengan jumlah sampel dan dimensi yang sangat besar [16].

Penelitian ini menggunakan seleksi fitur. Seleksi fitur adalah proses yang umum digunakan dalam pembelajaran mesin di mana sekumpulan fitur yang dimiliki oleh data digunakan untuk pembelajaran algoritma [17]. Untuk mengurangi dimensi model, seleksi fitur sering digunakan. Pemilihan fitur membantu mengurangi jumlah fitur domain dan mengurangi jumlah fitur yang berlebihan [18]. Seleksi fitur yang digunakan pada penelitian ini adalah *Forward Selection*. adalah salah satu teknik permodelan pembangunan model linier yang digunakan untuk menentukan kombinasi peubah terbaik dari gugus peubah. Selama prosedur ini, variabel yang masuk ke dalam persamaan tidak dapat dihilangkan. Selain itu, variabel bebas yang memiliki korelasi yang paling erat dengan variabel tak bebasnya dan variabel yang paling mungkin memiliki hubungan linier dengan Y juga dapat dimasukkan melalui pilihan maju [19] Di mana hanya fitur berkualitas tinggi yang digunakan untuk model [20].

Beberapa penelitian terdahulu sudah pernah dilakukan dengan menggunakan metode *K-Nearest Neighbor* (K-NN) untuk memprediksi masa belajar mahasiswa, penelitian ini menghasilkan akurasi sebesar 93,2% [5]. Penelitian lainnya juga menggunakan metode K-NN untuk memprediksi kinerja mahasiswa berdasarkan nilai, hasil dari penelitian ini menghasilkan nilai akurasi sebesar 94,3% [21]. Penelitian lainnya yaitu membandingkan dua metode yaitu *Support Vector Machine* (SVM) dan C4.5, penelitian ini menghasilkan SVM dengan akurasi 85% dan C4.5 80% [22].

Penelitian ini dilakukan pengujian perbandingan kinerja nilai akurasi algoritma *Random Forest* (RF) dan *Support Vector Machine* (SVM) sebagai metode klasifikasi untuk memprediksi capaian studi mahasiswa. Kebaharuan dari penelitian ini adalah menggunakan proses untuk mengoptimalkan model RF dan SVM serta menghasilkan atribut yang berpengaruh terhadap akurasi dalam memprediksi capaian studi mahasiswa program studi S1 Ilmu Hukum Fakultas Hukum Universitas Sebelas Maret Surakarta dengan seleksi fitur menggunakan *Forward Selection*. FS

## II. METODE PENELITIAN

### 2.1 Data Mining

Data mining adalah metode analisis data yang digunakan untuk mengekstrak informasi dari kumpulan data yang besar. Ini melibatkan berbagai proses seperti pengumpulan data, pengumpulan dan integrasi data, pengumpulan dan ekstraksi data, transformasi data, dan analisis data. Penambangan data juga melibatkan konversi data, yang merupakan proses algoritmik dan evaluasi data [23] [24]. Penambangan data adalah proses interaktif dan berulang yang digunakan untuk membuat model ideal dan dapat dibaca dalam *database* besar, membantu dalam pengambilan keputusan dan dianalisis dengan teknik khusus untuk pembelajaran yang lebih baik [25].

### 2.2 Klasifikasi

Klasifikasi adalah metode matematika seperti pemrograman linier, pohon keputusan, jaringan syaraf tiruan, dan statistik, yang digunakan untuk mengkategorikan setiap item dalam data pembelajaran mesin [26]. Klasifikasi termasuk metode pembelajaran terawasi karena memerlukan data pelatihan untuk membuat aturan yang mengklasifikasikan data pengujian ke dalam kelas atau kelompok tertentu. Teknik klasifikasi juga dikenal sebagai teknik pengklasifikasi adalah pendekatan sistematis untuk membuat model klasifikasi dari sekumpulan data input [27].

### 2.3 Random Forest

Adalah metode pembelajaran mesin yang digunakan untuk memecahkan masalah yang membutuhkan banyak data. Algoritma random forest adalah salah satu dari banyak algoritma berbasis pohon yang diusulkan oleh Tin Kam Ho pada tahun 1995 yang merupakan kombinasi dari  $x$  pohon dan dijadikan satu model [28]. Metode random

forest adalah model klasifikasi yang menggunakan seleksi data dan variabel acak untuk menghasilkan beberapa pohon keputusan. Sekumpulan pohon acak yang dihasilkan dari metode ini adalah kelas yang paling banyak dibuat oleh pohon acak yang ada [12]. Random forest terdiri dari sekelompok pohon regresi atau klasifikasi yang tidak dipangkat yang dibuat dengan menggunakan sampel acak dari data, dan memiliki keunggulan seperti kemampuan untuk mendeteksi kesalahan yang relatif besar, kinerja klasifikasi yang baik, kemampuan untuk menangani data dengan jumlah sampel yang lebih kecil, dan metode yang efektif untuk mengestimasi data yang sangat kecil [29].

## 2.4 Support Vector Machine

SVM adalah teknik perhitungan data dengan menggunakan pembelajaran yang diharapkan dan statistik mengembangkan kemampuan untuk memprediksi, dengan bantuan kernel SVM menghubungkan input pelatihan ke dimensi ruang yang lebih luas dan mengidentifikasi *hyperplane* sebagai ruang pemisah [14] SVM memiliki karakteristik *linier classifier*, metode ini dibuat pada tahun 1995. SVM dapat memecahkan masalah prediksi dalam kasus regresi dan klasifikasi yang sangat populer saat ini [30].

## 2.5 Feature Selection

Adalah proses yang umum digunakan dalam pembelajaran mesin di mana sekumpulan fitur yang dimiliki oleh data digunakan untuk pembelajaran algoritma [17]. Untuk menyempit atau mengurangi dimensi antar atribut, seleksi fitur adalah teknik yang umum digunakan untuk menghasilkan beberapa atribut yang dianggap relevan untuk proses [31]. *Feature Selection* membantu memilih fitur yang penting (fitur optimal) dan mengesampingkan fitur yang tidak penting, sehingga proses pemodelan lebih cepat [32].

## 2.6 Forward Selection

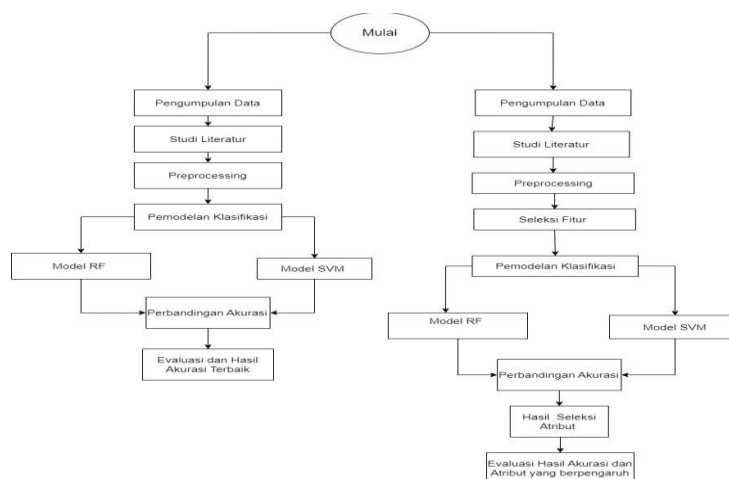
Adalah salah satu teknik permodelan pembangunan model linier yang digunakan untuk menentukan kombinasi peubah terbaik dari gugus peubah. Selama prosedur ini, variabel yang masuk ke dalam persamaan tidak dapat dihilangkan. Selain itu, variabel bebas yang memiliki korelasi yang paling erat dengan variabel tak bebasnya dan variabel yang paling mungkin memiliki hubungan linier dengan Y juga dapat dimasukkan melalui pilihan maju [19].

## 2.7 RapidMiner

Rapidminer adalah perangkat lunak sumber terbuka yang memungkinkan penambangan data, penambangan teks, dan analisis prediksi. Teknik deskriptif dan prediktif meningkatkan proses pengambilan keputusan. RapidMiner ditulis dalam jawa dan memiliki lebih dari 500 operator penambangan data, dan mudah digunakan, dapat diintegrasikan ke dalam sistem produksi apa pun [30]. Dengan menggunakan RapidMiner, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. RapidMiner dikhususkan untuk penggunaan data mining [33].

## 2.8 Tahapan Alur Penelitian

Penelitian ini dilakukan dalam beberapa tahapan yaitu pengumpulan data, studi literatur, *preprocessing*, proses seleksi fitur, perbandingan nilai akurasi dan evaluasi hasil. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar. 1. Tahapan Penelitian

## 2.9 Pengumpulan Data

Penelitian ini menggunakan data primer yang didapatkan dengan observasi melalui sistem informasi akademik. Data mahasiswa yang digunakan sejumlah 433 data tahun angkatan 2021. Data akan dilakukan *preprocessing* untuk pengecekan *missing value*. Data terbagi menjadi data latih dan data uji masing-masing 70% untuk data latih dan 30% untuk data uji. Adapun dibawah ini contoh dari data mahasiswa yang digunakan dapat dilihat pada Gambar 2.

JENIS KELAMIN	ASAL SEKOLAH	STATUS SEKOLAH	JALUR SELEKSI	ASAL DAERAH	IPS 1	IPS 2	IPS 3	IPS 4	SKS 1	SKS 2	SKS 3	SKS 4	TOTAL SKS	IPK	STATUS
PEREMPUAN	SMA	NEGERI	NASIONAL	Pulau Jawa	3.87	3.69	3.71	3.83	20	21	24	23	88	3.77	TERCAPAI
PEREMPUAN	SMA	NEGERI	NASIONAL	Pulau Jawa	3.83	3.73	3.89	3.86	20	21	24	23	88	3.83	TERCAPAI
LAKI-LAKI	SMA	NEGERI	MANDIRI	Pulau Jawa	3.72	3.68	3.81	3.8	20	21	22	23	86	3.75	TIDAK TERCAPAI
PEREMPUAN	SMA	NEGERI	NASIONAL	Pulau Jawa	3.91	3.86	3.81	3.92	20	21	24	23	88	3.87	TERCAPAI
PEREMPUAN	SMA	NEGERI	MANDIRI	Pulau Jawa	3.94	3.74	3.8	3.95	20	21	24	23	88	3.86	TERCAPAI
PEREMPUAN	SMA	NEGERI	NASIONAL	Pulau Jawa	4	3.65	3.95	3.97	20	21	24	23	88	3.9	TERCAPAI
LAKI-LAKI	SMA	NEGERI	MANDIRI	Pulau Jawa	3.94	3.93	3.64	3.92	20	21	20	23	84	3.86	TIDAK TERCAPAI
PEREMPUAN	SMA	NEGERI	MANDIRI	Pulau Jawa	3.84	3.89	3.85	3.8	20	21	24	23	88	3.84	TERCAPAI
PEREMPUAN	SMA	NEGERI	MANDIRI	Pulau Jawa	3.7	3.96	3.66	3.81	20	21	22	23	86	3.78	TIDAK TERCAPAI
LAKI-LAKI	SMA	NEGERI	MANDIRI	Pulau Jawa	2.51	3.63	3.64	3.85	20	21	24	21	86	3.43	TIDAK TERCAPAI
PEREMPUAN	SMA	NEGERI	NASIONAL	Pulau Jawa	3.91	3.97	3.86	3.95	20	21	24	23	88	3.92	TERCAPAI
LAKI-LAKI	SMA	NEGERI	MANDIRI	Pulau Jawa	3.27	3.56	3.57	3.5	20	21	20	21	82	3.48	TIDAK TERCAPAI
LAKI-LAKI	SMA	SWASTA	MANDIRI	Pulau Jawa	3.56	3.55	3.48	3.8	20	21	22	23	86	3.6	TIDAK TERCAPAI
LAKI-LAKI	SMA	NEGERI	MANDIRI	Pulau Jawa	3.64	3.64	3.44	3.66	20	21	20	21	82	3.6	TIDAK TERCAPAI
LAKI-LAKI	SMA	NEGERI	MANDIRI	Pulau Jawa	3.77	3.66	3.7	3.68	20	21	22	23	86	3.7	TIDAK TERCAPAI
LAKI-LAKI	SMA	SWASTA	NASIONAL	Pulau Jawa	3.78	3.81	3.72	3.86	20	21	24	23	88	3.79	TERCAPAI
PEREMPUAN	SMA	NEGERI	NASIONAL	Pulau Jawa	3.44	3.71	3.7	3.9	20	21	24	23	88	3.69	TERCAPAI
PEREMPUAN	SMA	NEGERI	NASIONAL	Pulau Jawa	3.81	3.55	2.3	1.68	20	21	20	17	78	2.89	TIDAK TERCAPAI
PEREMPUAN	SMA	NEGERI	NASIONAL	Pulau Jawa	3.83	3.84	3.85	3.83	20	21	22	23	86	3.84	TIDAK TERCAPAI
PEREMPUAN	SMA	NEGERI	MANDIRI	Pulau Jawa	3.74	3.66	3.86	3.85	20	21	24	23	88	3.78	TERCAPAI
LAKI-LAKI	SMA	NEGERI	NASIONAL	Pulau Jawa	3.5	3.79	3.54	3.85	20	21	20	21	82	3.67	TIDAK TERCAPAI
LAKI-LAKI	SMA	NEGERI	MANDIRI	Pulau Jawa	3.7	3.66	3.74	3.83	20	21	24	21	86	3.73	TIDAK TERCAPAI
LAKI-LAKI	SMA	NEGERI	MANDIRI	Pulau Jawa	3.4	3.52	3.8	3.69	20	21	24	23	88	3.53	TIDAK TERCAPAI

Gambar. 2. Data Mahasiswa

Untuk memprediksi capaian studi mahasiswa, data yang digunakan untuk melakukan perbandingan tingkat akurasi digunakan algoritma klasifikasi Random Forest dan Support Vector Machine. Penelitian dimulai dengan pengumpulan informasi. Data berjumlah 433 dikumpulkan di Bagian Akademik Program Studi S1 Ilmu Hukum Fakultas Hukum Universitas Sebelas Maret angkatan 2021. Jenis Kelamin, Asal Sekolah, Status Sekolah, Jalur Seleksi, Asal Daerah, Indeks Prestasi Semester 1, Indeks Prestasi Semester 2, Indeks Prestasi Semester 3, Indeks Prestasi Semester 4, Satuan Kredit Semester 1, Satuan Kredit Semester 2, Satuan Kredit Semester 4, Total SKS, Indeks Prestasi Kumulatif, dan Status termasuk dalam dataset pengujian. Tabel 1 menunjukkan atribut dan tipe data yang digunakan untuk penelitian.

TABEL I VARIABEL DAN TIPE DATA

No	Variabel	Tipe Data	Keterangan
1	Jenis Kelamin	Polynomial	L = Laki-laki P = Perempuan
2	Asal Sekolah	Integer	SMA SMK
3	Status Sekolah	Integer	Negeri Swasta
4	Jalur Sekolah	Integer	Nasional Mandiri
5	Asal Daerah	Integer	Pulau Jawa Luar Pulau Jawa
6	IPS 1	Real	Indeks Prestasi Semester
7	IPS 2	Real	
8	IPS 3	Real	
9	IPS 4	Real	
10	SKS 1	Integer	Satuan Kredit Semester
11	SKS 2	Integer	
12	SKS 3	Integer	
13	SKS 4	Integer	
14	Total SKS	Integer	

15	IPK	Real	
16	Status	Polynomial	Tercapai Tidak Tercapai

### 2.10 Proses Seleksi Fitur

Pada tahapan ini dilakukan eksperimen dengan seleksi fitur menggunakan *Forward Selection* dengan tujuan untuk mengetahui atribut-atribut yang memiliki pengaruh terhadap klasifikasi capaian studi mahasiswa dan dalam peningkatan performa akurasi.

### 2.11 Pemodelan dan Evaluasi

Pada tahap ini dilakukan pemodelan perbandingan metode *Random Forest* dan *Support Vector Machine* untuk memprediksi capaian studi mahasiswa. Data dibagi menjadi data latihan dan data uji. Kemudian, dataset diolah ke dalam pembelajaran mesin dan dibuat model klasifikasi dengan membandingkan dua algoritma klasifikasi untuk mendapatkan hasil yang paling akurat serta atribut yang berpengaruh dalam klasifikasi memprediksi capaian studi mahasiswa. Alat yang digunakan untuk membuat pemodelan adalah aplikasi RapidMiner. Setelah dilakukan pemodelan dari kedua algoritma, tahapan terakhir yaitu evaluasi dengan menggunakan *Confusion Matrix*. Adapun *Confusion Matrix* dapat dilihat pada Tabel 2 [6].

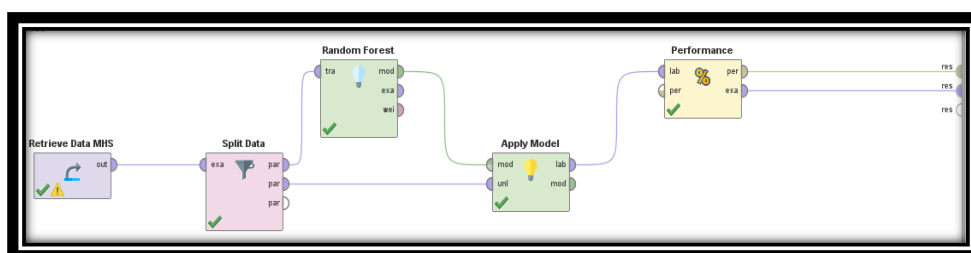
TABEL 2 CONFUSION MATRIX

Prediction	Actual	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

## III. HASIL DAN PEMBAHASAN

### 3.1 Pemodelan Klasifikasi *Random Forest*

Pemodelan dilakukan dengan menggunakan aplikasi RapidMiner. Prosesnya dimulai dengan mengambil data latihan yang telah disediakan sebelumnya dengan mengklik tab *Import Data* dan memilih folder lokal tempat data disimpan. Ada pun tampilan pemodelan *Random Forest* dapat dilihat pada Gambar 3.



Gambar. 3. Pemodelan Random Forest

Dataset penelitian dibaca dalam format Microsoft Excel dengan operator *Read Excel*. Kemudian, dataset maha-siswa dibagi menjadi partisi data latihan dan data uji, masing-masing sebesar 70% dan 30%. Operator *Split Data* digunakan untuk membagi data latihan dan data uji, dan operator *Performance* digunakan untuk mengukur akurasi dari dataset.

### 3.2 Evaluasi Hasil *Random Forest*

Berdasarkan pengujian menggunakan aplikasi RapidMiner dengan algoritma klasifikasi *Random Forest*, adapun hasil akurasi dapat dilihat pada Tabel 3.

Tabel 3. HASIL RANDOM FOREST

Prediksi	Aktual	
	Tercapai	Tidak Tercapai
Tercapai	89	0
Tidak Tercapai	3	37

accuracy: 97.67%			
	true TERCAPAI	true TIDAK TERCAPAI	class precision
pred. TERCAPAI	89	0	100.00%
pred. TIDAK TERCAPAI	3	37	92.50%
class recall	96.74%	100.00%	

Gambar. 4. Hasil Akurasi Random Forest

weighted_mean_recall: 98.37%, weights: 1, 1			
	true TERCAPAI	true TIDAK TERCAPAI	class precision
pred. TERCAPAI	89	0	100.00%
pred. TIDAK TERCAPAI	3	37	92.50%
class recall	96.74%	100.00%	

Gambar. 5. Recall Random Forest

weighted_mean_precision: 96.25%, weights: 1, 1			
	true TERCAPAI	true TIDAK TERCAPAI	class precision
pred. TERCAPAI	89	0	100.00%
pred. TIDAK TERCAPAI	3	37	92.50%
class recall	96.74%	100.00%	

Gambar. 6. Precision Random Forest

Pada Tabel 3 diketahui nilai dari prediksi tercapai 89 yang merupakan nilai *true positive* dan nilai *true negative* mendapatkan nilai prediksi 37. Sedangkan nilai *false negative* pada prediksi tidak tercapai mendapatkan nilai 3 dan *false positive* mendapatkan nilai 0. Hasil perhitungan kinerja algoritma *Random Forest* adalah sebagai berikut :

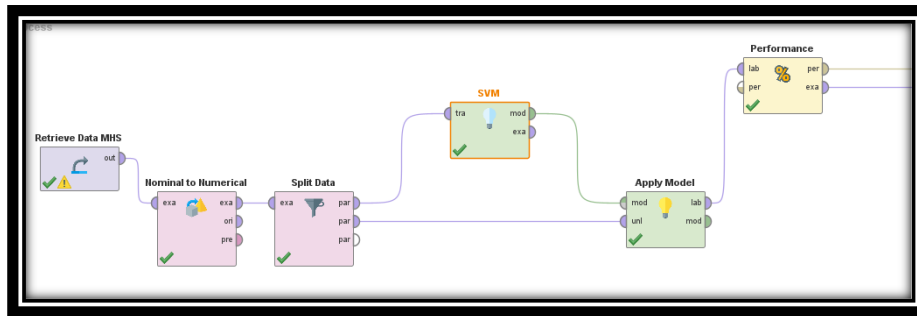
$$Accuracy : \frac{89+37}{89+37+3+0} = \frac{126}{129} = 0,9767 \times 100\% = \mathbf{97,67\%}$$

$$Precision : \frac{TP}{TP+FN} = \frac{89}{89+3} = \frac{89}{92} = 0,9673 \times 100\% = \mathbf{96,73\%}$$

$$Recall : \frac{TP}{TP+FP} = \frac{89}{89+0} = 1 \times 100\% = \mathbf{100\%}$$

### 3.3 Pemodelan Klasifikasi *Support Vector Machine*

Berdasarkan pengujian menggunakan RapidMiner dengan algoritma *Support Vector Machine*. Gambar 5 menunjukkan hasil kinerja akurasi



Gambar. 5. Pemodelan Support Vector Machine

Dataset penelitian dibaca dalam format Microsoft Excel dengan operator *Read Excel*. Kemudian, dataset siswa dibagi menjadi partisi data latihan dan data uji, masing-masing sebesar 70% dan 30%. Operator *Split Data* digunakan untuk membagi data latihan dan data uji, dan operator *Performance* digunakan untuk menghubungkan metode *Support Vector Machine*.

### 3.4 Evaluasi Hasil *Support Vector Machine*

Berdasarkan pengujian menggunakan aplikasi RapidMiner dengan algoritma klasifikasi *Support Vector Machine*, adapun hasil akurasi dapat dilihat pada Tabel 4.

Tabel 4. HASIL SUPPORT VECTOR MACHINE

	Prediksi		Aktual	
	Tercapai	Tidak Tercapai	Tercapai	Tidak Tercapai
Tercapai	81	0		
Tidak Tercapai	11	37		

weighted\_mean\_recall: 94.02%, weights: 1, 1

	true TERCAPAI	true TIDAK TERCAPAI	class precision
pred. TERCAPAI	81	0	100.00%
pred. TIDAK TERCAPAI	11	37	77.08%
class recall	88.04%	100.00%	

Gambar. 6. Hasil Recall Support Vector Machine

weighted\_mean\_precision: 88.54%, weights: 1, 1

	true TERCAPAI	true TIDAK TERCAPAI	class precision
pred. TERCAPAI	81	0	100.00%
pred. TIDAK TERCAPAI	11	37	77.08%
class recall	88.04%	100.00%	

Gambar. 7. Hasil Precision Support Vector Machine

Pada Tabel 4 diketahui nilai dari prediksi tercapai 81 yang merupakan nilai *true positive* dan nilai *true negative* mendapatkan nilai prediksi 37. Sedangkan nilai *false negative* pada prediksi tidak tercapai mendapatkan nilai 11 dan *false positive* mendapatkan nilai 0. Hasil perhitungan kinerja algoritma *Support Vector Machine* adalah sebagai berikut:



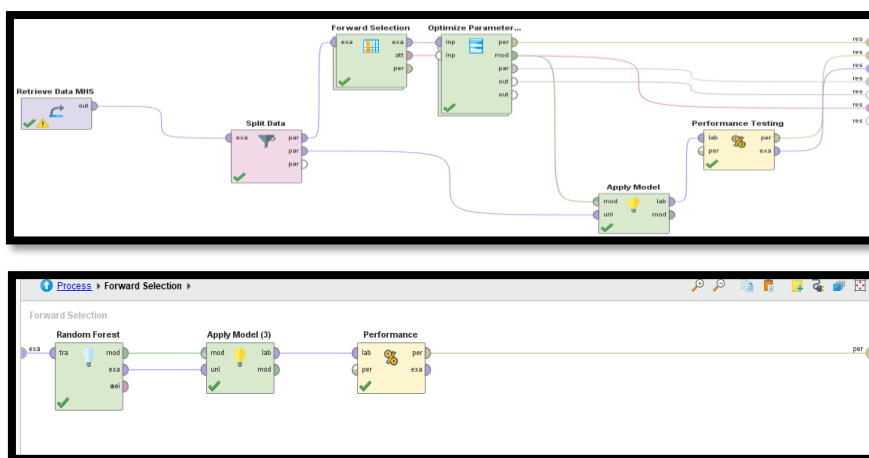
$$Accuracy : \frac{81+37}{81+37+0+11} = \frac{118}{129} = 0,9147 \times 100\% = \mathbf{91,47\%}$$

$$Precision : \frac{TP}{TP+FN} = \frac{81}{81+11} = \frac{81}{100} = 0,81 \times 100\% = \mathbf{81,00\%}$$

$$Recall : \frac{TP}{TP+FP} = \frac{81}{81+0} = \frac{81}{81} = 1 \times 100\% = \mathbf{100\%}$$

### 3.5 Pemodelan Seleksi Fitur Dengan *Forward Selection*

Pada tahap ini, seleksi fitur dilakukan dengan menggunakan seleksi maju, mulai dari atribut yang berdampak terbesar terhadap hasil prediksi capaian studi mahasiswa, dan atribut yang berdampak terbesar terhadap prediksi capaian studi mahasiswa. *Random Forest Classification* dan *Support Vector Machine* adalah dua algoritma yang menggunakan pemodelan seleksi fitur. Gambar 8 menunjukkan pemodelan seleksi fitur *Random Forest*.



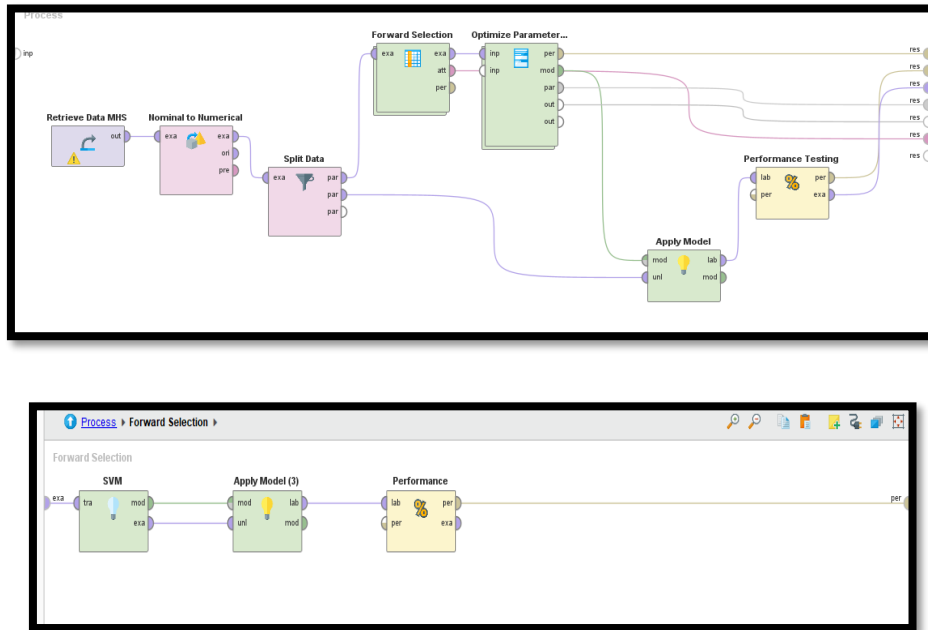
Gambar. 8. Pemodelan Seleksi Fitur Random Forest

Seleksi fitur dengan menggunakan *Forward Selection* pada algoritma klasifikasi *Random Forest* tidak merubah hasil akurasi yang sebelumnya mendapatkan akurasi sebesar 97,67%. Dari hasil pemodelan seleksi fitur, maka didapatkan atribut yang berpengaruh dalam memprediksi capaian studi mahasiswa. Tabel 5 menunjukkan atribut yang berpengaruh.

Tabel 5. SELEKSI ATRIBUT RANDOM FOREST

Algoritma Klasifikasi	Atribut	Akurasi Sebelum Seleksi Fitur	Akurasi Setelah Seleksi Fitur
<i>Random Forest</i>	Jenis Kelamin	97,67%	97,67%
	IPS 1		
	IPS 4		
	Total SKS		

Sedangkan pemodelan seleksi fitur *Support Vector Machine* dapat dilihat pada Gambar 11



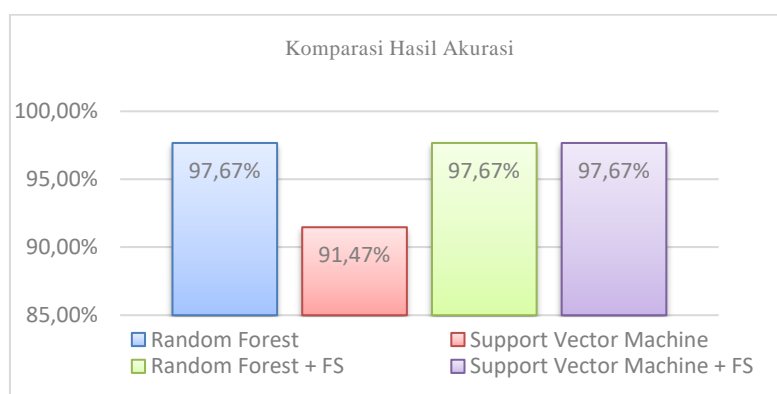
Gambar. 9. Pemodelan Seleksi Fitur Support Vector Machine

Dengan menggunakan *Forward Selection* pada algoritma klasifikasi *Support Vector Machine*, hasil akurasi yang sebelumnya sebesar 91,47% dapat ditingkatkan menjadi 97,67%. Faktor yang berpengaruh dalam memprediksi prestasi akademik siswa diidentifikasi melalui hasil pemodelan seleksi fitur. Tabel 6 menunjukkan atribut yang berpengaruh.

Tabel 6. SELEKSI ATRIBUT SUPPORT VECTOR MACHINE

Algoritma Klasifikasi	Atribut	Akurasi Sebelum Seleksi Fitur	Akurasi Setelah Seleksi Fitur	Kenaikan Akurasi
<i>Support Vector Machine</i>	Total SKS	91,47%	97,67%	6,2%

Dari hasil pengujian pemodelan perbandingan metode klasifikasi *Random Forest* dan *Support Vector Machine* sebelum dan sesudah menggunakan seleksi fitur dengan *Forward Selection*, maka didapatkan hasil yang dapat dilihat pada Gambar 10.



Gambar. 10. Komparasi Hasil Akurasi

Hasil komparasi algoritma klasifikasi dengan seleksi fitur yang digunakan pada data capaian studi mahasiswa diperoleh bahwa algoritma klasifikasi *Random Forest* tidak merubah hasil akurasi dengan nilai sebesar 97,67% dan menghasilkan empat atribut yang berpengaruh yaitu jenis kelamin, IPS 1, IPS 4, dan Total SKS, sedangkan penerapan seleksi fitur pada algoritma klasifikasi *Support Vector Machine* dapat meningkatkan nilai akurasi yang semula 91,47% setelah menggunakan seleksi fitur menjadi 96,67% dengan menghasilkan satu atribut yang berpengaruh dalam memprediksi capaian studi mahasiswa yaitu Total SKS. Perbandingan ini

dilakukan dengan membandingkan hasil penelitian sebelumnya. Hasil penelitian model memperoleh nilai akurasi yang baik dan memiliki kontribusi dari penelitian sebelumnya.

Seperti hasil dari penelitian yang dilakukan oleh [5] Dengan menganalisis 1.989 data nilai mahasiswa di Fakultas Ilmu Komputer BINUS University, penelitian ini mengusulkan algoritma K-NN untuk mengklasifikasikan berdasarkan perkiraan jarak nilai siswa. Hasil analisis metode K-NN menunjukkan nilai akurasi sebesar 93,2% dalam memprediksi status tepat waktu belajar, 91,5% dalam memprediksi total tahun studi, dan 75,62% dalam memprediksi total semester studi.

Penelitian lainnya [21] menggunakan Data yang digunakan bersumber dari *database* Kementerian Pendidikan di Palestina. Data tersebut memiliki 13 nilai atribut. 11 atribut berisi nilai mata pelajaran, rata-rata dan nilai. Metode yang digunakan yaitu klasifikasi K-NN dengan hasil K-NN *Weighted* 94,3% dan K-NN *Cosine* 81,3%. Penelitian oleh [22] Data Universitas Islam Al-Hidayah memuat 5 atribut: Jenis kelamin, status, status pekerjaan, IPK, prediksi, dengan 97 data menggunakan metode klasifikasi SVM dan C4.5 dengan hasil SVM 85% dan C.45 80%.

Penelitian lainnya yang dilakukan oleh [1] dengan 342 data Jurusan Teknik Informatika dan Sistem Informasi tahun akademik, 2010/2011, 2011/2012, 2012/2013, dan 2013/2014, menghasilkan akurasi C.45 98,04%, *Naïve Bayes* 96,00% dan K-NN 90%.

Penelitian [34] Data Universitas Bumigora yang terdiri dari 302 data mahasiswa angkatan 2009-2012 dianalisis menggunakan metode SVM, KNN Imputasi, dan *K-Means-Smote*. Menghasilkan akurasi sebesar 83,9%.

Penelitian lainnya [35] dengan Data IT Center Universitas Pamulang tahun 2017/2018 mencakup enam model k-nearest neighbour (kNN), antara lain DT, NB, kNN, LR, NN, MLP, dan SVM menghasilkan model klasifikasi DT, NB, k-NN, LR, NN, MLP, dan SVM, menunjukkan kinerja serupa di berbagai kumpulan data dan tingkat rata-rata yang bervariasi. Penelitian lainnya yang serupa oleh [36] dengan Data mahasiswa yang digunakan adalah usia, jenis kelamin, IPK, IPS 1, IPS 2, IPS 3, IPS 4, jurusan, SKS dan lama masa studi sebanyak 2181 data dengan algoritma *Support Vector Machine*, menghasilkan akurasi sebesar 94,4%.

Penelitian yang selanjutnya oleh [37] Penelitian menggunakan data akademik dari 1530 dataset dengan 7 atribut menggunakan metode SVM, *Decision Tree*, dan KNN dengan hasil SVM : 95%, *Decision Tree* : 93% dan K-NN : 92%. Penelitian [7] Penelitian ini memanfaatkan data 2293 mahasiswa program Ilmu Komputer, dengan menggunakan metode *Naïve Bayes*, *Random Forest*, *Decision Tree*, K-NN, dan SVM menghasilkan eksperimen menggunakan *Naïve Bayes*, *Random Forest*, *Decision Tree*, K-NN, dan SVM menunjukkan akurasi yang lebih baik dengan *Random Forest*, mencapai 77.35% dengan klasifikasi cukup. Penelitian yang terakhir yang dilakukan oleh [2] penelitian ini menggunakan Dataset yang digunakan adalah data mahasiswa angkatan tahun akademik 2018/2019 yang diambil dari Bagian Akademik Institu Bisnis dan Multimedia asmi dengan menggunakan metode *Random Forest* dan C4.5, menghasilkan akurasi RF : 92.4% dan C4.5 87.1%.

#### IV. KESIMPULAN

Metode seleksi maju tidak dapat meningkatkan akurasi algoritma *Random Forest* dan menghasilkan hasil sebesar 97,67%. Pemodelan dengan *Random Forest* sebelum seleksi fitur dan *Support Vector Machine*, yang menggunakan 433 data mahasiswa angkatan tahun 2021, yang dibagi menjadi 30% data uji dan 70% data latih, menghasilkan hasil akurasi sebesar 91,47%. Dalam pemodelan *Support Vector Machine*, seleksi fitur dapat meningkatkan nilai akurasi sebesar 6,2%. Akibatnya, akurasi SVM setelah dilakukan seleksi fitur adalah sebesar 97,67% dengan menghasilkan satu atribut yang berpengaruh, yaitu total SKS, pada klasifikasi prediksi capaian studi mahasiswa. Perbandingan metode klasifikasi RF dan SVM setelah dilakukan seleksi fitur menghasilkan akurasi yang sama, yaitu 97,67%. Oleh karena itu, temuan penelitian ini termasuk dalam kategori Hasil penelitian diharapkan dapat membantu pengelola program studi memberikan perawatan atau bimbingan khusus kepada siswa yang diharapkan tidak mencapai tujuan akademik mereka.

#### DAFTAR PUSTAKA

- [1] A. Budiyantara, I. Irwansyah, E. Prengki, P. Pratama Ahmad, and N. Wiliani, "Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu," 2020, [Online]. Available: [www.bri-institute.ac.id](http://www.bri-institute.ac.id)
- [2] S. Linawati, S. Nurdiani, H. Kartika, and L. Latifah, "Prediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma Random Forest Dan C4.5," vol. VIII, no. 1, 2020, [Online]. Available: [www.bsi.ac.id](http://www.bsi.ac.id)
- [3] I. M. N. O. Mahardika, "PINTU : Pusat Penjamin Mutu," vol. 2, no. 2, 2021.
- [4] A. M. F. Achmad, "Prediksi Performansi Mahasiswa dengan Mempertimbangkan Motivasi Intrinsik Menggunakan Machine Learning," *J. Serambi Eng.*, vol. 8, no. 4, pp. 6976–6982, 2023, doi: 10.32672/jse.v8i4.6722.
- [5] T. Asril and S. M. Isa, "Prediction of Students Study Period using K-Nearest Neighbor Algorithm," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 6, pp. 2585–2593, 2020, doi: 10.30534/ijeter/2020/60862020.

- [6] 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.
- [7] J. Zeniarja, A. Salam, and F. A. Ma’ruf, “Seleksi Fitur dan Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa,” *J. Rekayasa Elektr.*, vol. 18, no. 2, pp. 102–108, 2022, doi: 10.17529/jre.v18i2.24047.
- [8] N. Hidayati and A. Hermawan, “K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation,” *J. Eng. Appl. Technol.*, vol. 2, no. 2, pp. 86–91, 2021, doi: 10.21831/jeatech.v2i2.42777.
- [9] R. P. S. Putri and I. Waspada, “Penerapan Algoritma C4.5 pada Aplikasi Prediksi Kelulusan Mahasiswa Prodi Informatika,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 1–7, 2018, doi: 10.23917/khif.v4i1.5975.
- [10] S. F. Selegi, “Evaluasi Pelaksanaan Standar Proses Pembelajaran KKN Program Studi Pendidikan Geografi Universitas PGRI Palembang,” vol. 4, no. 2, 2019.
- [11] T. S. Guntoro, “MULTILATERAL : Jurnal Pendidikan Jasmani dan Olahraga mahasiswa olahraga Tri Setyo Guntoro PENDAHULUAN Dalam Permendikbud No. 3 tahun 2020 disebutkan bahwa rumusan keterampilan umum untuk jenjang sarjana adalah mahasiswa yaitu dapat menerapkan pemikiran,” vol. 22, no. 1, pp. 1–12, 2023.
- [12] I. K. Hasan, R. Resmawan, and J. Ibrahim, “Perbandingan K-Nearest Neighbor dan Random Forest dengan Seleksi Fitur Information Gain untuk Klasifikasi Lama Studi Mahasiswa,” *Indones. J. Appl. Stat.*, vol. 5, no. 1, p. 58, 2022, doi: 10.13057/ijas.v5i1.58056.
- [13] A. F. A. Naibaho and A. Zahra, “Prediksi Kelulusan Siswa Sekolah Menengah Pertama Menggunakan Machine Learning,” *J. Inform. dan Tek. Elektro Terap.*, vol. 11, no. 3, 2023, doi: 10.23960/jitet.v11i3.3056.
- [14] N. W. E. Rosiana Dewi, I. G. Aris Gunadi, and G. Indrawan, “Detection of Class Regularity with Support Vector Machine methods,” *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 11, no. 1, p. 20, 2020, doi: 10.24843/lkjiti.2020.v11i1.i01.p03.
- [15] T. Tumini, “Support vector machine,” *Mach. Learn. Methods Appl. to Brain Disord.*, vol. 4, no. 2, pp. 101–121, 2019, doi: 10.1016/B978-0-12-815739-8.00006-7.
- [16] H. Hermanto, A. Mustopa, and A. Y. Kuntoro, “Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 211–220, 2020, doi: 10.33480/jitk.v5i2.1181.
- [17] M. F. Nugroho and S. Wibowo, “Fitur Seleksi Forward Selection Untuk Menentukan Atribut Yang Berpengaruh Pada Klasifikasi Kelulusan Mahasiswa Fakultas Ilmu Komputer UNAKI Semarang Menggunakan Algoritma Naive Bayes,” *J. Inform. Upgris*, vol. 3, no. 1, pp. 63–70, 2017, doi: 10.26877/jiu.v3i1.1669.
- [18] T. H. Hasibuan and D. Mahdiana, “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta,” *Skanika*, vol. 6, no. 1, pp. 61–74, 2023, doi: 10.36080/skanika.v6i1.2976.
- [19] E. Nurlia and U. Enri, “Penerapan Fitur Seleksi Forward Selection Untuk Menentukan Kematian Akibat Gagal Jantung Menggunakan Algoritma C4.5,” *J. Tek. Inform. Musirawas Elin Nurlia*, vol. 6, no. 1, p. 42, 2021.
- [20] T. A. Y. Siswa and R. P. Wibowo, “Komparasi Metode Seleksi Fitur Dalam Prediksi Keterlambatan Pembayaran Biaya Kuliah,” *Teknika*, vol. 12, no. 1, pp. 73–82, 2023, doi: 10.34148/teknika.v12i1.601.
- [21] S. S. Alfere and A. Y. Maghari, “Prediction of Student’s Performance Using Modified KNN Classifiers,” *First Int. Conf. Eng. Futur. Technol. (ICEFT 2018)*, no. Icept, pp. 143–150, 2018.
- [22] A. Mailana, A. A. Putra, S. Hidayat, and A. Wibowo, “Comparison of C4.5 Algorithm and Support Vector Machine in Predicting the Student Graduation Timeliness,” *J. Online Inform.*, vol. 6, no. 1, p. 11, 2021, doi: 10.15575/join.v6i1.608.
- [23] O. W. Yuda, D. Tuti, I. S. Yee, and Susanti, “Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest,” *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, 2022, doi: 10.33372/stn.v8i2.885.
- [24] A. Andika, S. Syarli, and C. R. Sari, “Data Mining Klasifikasi Kelulusan Mahasiswa Menggunakan Metode Naive Bayes,” *J. Pegguruang Conf. Ser.*, vol. 4, no. 1, p. 423, 2022, doi: 10.35329/jp.v4i1.2358.
- [25] N. Khasanah, A. Salim, N. Afni, R. Komarudin, and Y. I. Maulana, “Prediksi Kelulusan Mahasiswa Dengan Metode Naive Bayes,” *Technol. J. Ilm.*, vol. 13, no. 3, p. 207, 2022, doi: 10.31602/tji.v13i3.7312.
- [26] M. Raharjo, R. Ridwa, J. L. Putra, and T. Ifian A. Sandi, “Implementasi Metode Decision Tree Klasifikasi Data Mining Untuk Prediksi Peminatan Jurusan Robotika oleh Mahasiswa,” *J. Tek. Komput. AMIK BSI*, 2019, doi: 10.31294/jtk.v4i2.
- [27] E. E. Barito, J. T. Beng, and D. Arisandi, “Penerapan Algoritma C4.5 Untuk Klasifikasi Mahasiswa Penerima Bantuan Sosial Covid-19,” *J. Ilmu Komput. dan Sist. Inf.*, vol. 10, no. 1, 2022, doi: 10.24912/jiksi.v10i1.17819.
- [28] M. L. Mu’tashim, A. Zaidiah, and B. S. Yulistiawan, “Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta),” *Semin. Nas. Mhs. Bid. Ilmu Komput. dan Apl.*, vol. 4, no. 1, pp. 155–166, 2023.
- [29] N. A. P. Indaryono, R. R. Saedudin, and F. Hamami, “Analisa Perbandingan Algoritma Random Forest Dan Naive Bayes Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Di Indonesia,” vol. 9, no. 1, pp. 158–167, 2024, doi: 10.29100/jipi.v9i1.4421.
- [30] M. F. Abdullah, K. Kusriani, and M. R. Arief, “Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Menggunakan Metode Support Vector Machine,” *J. Sains dan Teknol.*, vol. 14, no. 1, p. 35, 2022.
- [31] F. N. Fajriyan, M. Ahsan, and W. Harianto, “Komparasi Tingkat Akurasi Information Gain Dan Gain Ratio Pada Metode K-Nearest Neighbor,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 6, no. 1, pp. 386–391, 2022, doi: 10.36040/jati.v6i1.4694.
- [32] M. N. Akbar, H. Hariani, and A. I. Syahyadi, “Analisis Prediksi Ketepatan Masa Studi Mahasiswa Menggunakan Algoritma Naive Bayes Classifier Dan Feature Selection,” *J. INSTEK (Informatika Sains dan Teknol.)*, vol. 7, no. 2, pp. 306–315, 2022, doi: 10.24252/instek.v7i2.32576.
- [33] D. B. Srisulistiwati, M. Khaerudin, and S. Rejeki, “Sistem Informasi Prediksi Penjualan Alat Tulis Kantor Dengan Metode Fp-Growth (Studi Kasus Toko Koperasi Sekolah Bina Mulia),” *J. Sist. Inf. Univ. Suryadarma*, vol. 8, no. 2, 2021, doi: 10.35968/jsi.v8i2.739.
- [34] H. Hairani, “Peningkatan Konejra Metode SVM Menggunakan Metode KNN Imputasi dan K-Means-Smote untuk Klasifikasi Kelulusan Mahasiswa Universitas Bumigora,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 4, pp. 713–718, 2021, doi: 10.25126/jtiik.2021843428.
- [35] R. Rachmatika and A. Bisri, “Perbandingan Model Klasifikasi untuk Evaluasi Kinerja Akademik Mahasiswa,” *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 417, 2020, doi: 10.26418/jp.v6i3.43097.
- [36] E. Haryatmi and S. P. 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.

- [37] S. Wiyono, D. S. Wibowo, M. F. Hidayatullah, and Dairoh, “Comparative Study of KNN, SVM and Decision Tree Algorithm for Student’s Performance Prediction,” *Int. J. Comput. Sci. Appl. Math.*, vol. 6, no. 2, p. 50, 2020, doi: 10.12962/j24775401.v6i2.4360.