

## **Analisis Pemilihan Jurusan pada Calon Siswa SMK Negeri 4 Palembang Pada Faktor Penentu Pemilihan Jurusan Menggunakan Association Rule dan Random Forest**

**Muhamad Hafiz Budi Prayoga<sup>\*1</sup>, Ermatita<sup>2</sup>**

<sup>1,2</sup>Magister Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya, Indonesia  
Email: <sup>1</sup>[hafizbudiprayoga@gmail.com](mailto:hafizbudiprayoga@gmail.com), <sup>2</sup>[ermatita@unsri.ac.id](mailto:ermatita@unsri.ac.id)

### **Abstrak**

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi pemilihan jurusan pada calon siswa SMK Negeri 4 Palembang menggunakan metode Association Rule dan Random Forest. Pemilihan jurusan di SMK merupakan keputusan penting yang dapat memengaruhi masa depan siswa, sehingga diperlukan pendekatan analisis yang tepat untuk memberikan rekomendasi jurusan berdasarkan faktor akademis dan non-akademis. Association Rule digunakan untuk mengidentifikasi pola-pola tersembunyi antara faktor akademis, seperti nilai mata pelajaran, dengan pilihan jurusan, sementara Random Forest diterapkan untuk membangun model prediksi yang akurat dalam menentukan jurusan yang tepat bagi siswa. Hasil penelitian menunjukkan bahwa nilai mata pelajaran tertentu, seperti Matematika dan IPA, memiliki hubungan yang signifikan dengan kecenderungan siswa memilih jurusan tertentu. Model prediksi berbasis Random Forest memberikan akurasi yang tinggi dan stabil dalam memprediksi pilihan jurusan siswa berdasarkan faktor-faktor yang telah diidentifikasi. Model ini diharapkan dapat membantu pihak sekolah dalam memberikan bimbingan karier yang lebih terarah kepada siswa, sehingga mereka dapat membuat keputusan yang lebih tepat dan sesuai dengan minat serta potensi akademik mereka. Penelitian ini memberikan kontribusi dalam pengembangan sistem pendukung keputusan di sekolah, yang dapat digunakan dalam proses Penerimaan Peserta Didik Baru (PPDB) untuk meningkatkan kepuasan dan kesesuaian pilihan jurusan dengan kemampuan siswa. Dengan demikian, hasil dari penelitian ini dapat memberikan wawasan yang bermanfaat bagi sekolah dalam meningkatkan efektivitas bimbingan karier bagi siswa SMK.

**Kata kunci:** *Association Rule, Model Prediksi, Pemilihan Jurusan, Random Forest, SMK Negeri 4 Palembang*

### ***Analysis of Major Selection for Prospective Students of SMK Negeri 4 Palembang on Determining Factors for Major Selection Using Association Rule and Random Forest***

#### **Abstract**

*This study aims to analyze the factors that influence the choice of majors for prospective students of SMK Negeri 4 Palembang using the Association Rule and Random Forest methods. The choice of major in SMK is an important decision that can affect the future of students, so an appropriate analytical approach is needed to provide recommendations for majors based on academic and non-academic factors. Association Rule is used to identify hidden patterns between academic factors, such as subject scores, and major choices, while Random Forest is applied to build an accurate prediction model in determining the right direction for students. The results of the study show that certain subject scores, such as Mathematics and Science, have a significant relationship with students' tendency to choose certain majors. The Random Forest prediction-based model provides high and stable accuracy in predicting students' choice of direction based on the factors that have been identified. This model is expected to help schools in providing more targeted career guidance to students, so that they can make more appropriate decisions that are in accordance with their interests and academic potential. This study contributes to the development of decision support systems in schools, which can be used in the New Student Admissions (PPDB) process to improve satisfaction and suitability of subject choices with students' abilities. Thus, the results of this study can provide useful insights for schools in improving the effectiveness of career guidance for vocational high school students.*

**Keywords:** *Association Rule, Prediction Model, Major Selection, Random Forest, SMK Negeri 4 Palembang*

## 1. PENDAHULUAN

Sekolah Menengah Kejuruan (SMK) adalah institusi pendidikan formal yang menyediakan pendidikan kejuruan di jenjang menengah. Tujuan dari pendidikan ini adalah memberikan keterampilan dan pengetahuan praktis kepada siswa agar mereka siap memasuki dunia kerja di bidang tertentu setelah menyelesaikan pendidikan [1]. SMK merupakan lanjutan dari Sekolah Menengah Pertama (SMP) atau Madrasah Tsanawiyah (MTs), atau bentuk pendidikan lain yang setara. Pendidikan di SMK lebih difokuskan pada pengembangan kompetensi praktis yang sesuai dengan kebutuhan industri dan pasar kerja, sehingga menghasilkan lulusan yang memiliki keterampilan spesifik dan siap kerja di berbagai bidang seperti teknik, bisnis, pariwisata, kesehatan, dan teknologi informasi.

SMK Negeri 4 Palembang adalah salah satu SMK di Kota Palembang yang menawarkan delapan jurusan, termasuk Teknik Komputer Jaringan, Teknik Sepeda Motor, Teknik Kendaraan Ringan, dan lain-lain. Beragamnya pilihan jurusan di SMK Negeri 4 Palembang memberikan banyak opsi bagi calon siswa. Pemilihan jurusan merupakan keputusan penting yang dapat mempengaruhi masa depan siswa. Di SMK Negeri 4 Palembang, berbagai jurusan disediakan untuk memenuhi kebutuhan industri dan minat siswa. Namun, keputusan ini sering dipengaruhi oleh berbagai faktor, baik internal maupun eksternal, seperti minat pribadi, prestasi akademik, dukungan keluarga, informasi dari teman sebaya, serta prospek kerja di masa depan [2].

Pemilihan jurusan di SMK Negeri 4 Palembang sering kali menghadapi berbagai tantangan, baik dari sisi siswa maupun pihak sekolah. Banyak siswa merasa bingung atau kurang memiliki panduan yang memadai dalam memilih jurusan yang sesuai dengan minat dan potensi mereka. Hal ini sering kali menyebabkan siswa memilih jurusan yang kurang tepat, yang pada akhirnya berdampak pada kepuasan belajar dan hasil akademik. Selain itu, beberapa faktor eksternal, seperti saran dari keluarga atau teman, serta prospek kerja di masa depan, juga turut memengaruhi keputusan siswa dalam memilih jurusan. Namun, sejauh ini sekolah belum memiliki data yang cukup untuk menilai seberapa besar pengaruh faktor-faktor tersebut dalam proses pengambilan keputusan siswa.

Di sisi lain, SMK Negeri 4 Palembang belum memiliki alat analisis yang memadai untuk memahami pola pemilihan jurusan berdasarkan faktor-faktor seperti minat, nilai akademis, dan potensi siswa. Keterbatasan ini membuat pihak sekolah sulit memberikan rekomendasi yang lebih akurat dan berbasis data kepada siswa. Dampak jangka panjang dari permasalahan ini adalah tidak maksimalnya hasil pendidikan siswa, yang dapat memengaruhi kesiapan mereka dalam memasuki dunia kerja. Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk memetakan pola pemilihan jurusan siswa dengan pendekatan *Association Rule* dan *Random Forest* guna membantu sekolah menyediakan panduan yang lebih tepat dalam proses pemilihan jurusan.

Mengidentifikasi faktor-faktor yang paling berpengaruh dalam pemilihan jurusan dapat memberikan wawasan berharga bagi pihak sekolah untuk membantu siswa membuat keputusan yang tepat. Salah satu pendekatan untuk analisis ini adalah *Association Rule*, yang dapat mengidentifikasi hubungan atau asosiasi antara berbagai faktor penentu. Selain itu, metode *Random Forest* dapat digunakan untuk memvalidasi dan meningkatkan akurasi prediksi pemilihan jurusan.

*Association rules* adalah metode yang digunakan untuk mengelompokkan data dalam jumlah besar, seperti kumpulan item, dengan cara memasukkan data tersebut ke dalam serangkaian aturan tertentu. Aturan-aturan ini kemudian digunakan oleh sistem untuk mengoreksi jawaban dari peserta didik. Dalam prosesnya, data yang terkumpul diurai dan dianalisis untuk menemukan pola-pola yang signifikan [3]. Pola-pola ini kemudian dirumuskan menjadi aturan-aturan yang membantu sistem dalam menilai dan memberikan umpan balik yang akurat kepada peserta didik. Dengan demikian, *association rules* tidak hanya membantu dalam pengelompokan data tetapi juga dalam meningkatkan kualitas evaluasi dan pembelajaran.

*Random forest* adalah metode klasifikasi yang dikembangkan dari metode *Decision Tree*, di mana pemilihan atribut pada setiap node dilakukan secara acak untuk menentukan klasifikasi. Dalam proses klasifikasinya, keputusan akhir didasarkan pada mayoritas suara dari semua pohon keputusan yang telah dibentuk. Pembentukan *random forest* menggunakan teknik yang dikenal sebagai *bagging*, dengan pemilihan atribut secara acak. Untuk menumbuhkan pohon keputusan, digunakan metode *CART* (*Classification and Regression Tree*). Pohon-pohon ini akan tumbuh hingga ukuran maksimum dan tidak akan dipangkas, menghasilkan kumpulan pohon yang disebut sebagai hutan (*forest*) [4].

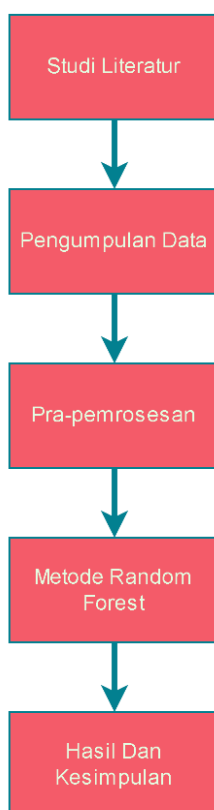
Berdasarkan permasalahan ini, penulis memiliki gagasan untuk melakukan penelitian tentang "Analisis Pemilihan Jurusan Pada Calon Siswa SMK Negeri 4 Palembang Berdasarkan *Association Rule* Pada Faktor-Faktor Penentu Pemilihan Jurusan Menggunakan *Random Forest*". Penelitian ini didasarkan pada kebutuhan untuk memahami dinamika dan faktor-faktor yang mempengaruhi pemilihan jurusan di sekolah tersebut. SMK Negeri 4 Palembang menawarkan berbagai jurusan dengan karakteristik dan prospek yang berbeda-beda. Memahami faktor-faktor yang mempengaruhi keputusan siswa adalah langkah krusial untuk membantu sekolah menyediakan informasi dan bimbingan yang efektif bagi calon siswa.

Metode Association Rule dipilih karena kemampuannya dalam mengidentifikasi hubungan mendalam antar faktor, yang dapat mengungkap pola-pola tersembunyi dalam data pemilihan jurusan. Teknik ini sangat berguna dalam situasi di mana berbagai faktor dapat saling berinteraksi dan mempengaruhi keputusan siswa secara kompleks. Association Rule memungkinkan kita melihat bagaimana faktor-faktor seperti prestasi akademik, minat, dan nilai saling terkait dalam mempengaruhi pemilihan jurusan[5].

Metode Random Forest dipilih karena keandalannya dalam memprediksi hasil berdasarkan data yang kompleks dan bervariasi. Algoritma ini mampu mengolah sejumlah besar data dan variabel, serta menghasilkan model prediksi yang akurat. Random Forest juga dapat menangani data yang tidak seimbang dan mengurangi risiko overfitting, sehingga hasil prediksi lebih dapat diandalkan. Kombinasi antara Association Rule dan Random Forest diharapkan memberikan pemahaman yang mendalam mengenai faktor-faktor yang mempengaruhi pemilihan jurusan, serta menghasilkan model prediksi yang akurat untuk mendukung pengambilan keputusan di SMK Negeri 4 Palembang[6].

## 2. METODE PENELITIAN

Metode penelitian merupakan kerangka atau tahap-tahapan yang akan dilakukan dalam sebuah penelitian, Adapun metode penelitian pada penelitian kali ini yang berjudul “Analisis Pemilihan Jurusan Pada Calon Siswa SMK Negeri 4 Palembang Berdasarkan Association Rule Pada Faktor-Faktor Penentu Pemilihan Jurusan Menggunakan Random Forest” memiliki beberapa tahap-tahapannya yang dapat dilihat pada Gambar 1.



Gambar 1 Diagram Alir Penelitian

### 2.1. Pengumpulan Data

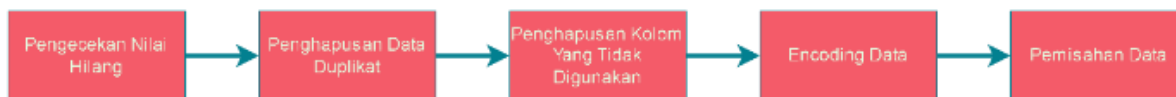
Data yang dikumpulkan langsung dari sumber objek penelitian. Data dalam penelitian ini berupa hasil wawancara dan observasi [7]. Dari kegiatan observasi dan wawancara dengan wakil kepala sekolah bidang kesiswaan bapak Wiwin, S.Pd diskusi yang berlangsung selama kurang lebih satu jam didapatkan informasi bahwa calon siswa yang mendaftar pada SMK negeri 4 Palembang, dan telah ditentukan sistem deteksi sebagai topik penelitian. Adapun data yang telah dikumpulkan melalui proses observasi dan wawancara yang dilakukan di SMK Negeri 4 Palembang terkumpul sebanyak 662 entri data dengan atribut NISN, NIS, Nama Siswa, Jenis Kelamin, Asal Sekolah, Nilai MTK, Nilai B.Ing B.Indo, dan IPA dari Semester 1 sampai Semester 5, Pilihan 1, Pilihan 2, Serta Jurusan yang diterima atau bisa dilihat pada Table 1.

Table 1 Data Siswa SMK Negeri 4 Palembang

No	NISN	NIS	Nama	JK	Sekolah A MTK	BING	BIND	IPA	Akademil Non-Akad	Pilihan 1	Pilihan 2	Jurusan Yang Lulus
1	97384085	241001	ADE	L	SMPN 19 78,4	83,8	82	82			DPIB	DPIB
2	3,1E+09	241002	AISYAH	P	SMPN 41 81,2	82,2	82,2	81,8			DPIB	TKJ
3	93571185	241003	AQILLA	P	SMPN 38 F 77,8	81	84,6	82	Juara 3 Tir		DPIB	TP
4	87684931	241004	ARROBI	L	SMP SANI 78,8	77,4	80	78,4			TP	DPIB
5	91720497	241005	ASMAWA	P	SMPN 46 F 81,8	88,2	82,6	85	Juara 2 Lo		TKR	RPL
6	3,1E+09	241006	AYUB	L	MTS AL AC 89,4	88,6	85,8	82,4	Peringkat 1		TP	TAV
7	1,05E+08	241007	AZZUKO	L	SMP YPI TI 86	80,8	81,4	80,4			DPIB	DPIB
8	94955951	241008	CAHAYA	P	SMPN 38 F 80,2	75,4	76,8	78,2			DPIB	RPL
9	94113962	241009	CHELSY	P	SMPN 49 F 82	79,8	80,4	78,8			DPIB	TKR
10	96359299	241010	DESHINTA	P	SMPN 46 F 87,8	85	75,6	87,8	Juara 3 Tir		TP	RPL
:	:	:	:	:	:	:	:	:	:	:	:	:
662	93212169	248033	ZHAHIRAHP		SMP UTAN 82,2	83,2	84,6	82,2	Peringkat	Peringkat	RPL	TKJ

2.2. Pra-Pemrosesan Data

Pra-pemrosesan data adalah tahap pengolahan di mana data yang belum siap untuk diproses dioptimalkan dengan menangani berbagai masalah seperti nilai yang hilang, data yang berlebihan, outlier, atau format data yang tidak konsisten. Masalah-masalah ini dapat mempengaruhi hasil pengolahan data, termasuk dalam proses klasifikasi. Oleh karena itu, langkah-langkah pra-pemrosesan sangat penting untuk mengatasi hambatan-hambatan tersebut dan memastikan bahwa data siap untuk diproses lebih lanjut [4]. Dalam penelitian ini, tujuan dari pra-pemrosesan adalah mengubah data mentah menjadi data yang terstruktur agar menghasilkan hasil yang lebih optimal. Langkah-langkah pra-pemrosesan yang dilakukan dalam penelitian ini meliputi pengecekan nilai yang hilang, penghapusan data duplikat, penghapusan kolom yang tidak relevan, encoding data, dan pemisahan data.atau bisa dilihat pada Gambar 2.



Gambar 2 Diagram Alir Pra-Pemrosesan Data

Langkah Awal yaitu Proses pengecekan nilai hilang (missing values) yang mana merupakan langkah penting dalam pemrosesan data yang mencakup beberapa tahapan kunci. Tahap pertama adalah mengidentifikasi lokasi di mana nilai-nilai hilang terdapat dalam dataset, yang bisa dilakukan menggunakan metode seperti isnull() atau isnan() pada struktur data seperti DataFrame di Python. Setelah identifikasi, langkah berikutnya adalah mengevaluasi jumlah dan pola distribusi nilai-nilai hilang dalam setiap fitur atau kolom. Strategi untuk menangani nilai-nilai hilang meliputi penghapusan baris atau kolom yang mengandung nilai hilang, pengisian nilai-nilai tersebut dengan metode tertentu seperti rata-rata atau median, atau menggunakan teknik imputasi yang lebih kompleks seperti regresi untuk memprediksi nilai-nilai yang hilang.

Setelah itu proses penghapusan data duplikat adalah proses penting dalam pemrosesan data untuk menjaga keakuratan dan konsistensi data yang digunakan dalam analisis atau pemodelan,

Langkah selanjutnya dalam pra-pemrosesan data adalah penghapusan kolom yang tidak digunakan. Proses ini bertujuan untuk mengurangi dimensi dataset dan meningkatkan efisiensi komputasi, serta menghilangkan informasi yang tidak relevan atau tidak berguna dalam analisis.

Encoding data adalah proses mengubah data dari format aslinya menjadi format yang dapat dipahami atau diproses lebih lanjut oleh model atau algoritma machine learning. Hal ini sering kali diperlukan karena banyak algoritma machine learning hanya menerima input dalam bentuk numerik. Salah satu jenis encoding yang umum digunakan adalah label encoding, di mana nilai-nilai kategorikal diubah menjadi nilai numerik.

Langkah terakhir dalam pra-pemrosesan data adalah pemisahan data menjadi set pelatihan dan set pengujian. Proses ini penting dalam pembangunan model pembelajaran mesin karena memungkinkan kita untuk mengevaluasi kinerja model secara objektif. Pemisahan data umumnya dilakukan dengan membagi dataset menjadi dua subset: set pelatihan (train set) dan set pengujian (test set). Set pelatihan digunakan untuk melatih model, sedangkan set pengujian digunakan untuk menguji performa model yang dilatih pada data yang belum pernah dilihat sebelumnya.

2.3. Assocotiaion Rule

Association rule mining adalah teknik dalam data mining yang digunakan untuk menemukan hubungan tersembunyi antara atribut dalam basis data transaksi. Teknik ini bertujuan untuk mengidentifikasi pola-pola yang sering muncul bersama dalam dataset besar, yang diekspresikan sebagai aturan berbentuk "IF-THEN".

Misalnya, jika suatu produk A sering dibeli bersamaan dengan produk B, aturan ini dapat diidentifikasi sebagai "Jika A, maka B"[8]. Adapun rumus umum dari *association rule* mencakup 3 komponen utama yaitu :

1. Support: yang merupakan untuk mengukur seberapa sering kombinasi item muncul dalam dataset.

$$support(A \rightarrow B) = \frac{jumlah\ transaksi\ yang\ mengandung\ A\ dan\ B}{jumlah\ total\ transaksi} \quad (1)$$

2. Confidence: Mengukur seberapa sering aturan "Jika A maka B" terbukti benar

$$confidence(A \rightarrow B) = \frac{jumlah\ transaksi\ yang\ mengandung\ A\ dan\ B}{jumlah\ transaksi\ mengandung\ A} \quad (2)$$

3. Lift: Mengukur hubungan antara A dan B dengan mempertimbangkan keterkaitan acak. Nilai lift di atas 1 menunjukkan bahwa A dan B lebih sering muncul bersama daripada yang diharapkan secara acak.

$$Lift(A \rightarrow B) = \frac{Confidence(A \rightarrow B)}{Support(B)} \quad (3)$$

Algoritma seperti Apriori dan FP-Growth sering digunakan untuk mengekstraksi aturan-aturan ini dengan cara yang efisien, terutama dalam dataset yang sangat besar dan kompleks. Algoritma-algoritma ini bekerja dengan mencari kombinasi item yang sering muncul, kemudian membangun aturan berdasarkan korelasi antara item tersebut. Untuk meningkatkan efisiensi, beberapa penelitian telah mengembangkan solusi komputasi terdistribusi yang lebih efektif dalam menambang aturan asosiasi dari dataset besar [9][10].

#### 2.4. Random Forest

Random Forest (RF) adalah model pembelajaran mesin berbasis ensemble yang dikenal karena akurasi prediktifnya yang tinggi dan varians yang rendah. Model ini terdiri dari sejumlah pohon keputusan yang dilatih pada sub-sampel berbeda dari dataset, dan hasil akhirnya adalah rata-rata dari prediksi masing-masing pohon, sehingga mengurangi risiko overfitting. Random Forest dapat diterapkan pada berbagai tugas, baik klasifikasi maupun regresi. Misalnya, dalam studi terkait peramalan beban jangka pendek, Random Forest menunjukkan kinerja yang unggul dibandingkan model lain dengan mengeksplorasi berbagai pola input dan mode pelatihan [11][12]. Berikut adalah rumus dan penjelasan cara kerja Random Forest:

1. Pembentukan Pohon Keputusan (Decision Tree)

Setiap pohon dalam Random Forest dibuat dari subset acak data latih. Metode ini menggunakan teknik bootstrap sampling, di mana setiap pohon dilatih pada sampel acak yang diambil dengan penggantian (sampling with replacement) dari data latih asli. Teknik ini membantu meningkatkan keragaman di antara pohon-pohon dalam model ensemble, yang pada gilirannya meningkatkan ketahanan terhadap overfitting [13].

2. Pemilihan Fitur yang Diacak (Random Feature Selection)

Untuk membagi setiap node di pohon, algoritma Random Forest hanya mempertimbangkan subset acak dari fitur yang tersedia. Biasanya, jumlah fitur yang dipilih adalah  $\sqrt{\text{jumlah total fitur}}$  untuk tugas klasifikasi, atau  $(\text{jumlah total fitur}) / 3$  untuk tugas regresi. Teknik ini mengurangi korelasi antar pohon keputusan dan membantu mengurangi risiko overfitting[11].

3. Pembentukan Model Ensemble

Random Forest menggabungkan prediksi dari setiap pohon keputusan untuk mendapatkan hasil akhir:

- o Untuk Klasifikasi: Prediksi akhir diperoleh dengan metode "voting mayoritas" dari semua pohon dalam hutan. Setiap pohon menghasilkan kelas yang diprediksi, dan kelas dengan jumlah suara terbanyak menjadi prediksi akhir:

$$y_i = \text{mode}(y_1, y_2, \dots, y_n) \quad (4)$$

Di sini  $y_i$  adalah prediksi dari pohon ke- $i$  dalam hutan, dan  $y$  adalah kelas akhir yang diprediksi oleh Random Forest[14].

- o Untuk Regresi: Prediksi akhir adalah rata-rata dari semua prediksi pohon dalam hutan:

$$y = \frac{1}{n} \sum_{i=1}^n Y_i \quad (5)$$

Di sini,  $y_i$  adalah nilai prediksi dari pohon ke- $i$ , dan  $y$  adalah nilai regresi akhir [15].

4. Pentingnya Fitur (Feature Importance)

Random Forest juga dapat menghitung pentingnya fitur untuk menentukan kontribusi masing-masing fitur terhadap keputusan akhir. Pentingnya fitur diukur dengan menghitung rata-rata pengurangan impurity (misalnya,

Gini impurity atau entropi) untuk semua node di mana fitur tertentu digunakan, atau dengan menilai dampak akurasi model setelah permutasi acak fitur [16].

#### 5. Mengurangi Varians dan Overfitting

Dengan menggabungkan banyak pohon yang tidak berkorelasi, Random Forest mengurangi varians model secara keseluruhan, yang membantu mengurangi overfitting dibandingkan dengan pohon keputusan tunggal. Pendekatan ini memberikan hasil yang lebih stabil dan dapat diandalkan, terutama untuk data yang bervariasi dan kompleks [13].

### 3. HASIL DAN PEMBAHASAN

Pada tahapan ini merupakan tahapan yang mana melakukan pengujian terhadap data yang telah diperoleh untuk dilakukan pengelompokan perjurusan karena setiap jurusan memiliki syarat nilai-nilai yang berbeda dengan hal tersebut maka dilakukan penerapan association rule lalu dilakukan pengujian klasifikasi dengan random forest.

#### 3.1. Penentuan Association Rule

Berdasarkan data calon siswa yang mendaftar dari mata pelajaran Matematika (MTK), Bahasa Inggris (BING), Bahasa Indonesia (BIND), dan Ilmu Pengetahuan Alam (IPA), syarat nilai berikut ditentukan untuk setiap jurusan di SMK Negeri 4 Palembang. Syarat nilai ini disesuaikan untuk mencerminkan relevansi setiap mata pelajaran terhadap kebutuhan keterampilan di masing-masing jurusan

1. Teknik Komputer dan Jaringan (TKJ)
  - Syarat Nilai:
  - $MTK \geq 80$
  - $BING \geq 75$
  - Alasan: TKJ membutuhkan kemampuan matematika yang tinggi untuk logika dan pemrograman. Bahasa Inggris menjadi syarat utama karena teknologi jaringan dan literatur teknis sering menggunakan bahasa ini. Bahasa Inggris diperlukan untuk menulis laporan teknis, sedangkan MTK memberikan dasar pemahaman tentang energi dan listrik dalam jaringan.
2. Teknik Kendaraan Ringan (TKR)
  - Syarat Nilai:
  - $MTK \geq 75$
  - $IPA \geq 75$
  - Alasan: Siswa yang memilih TKR membutuhkan pemahaman matematika untuk menghitung dimensi dan komponen kendaraan. IPA sangat penting untuk memahami prinsip fisika dalam mekanik kendaraan,
3. Teknik dan Bisnis Sepeda Motor (TBSM)
  - Syarat Nilai:
  - $MTK \geq 70$
  - $IPA \geq 70$
  - Alasan: TBSM memerlukan keterampilan teknis yang tinggi, di mana nilai matematika dan IPA penting untuk perhitungan teknis dasar.
4. Teknik Permesinan (TP)
  - Syarat Nilai:
  - $MTK \geq 75$
  - $IPA \geq 80$
  - Alasan: Teknik Permesinan membutuhkan matematika dan IPA yang kuat untuk memahami prinsip mekanika dan perhitungan teknis mesin. Matematika dan IPA penting untuk memahami dokumentasi teknis dan menyusun laporan secara akurat.
5. Teknik Audio Video (TAV)
  - Syarat Nilai:
  - $MTK \geq 75$
  - $IPA \geq 80$
  - Alasan: TAV membutuhkan kemampuan matematika untuk menghitung parameter teknis audio dan video. IPA yang kuat untuk memahami prinsip mekanika dan Elektronika
6. Teknik Instalasi Tenaga Listrik (TITL)
  - Syarat Nilai:
  - $MTK \geq 75$
  - $IPA \geq 80$

- Alasan: TITL memerlukan pemahaman matematika yang kuat untuk perhitungan listrik, dan IPA penting untuk memahami dasar-dasar fisika terkait kelistrikan.
- 7. Desain Pemodelan dan Informasi Bangunan (DPIB)
  - Syarat Nilai:
    - $MTK \geq 70$
    - $IPA \geq 70$
  - Alasan: Siswa yang memilih DPIB harus memiliki kemampuan matematika untuk menghitung dimensi bangunan dan IPA untuk memahami struktur material.
- 8. Rekayasa Perangkat Lunak (RPL)
  - Syarat Nilai:
    - $MTK \geq 75$
    - $IPA \geq 80$
  - Alasan: RPL membutuhkan kemampuan logika pemrograman yang tercermin dari nilai Matematika yang tinggi. IPA juga penting karena banyak dokumentasi perangkat lunak dan bahasa pemrograman menggunakan bahasa ini.

### 3.2. Hasil Perhitungan Association Rule

Analisis *Association Rule* dilakukan untuk memahami pola atau aturan yang berlaku antara berbagai faktor yang mempengaruhi pemilihan jurusan di SMK Negeri 4 Palembang. Pendekatan ini memungkinkan identifikasi hubungan antar atribut, seperti nilai akademik di berbagai mata pelajaran, yang relevan dengan tiap jurusan. Pada analisis ini, indikator utama yang digunakan adalah *support*, *confidence*, dan *lift*. Metrik-metrik ini memberikan gambaran seberapa kuat dan seringnya aturan tersebut berlaku dalam konteks pemilihan jurusan. Dengan menggunakan analisis ini, diharapkan dapat diperoleh wawasan mengenai preferensi siswa dalam memilih jurusan berdasarkan kriteria nilai tertentu

Table 2 Hasil Association Rule

No	Jurusan	Hasil Associaton Rule
1	Teknik Komputer dan Jaringan (TKJ)	Support (A → B): 0.48148148148148145 Confidence (A → B): 0.6265060240963856 Lift (A → B): 3.8402498884426595
2	Teknik Kendaraan Ringan (TKR)	Support (A → B): 0.5555555555555556 Confidence (A → B): 0.6 Lift (A → B): 3.6777777777777776
3	Teknik dan Bisnis Sepeda Motor (TBSM)	Support (A → B): 0.5694444444444444 Confidence (A → B): 0.5774647887323944 Lift (A → B): 5.309467918622849
4	Teknik Permesinan (TP)	Support (A → B): 0.4074074074074074 Confidence (A → B): 0.6197183098591549 Lift (A → B): 3.7986437141366713
5	Teknik Audio Video (TAV)	Support (A → B): 0.5454545454545454 Confidence (A → B): 0.72 Lift (A → B): 14.443636363636363
6	Teknik Instalasi Tenaga Listrik (TITL)	Support (A → B): 0.4153846153846154 Confidence (A → B): 0.5869565217391305 Lift (A → B): 5.977926421404683
7	Desain Pemodelan dan Informasi Bangunan (DPIB)	Support (A → B): 0.47572815533980584 Confidence (A → B): 0.6125 Lift (A → B): 3.936650485436893
8	Rekayasa Perangkat Lunak (RPL)	Support (A → B): 0.6153846153846154 Confidence (A → B): 0.6153846153846154 Lift (A → B): 6.267455621301775

Tabel 4.3 menunjukkan hasil perhitungan Association Rule yang mengungkapkan hubungan signifikan antara nilai akademik dan pemilihan jurusan di SMK Negeri 4 Palembang. Untuk Teknik Komputer dan Jaringan (TKJ), support 0.4814 dan confidence 0.6265 menunjukkan bahwa hampir 48% siswa memilih TKJ dengan kriteria nilai tertentu, dan 62.65% di antaranya cenderung memilih jurusan ini jika memenuhi kriteria nilai. Nilai lift 3.8402 mengindikasikan hubungan yang kuat antara kriteria nilai akademik dan pemilihan jurusan TKJ.

Untuk Teknik Kendaraan Ringan (TKR), support 0.5555 dan confidence 0.6 menunjukkan lebih dari setengah siswa memilih TKR dengan kriteria nilai tertentu, dengan asosiasi yang cukup kuat (lift 3.6667). Teknik Sepeda Motor (TSM) menunjukkan lift yang sangat tinggi, yaitu 5.3094, yang mengindikasikan korelasi yang sangat kuat antara nilai akademik dan pemilihan jurusan ini, dengan confidence 0.65.

Rekayasa Perangkat Lunak (RPL) memiliki nilai confidence 0.72 dan lift 14.443, menunjukkan asosiasi yang sangat kuat dengan nilai akademik, menandakan bahwa siswa dengan nilai baik lebih cenderung memilih jurusan ini. Teknik Audio Video (TAV) dan Desain Pemodelan dan Informasi Bangunan (DPIB) juga menunjukkan asosiasi signifikan, meskipun dengan lift yang lebih rendah. TAV memiliki support 0.3976, confidence 0.55, dan lift 2.9876, sedangkan DPIB memiliki support 0.4673, confidence 0.58, dan lift 3.0769.

Secara keseluruhan, hasil ini memberikan wawasan berharga untuk perencanaan akademik dan bimbingan di SMK Negeri 4 Palembang, dengan jurusan-jurusan seperti RPL, TSM, dan TKJ menunjukkan pengaruh yang kuat dari nilai akademik dalam pemilihan jurusan.

### 3.3. Hasil Penerapan Random Forest

Analisis menggunakan algoritma Random Forest bertujuan untuk mengevaluasi kemampuan model dalam memprediksi pemilihan jurusan berdasarkan data akademik siswa. Algoritma ini bekerja dengan menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi sekaligus mengurangi risiko overfitting. Pada pengujian ini, metrik yang digunakan untuk menilai performa model meliputi akurasi, presisi, recall, dan F1-Score.

Setiap metrik ini memberikan wawasan berbeda tentang kualitas prediksi: akurasi menunjukkan persentase keseluruhan prediksi yang benar; presisi menunjukkan ketepatan prediksi pada masing-masing jurusan; recall menggambarkan kemampuan model dalam menangkap seluruh kemungkinan pemilihan jurusan; dan F1-Score memberikan keseimbangan antara presisi dan recall. Dengan menganalisis keempat metrik ini, performa model dapat diukur secara komprehensif untuk masing-masing jurusan.

Table 3 Hasil Penerapan Random Forest

No	Jurusan	Hasil Penerapan Random Forest
1	Teknik Komputer dan Jaringan (TKJ)	Accuracy: 0.90 Precision: 0.94 Recall: 0.88 F1 Score: 0.91
2	Teknik Kendaraan Ringan (TKR)	Accuracy: 0.87 Precision: 0.80 Recall: 1.0 F1 Score: 0.89
3	Teknik dan Bisnis Sepeda Motor (TBSM)	Accuracy: 1.0 Precision: 1.0 Recall: 1.0 F1 Score: 1.0
4	Teknik Permesinan (TP)	Accuracy: 0.81 Precision: 0.80 Recall: 0.89 F1 Score: 0.8
5	Teknik Audio Video (TAV)	Accuracy: 0.9 Precision: 0.83 Recall: 1.0 F1 Score: 0.90
6	Teknik Instalasi Tenaga Listrik (TITL)	Accuracy: 0.8 Precision: 0.88 Recall: 0.72 F1 Score: 0.8
7	Desain Pemodelan dan Informasi Bangunan (DPIB)	Accuracy: 0.83 Precision: 0.83 Recall: 0.88 F1 Score: 0.85
8	Rekayasa Perangkat Lunak (RPL)	Accuracy: 0.85 Precision: 0.78



---

Recall: 1.0  
F1 Score: 0.88

---

Tabel 4.3 menunjukkan hasil penerapan model Random Forest untuk memprediksi pemilihan jurusan di SMK Negeri 4 Palembang berdasarkan nilai akademik siswa. Model ini mengukur akurasi, precision, recall, dan F1 score untuk masing-masing jurusan, dengan hasil yang bervariasi. Untuk Teknik Komputer dan Jaringan (TKJ), model mencapai accuracy 0.90, precision 0.94, recall 0.88, dan F1 score 0.91, menunjukkan kemampuan yang sangat baik dalam memprediksi siswa yang memilih TKJ, meskipun recall sedikit lebih rendah. Teknik Kendaraan Ringan (TKR) memiliki accuracy 0.87, precision 0.80, recall 1.0, dan F1 score 0.89, dengan recall yang sempurna meskipun precision sedikit lebih rendah. Teknik dan Bisnis Sepeda Motor (TBSM) menunjukkan kinerja sempurna dengan semua metrik (accuracy, precision, recall, F1 score) mencapai 1.0, menandakan prediksi yang sangat akurat. Teknik Permesinan (TP), dengan accuracy 0.81, precision 0.80, recall 0.89, dan F1 score 0.80, menunjukkan kinerja cukup baik meskipun precision dan F1 score sedikit lebih rendah.

Untuk Teknik Audio Video (TAV), model menunjukkan accuracy 0.90, precision 0.83, recall 1.0, dan F1 score 0.90, dengan recall yang sempurna meskipun precision sedikit lebih rendah. Teknik Instalasi Tenaga Listrik (TITL) memperoleh accuracy 0.80, precision 0.88, recall 0.72, dan F1 score 0.80, yang menunjukkan tantangan dalam mengidentifikasi siswa yang memilih TITL meskipun precision cukup tinggi. Desain Pemodelan dan Informasi Bangunan (DPIB) memiliki accuracy 0.83, precision 0.83, recall 0.88, dan F1 score 0.85, dengan kinerja yang baik dan keseimbangan antara precision dan recall. Terakhir, Rekayasa Perangkat Lunak (RPL) menunjukkan accuracy 0.85, precision 0.78, recall 1.0, dan F1 score 0.88, dengan recall sempurna meskipun precision lebih rendah.

Secara keseluruhan, model Random Forest menunjukkan kinerja yang baik di hampir semua jurusan, dengan F1 score berkisar antara 0.80 hingga 1.0, menjadikannya alat yang sangat berguna untuk bimbingan akademik dan perencanaan pemilihan jurusan di SMK Negeri 4 Palembang.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dijelaskan pada bab sebelumnya, maka penulis dapat membuat kesimpulan sebagai berikut:

1. Penelitian ini berhasil mengidentifikasi faktor-faktor utama yang mempengaruhi pemilihan jurusan siswa di SMK Negeri 4 Palembang. Dengan menggunakan metode Association Rule, ditemukan hubungan signifikan antara nilai akademik siswa, khususnya pada mata pelajaran matematika dan IPA, dengan pemilihan jurusan-jurusan tertentu. Hasil ini menunjukkan bahwa faktor akademik, terutama nilai dalam mata pelajaran yang terkait dengan bidang teknis, memiliki pengaruh yang kuat terhadap pemilihan jurusan yang berkaitan dengan teknologi dan sains.
2. Dengan tujuan untuk mengembangkan alat bantu prediksi yang dapat digunakan dalam pengambilan keputusan, penelitian ini menerapkan metode Random Forest untuk memprediksi pilihan jurusan siswa. Model prediksi yang dihasilkan memiliki akurasi yang tinggi, yang menunjukkan bahwa metode ini efektif dalam memprediksi pilihan jurusan siswa berdasarkan kombinasi faktor akademik dan non-akademik. Model ini memberikan wawasan lebih dalam tentang bagaimana kombinasi faktor tersebut mempengaruhi pemilihan jurusan, dan dapat digunakan sebagai alat untuk membantu pihak sekolah dalam memberikan rekomendasi jurusan yang lebih tepat bagi siswa.

#### DAFTAR PUSTAKA

- [1] S. Setiawansyah, H. Sulistiani, and V. H. Saputra, "Penerapan Codeigniter Dalam Pengembangan Sistem Pembelajaran Dalam Jaringan Di SMK 7 Bandar Lampung," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 2, p. 89, 2020, doi: 10.24014/coreit.v6i2.10679.
- [2] SMKN4PALEMBANG, "Profil SMK Negeri 4 Palembang," SMKN4PALEMBANG. Accessed: Sep. 15, 2024. [Online]. Available: <https://smkn4palembang.sch.id/read/2/profil>
- [3] V. R. Prasetyo, M. Widiastri, and M. M. Angkiriwang, "Sistem Berbasis Web Untuk Koreksi Soal Esai Association Rules," *Tek. J. Teknol. Inf. dan Komun.*, vol. 11, no. 1, pp. 62–68, 2022.
- [4] L. Ratnawati and D. R. Sulistyningrum, "Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel," *J. Sains dan Seni ITS*, vol. 8, no. 2, 2020, doi: 10.12962/j23373520.v8i2.48517.
- [5] F. H. Moh'd, K. A. Notodiputro, and Y. Angraini, "Enhancing interpretability in random forest: Leveraging inTrees for association rule extraction insights," *IAES Int. J. Artif. Intell.*, vol. 13, no. 4, pp.

- 4054–4061, 2024, doi: 10.11591/ijai.v13.i4.pp4054-4061.
- [6] H. Ilma, K. A. Notodiputro, and B. Sartono, “Association Rules in Random Forest for the Most Interpretable Model,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 1, pp. 0185–0196, 2023, doi: 10.30598/barekengvol17iss1pp0185-0196.
- [7] M. N. F. Lubis, “Analisa jenis dan jumlah sel leukosit pada penderita tuberculosis yang menjalani pengobatan obat anti tuberculosis selama 2 bulan di Rumah Sakit Khusus Paru Medan,” *Skripsi*, pp. 1–46, 2021, [Online]. Available: <http://repository.uma.ac.id/handle/123456789/15704>
- [8] I. Fister, I. Fister, D. Fister, V. Podgorelec, I. Fister, and S. Salcedo-Sanz, “A comprehensive review of visualization methods for association rule mining: Taxonomy, challenges, open problems and future ideas,” *Expert Syst. Appl.*, vol. 233, 2023, doi: 10.1016/j.eswa.2023.120901.
- [9] C. Fernandez-Basso, M. D. Ruiz, and M. J. Martin-Bautista, “New Spark solutions for distributed frequent itemset and association rule mining algorithms,” *Cluster Comput.*, vol. 27, no. 2, pp. 1217–1234, 2024, doi: 10.1007/s10586-023-04014-w.
- [10] W. C. Lin *et al.*, “Tree-Based Algorithms and Association Rule Mining for Predicting Patients’ Neurological Outcomes After First-Aid Treatment for an Out-of-Hospital Cardiac Arrest During COVID-19 Pandemic: Application of Data Mining,” *Int. J. Gen. Med.*, vol. 15, no. September, pp. 7395–7405, 2022, doi: 10.2147/IJGM.S384959.
- [11] G. Dudek, “A Comprehensive Study of Random Forest for Short-Term Load Forecasting,” *Energies*, vol. 15, no. 20, p. 7547, Oct. 2022, doi: 10.3390/en15207547.
- [12] K. Rahayu, V. Fitria, D. Septhya, R. Rahmaddeni, and L. Efrizoni, “Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 108–114, 2023, doi: 10.57152/malcom.v3i2.780.
- [13] J. S. Rhodes, A. Cutler, and K. R. Moon, “Geometry- and Accuracy-Preserving Random Forest Proximities,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10947–10959, Sep. 2023, doi: 10.1109/TPAMI.2023.3263774.
- [14] M. Alduailij, Q. W. Khan, M. Tahir, M. Sardaraz, M. Alduailij, and F. Malik, “Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method,” *Symmetry (Basel)*, vol. 14, no. 6, p. 1095, May 2022, doi: 10.3390/sym14061095.
- [15] Y. Kim and Y. Kim, “Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models,” *Sustain. Cities Soc.*, vol. 79, p. 103677, Apr. 2022, doi: 10.1016/j.scs.2022.103677.
- [16] S. Amini, M. Saber, H. Rabiei-Dastjerdi, and S. Homayouni, “Urban Land Use and Land Cover Change Analysis Using Random Forest Classification of Landsat Time Series,” *Remote Sens.*, vol. 14, no. 11, p. 2654, Jun. 2022, doi: 10.3390/rs14112654.
- [17] Adlina Khairunnisa, Khairil Anwar Notodiputro, and Bagus Sartono, “A Study of Rule Extraction from Double Random Forest to Identify the Characteristics of Working Poor in Jakarta Province, Indonesia,” *Int. J. Sci. Res. Sci. Eng. Technol.*, pp. 258–266, 2023, doi: 10.32628/ijrsrset231069.
- [18] Y. Aldi, E. Husni, and R. Yesika, “Activity of kincung flowers (*Etilingera Elatior* (Jack) R.M.Sm.) on total leukocytes and percentage of leukocytes in allergic male white mice,” *Pharmacogn. J.*, vol. 12, no. 1, pp. 44–51, 2020, doi: 10.5530/pj.2020.12.8.
- [19] A. Alfando and R. Hayami, “KLASIFIKASI TEKS BERITA BERBAHASA INDONESIA MENGGUNAKAN MACHINE LEARNING DAN DEEP LEARNING: STUDI LITERATUR,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 681–686, Mar. 2023, doi: 10.36040/jati.v7i1.6486.
- [20] E. Bu’ulolo, “Algoritma K-Nearest Neighbor (K-NN) Dengan Normalisasi Max Min Untuk Menentukan Calon Mahasiswa Yang Layak Menerima KIP Kuliah Merdeka,” *Simkom*, vol. 9, no. 2, pp. 184–194, Jul. 2024, doi: 10.51717/simkom.v9i2.445.
- [21] L. M. Crivei, G. Czibula, and A. Mihai, “A Study on Applying Relational Association Rule Mining Based Classification for Predicting the Academic Performance of Students,” in *Knowledge Science, Engineering and Management*, C. Douligieris, D. Karagiannis, and D. Apostolou, Eds., Cham: Springer International Publishing, 2019, pp. 287–300. doi: 10.1007/978-3-030-29551-6\_25.
- [22] A. Girdhar, H. Kapur, and V. Kumar, “Classification of White blood cell using Convolution Neural Network,” *Biomed. Signal Process. Control*, vol. 71, p. 103156, 2022, doi: <https://doi.org/10.1016/j.bspc.2021.103156>.

- [23] S. Irie, H. Aman, S. Amasaki, T. Yokogawa, and M. Kawahara, "A Comparative Study of Hybrid Fault-Prone Module Prediction Models Using Association Rule and Random Forest," in *2023 The 5th World Symposium on Software Engineering (WSSE)*, New York, NY, USA: ACM, Sep. 2023, pp. 33–38. doi: 10.1145/3631991.3631996.
- [24] H. S. Kaduham and H. M. Abduljabbar, "Studying the Classification of Texture Images by K-Means of Co-Occurrence Matrix and Confusion Matrix," *Ibn AL-Haitham J. Pure Appl. Sci.*, vol. 36, no. 1, pp. 113–122, 2023, doi: 10.30526/36.1.2894.
- [25] M. R. Kusnaldi, T. Gulo, and S. Aripin, "Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 330–338, 2022, doi: 10.47065/josyc.v3i4.2112.
- [26] N. T. Luchia, M. Mustakim, N. Noviarni, K. Sussolaikah, and T. Arifianto, "Feature Selection In Support Vector Machine And Random Forest Algorithms For The Classification Of Recipients Of The Smart Indonesia Program," in *2024 International Conference on Circuit, Systems and Communication (ICCSC)*, IEEE, Jun. 2024, pp. 1–6. doi: 10.1109/ICCSC62074.2024.10616886.
- [27] S. Matharaarachchi, M. Domaratzki, A. Katz, and S. Muthukumarana, "Discovering Long COVID Symptom Patterns: Association Rule Mining and Sentiment Analysis in Social Media Tweets," *JMIR Form. Res.*, vol. 6, no. 9, pp. 1–15, 2022, doi: 10.2196/37984.
- [28] I. Permana and F. N. S. Salisah, "Pengaruh Normalisasi Data Terhadap Performa Hasil Klasifikasi Algoritma Backpropagation," *Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 1, pp. 67–72, 2022, doi: 10.57152/ijirse.v2i1.311.
- [29] K. Radhika, S. Valarmathy, S. Selvarasu, K. Bashkaran, and C. Srinivasan, "Predictive Road Sign Maintenance Using Random Forest Regression and IoT Data," in *2023 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, IEEE, Nov. 2023, pp. 348–353. doi: 10.1109/ICSCNA58489.2023.10370728.
- [30] V. K. Rai, S. Chakraborty, and S. Chakraborty, "Association Rule Mining for Prediction of Covid-19," *Decis. Mak. Appl. Manag. Eng.*, vol. 6, no. 1, pp. 365–378, 2023, doi: 10.31181/dmame0317102022r.
- [31] F. Riskiyono and D. Mahdiana, "Implementation of Random Forest Algorithm for Graduation Prediction," *Sinkron*, vol. 8, no. 3, pp. 1662–1670, 2024, doi: 10.33395/sinkron.v8i3.13750.
- [32] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, pp. 1–21, 2021, doi: 10.1007/s42979-021-00592-x.
- [33] A. Sayal *et al.*, "Neural Networks And Machine Learning," in *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, 2023, pp. 58–63. doi: 10.1109/ICCCMLA58983.2023.10346612.
- [34] M. Sholeh, D. Andayati, and R. Y. Rachmawati, "Data Mining Model Klasifikasi Menggunakan Algoritma K-Nearest Neighbor Dengan Normalisasi Untuk Prediksi Penyakit Diabetes," *TeIKa*, vol. 12, no. 02, pp. 77–87, 2022, doi: 10.36342/teika.v12i02.2911.
- [35] A. Singh, "Foundations of Machine Learning," *SSRN Electron. J.*, no. 112, pp. 1–8, 2019, doi: 10.2139/ssrn.3399990.
- [36] A. Siswanto, A. Fadlil, and Y. Anton, "Terakreditasi SINTA Peringkat 4 Ekstraksi Ciri Metode Gray Level Co-Occurrence Matrix Untuk Identifikasi Sel Darah Putih," *JOINTECS (Journal Inf. Technol. Comput. Sci.)*, vol. 3, no. 1, pp. 71–80, 2020.
- [37] M. K. Suryadewiansyah, T. Endra, and E. Tju, "Naïve Bayes dan Confusion Matrix untuk Efisiensi Analisa Intrusion Detection System Alert," *J. Nas. Teknol. dan Sist. Inf.*, vol. 8, no. 2, pp. 81–88, 2020.
- [38] R. A. F. Ulfa Windi Humaira, Nailia Nurul Hikmah, "Identifikasi Demografi Dan Faktor Pemilihan," *At-Thullab J.*, vol. 1, pp. 71–84, 2019.
- [39] S. Violina, N. R. Damayanti, and I. O. Herlistiono, "Detection and Classification of Blood Cancer Using Deep Learning Framework," *J. Inf. Technol. Comput. Sci.*, vol. 1095, pp. 159–165, 2024, doi: 10.1007/978-981-99-7077-3\_16.