



Integration of K-Means and Random Forest for Recommendation of Nutritious Food Menu (MBG) Based on Recipient Age

Integrasi K-Means dan Random Forest untuk Rekomendasi Menu Makanan Bergizi (MBG) Berdasarkan Usia Penerima

**Miftahul Jannah^{1*}, Ajang Sopandi², Ryan Zulham Ramadhani³,
Hengki Rusdianto⁴, Khelvin Ovella Putra⁵**

¹Program Studi Teknik Informatika, Politeknik Negeri Bengkalis, Indonesia

^{2,3,4}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang, Indonesia

⁵Program Studi Manajemen Informatika, Fakultas Vokasi, Institut Teknologi Mitra Gama, Indonesia

E-Mail: ¹miftahuljannah@polbeng.ac.id, ²ajangsopandi@ft-umt.ac.id, ³ryanzulham@ft-umt.ac.id,

⁴hengki.rusdianto@gmail.com, ⁵khelvinovellaputra@info@mitra-gama.ac.id

Received Apr 12th 2025; Revised May 12th 2026; Accepted Jun 09th 2026; Available Online Jun 20th 2026

Corresponding Author: Miftahul Jannah

Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

The Free Nutritious Meal Program (MBG) aims to improve the nutritional quality of children as an effort to support the growth and health of students [MJ1.1], however, the implementation of the MBG food menu provided every day is still not in accordance with the needs of students at school because it is not based on age groups. Differences in age, weight, and height affect nutritional needs so that a data-based approach is needed to produce appropriate MBG food menu recommendations. This study proposes the integration of the K-Means Clustering and Random Forest algorithms to group recipients based on student age groups so that the nutrition provided is in accordance with WHO nutritional status recommendations and the Nutritional Adequacy Rate (AKG) of the Indonesian Ministry of Health. The data used amounted to 1190 food menu data as an initial study for system development. The clustering results were used as the basis for grouping food menus based on nutritional categories from food menu data and using Random Forest in determining daily menus based on Nutritional Adequacy Rate (AKG) data. The results of the study show that the proposed method is able to provide recommendations for nutritious food menus automatically according to individual characteristics. This study shows that the integration of clustering and classification methods is effective in supporting a data-based nutritious food menu recommendation system.

Keywords: *Clustering, K-Means, Nutritious Food, Random Forest, Recommendation System*

Abstrak

Program Makan Bergizi Gratis (MBG) bertujuan meningkatkan kualitas gizi anak-anak sebagai upaya mendukung pertumbuhan dan kesehatan peserta didik, namun implementasinya menu makanan MBG yang diberikan setiap harinya masih belum sesuai dengan kebutuhan peserta didik disekolah karena tidak berdasarkan kelompok usia. Perbedaan usia, berat badan, dan tinggi badan memengaruhi kebutuhan nutrisi sehingga diperlukan pendekatan berbasis data untuk menghasilkan rekomendasi menu makanan MBG yang tepat. Penelitian ini mengusulkan integrasi algoritma K-Means Clustering dan Random Forest untuk mengelompokkan penerima berdasarkan kelompok usia peserta didik sehingga nutrisi yang diberikan sesuai dengan rekomendasi status gizi WHO serta Angka Kecukupan Gizi (AKG) Kementerian Kesehatan RI. Data yang digunakan berjumlah 1190 data menu pangan sebagai studi awal pengembangan sistem. Hasil clustering digunakan sebagai dasar pengelompokan menu makanan berdasarkan kategori gizi dari data menu pangan dan menggunakan Random Forest dalam menentukan menu harian berdasarkan data AKG. Hasil penelitian menunjukkan bahwa metode yang diusulkan mampu memberikan rekomendasi menu makanan bergizi secara otomatis sesuai dengan karakteristik individu. Penelitian ini menunjukkan bahwa integrasi metode clustering dan klasifikasi efektif dalam mendukung sistem rekomendasi menu makanan bergizi berbasis data.

Kata Kunci: *Clustering, K-Means, Makanan Bergizi, Random Forest, Rekomendasi Menu*

1. PENDAHULUAN

Makan Bergizi Gratis (MBG) merupakan salah satu program unggulan yang dicanangkan oleh pemerintahan Prabowo–Gibran yang bertujuan untuk mengatasi permasalahan kekurangan gizi dan upaya meningkatkan kualitas Sumber Daya Manusia (SDM) di Indonesia. Sasaran utama program MBG adalah anak-anak, pelajar dan ibu hamil. Hal tersebut berdasarkan pada data dari Kementerian Kesehatan dan Kemenko Pembangunan Manusia dan Kebudayaan yang menunjukkan 41% siswa mengalami kelaparan yang berdampak pada menurunnya kualitas Pendidikan [1]. Asupan gizi yang baik sangat diperlukan untuk mendukung efektivitas proses pembelajaran serta dalam upaya mencetak warga negara yang berkualitas. Penyediaan program makan bergizi menjadi salah satu langkah strategis dalam meningkatkan kualitas pendidikan sekaligus kualitas sumber daya manusia [2]. meningkatkan kualitas gizi anak-anak sebagai upaya mendukung pertumbuhan dan kesehatan peserta didik. Ini akan membantu meningkatkan fokus. Status gizi yang ideal juga penting untuk menjamin pertumbuhan dan perkembangan yang sehat [3].

Namun, dalam implementasinya masih ditemukan berbagai kendala, terutama terkait ketidaksesuaian menu makanan dengan kebutuhan gizi berdasarkan usia penerima dan implementasinya telah menimbulkan kekhawatiran publik tentang keamanan dan pengelolaan pangan. [4] Tercukupinya asupan makanan seimbang harian anak perlu memperhatikan pemenuhan zat gizi seperti karbohidrat, protein, lemak, vitamin dan mineral yang terkandung dalam nasi, lauk, sayur, buah, susu dan air. Asupan gizi yang baik akan menghasilkan status gizi anak yang baik pula untuk menciptakan sumberdaya manusia yang berkualitas yaitu sehat, cerdas, dan memiliki fisik yang tangguh serta produktif [5]. Oleh karena itu, pemberian menu makanan yang bersifat umum tanpa mempertimbangkan karakteristik tersebut dapat menyebabkan intervensi gizi menjadi kurang optimal. Perbedaan kebutuhan nutrisi pada setiap kelompok usia menuntut adanya penyesuaian komposisi makanan yang tepat agar intervensi gizi dan program MBG dapat berjalan secara optimal.

Untuk mengatasi masalah tersebut maka dibutuhkan bantuan teknologi dengan memanfaatkan teknik data mining dalam menganalisis data yang ada. Salah satu algoritma yang dapat digunakan adalah K-means *clustering* untuk mengelompokkan objek atau data berdasarkan kesamaan karakteristik tertentu. Teknik *clustering* merupakan salah satu metode yang utama dalam data mining, yang mana berfungsi untuk mengelompokkan data ke dalam beberapa *cluster* berdasarkan Tingkat kemiripan karakteristik yang dimilikinya. Dengan pendekatan ini, data yang memiliki karakteristik sama akan dikelompokkan ke dalam satu *cluster*, namun data dengan karakteristik berbeda akan ditempatkan dalam *cluster* yang berbeda. Salah satu algoritma *clustering* yang paling populer dan banyak digunakan adalah K-Means. Algoritma ini bekerja dengan membagi data ke dalam sejumlah *cluster* yang telah ditentukan sebelumnya, dengan tujuan mengurangi jarak antara data dalam satu *cluster* terhadap pusatnya (*centroid*) [6]. Algoritma K-Means merupakan pendekatan yang efektif dalam dunia analisis data dan pengelompokan objek-objek serupa ke dalam *cluster-cluster* yang terpisah [7]. Namun dengan mengintegrasikan algoritma random forest berpotensi untuk menghasilkan sistem yang lebih adaptif dan akurat dalam memberikan rekomendasi. Algoritma random forest ini menggunakan banyak pohon keputusan yang dibangun secara acak, dan hasilnya adalah kombinasi prediksi dari seluruh pohon. Cara kerja algoritma ini adalah dengan membuat beberapa pohon keputusan dan menyatukannya untuk menghasilkan prediksi yang lebih akurat dan stabil [8].

Penelitian sebelumnya yang dilakukam oleh [9] menunjukkan bahwa algoritma K-Means *clustering* dapat menentukan kelompok menu makanan berdasarkan kalori harian. Peneliti lainnya yang juga memanfaatkan algoritma K-Means *clustering* dalam memberikan rekomendasi resep yang sesuai dengan profil gizi pengguna dan berpotensi mendukung pola makan yang lebih terarah serta personal dengan hasil yang akurat [10]. Penelitian lain juga menunjukkan bahwa integrasi algoritma K-Means *clustering* dan Random Forest berhasil membangun system klasifikasi kerusakan jalan dengan menggabungkan algoritma K-Means untuk pengelompokkan dan Random Forest untuk klasifikasi [11]. Integrasi algoritma tersebut dapat memberikan akurasi yang tinggi.

Meskipun berbagai penelitian telah memanfaatkan K-Means maupun Random Forest secara terpisah, penerapan integrasi kedua metode tersebut dalam konteks rekomendasi menu makanan bergizi berdasarkan usia penerima masih terbatas. Padahal, pendekatan ini berpotensi menghasilkan sistem yang lebih adaptif dan akurat dalam memberikan rekomendasi. Oleh karena itu, diperlukan penelitian yang mengombinasikan kedua metode tersebut untuk mengatasi permasalahan ketidaktepatan menu dalam program makan bergizi gratis.

Berdasarkan latar belakang tersebut, tujuan dari penelitian ini adalah untuk mengembangkan model rekomendasi menu makanan bergizi dengan mengintegrasikan algoritma K-Means dan Random Forest. Secara khusus, penelitian ini bertujuan untuk mengelompokkan menu makanan bergizi gratis berdasarkan indikator antropometri seperti usia, berat badan, dan tinggi badan yang mengacu pada standar kebutuhan gizi dan klasifikasi status gizi WHO serta Angka Kecukupan Gizi (AKG) Kementerian Kesehatan RI [12]. Dengan demikian, diharapkan hasil penelitian ini dapat meningkatkan ketepatan pemberian menu makanan berdasarkan kelompok usia serta mendukung pengambilan keputusan berbasis data dalam program makan bergizi gratis.

2. METODOLOGI PENELITIAN

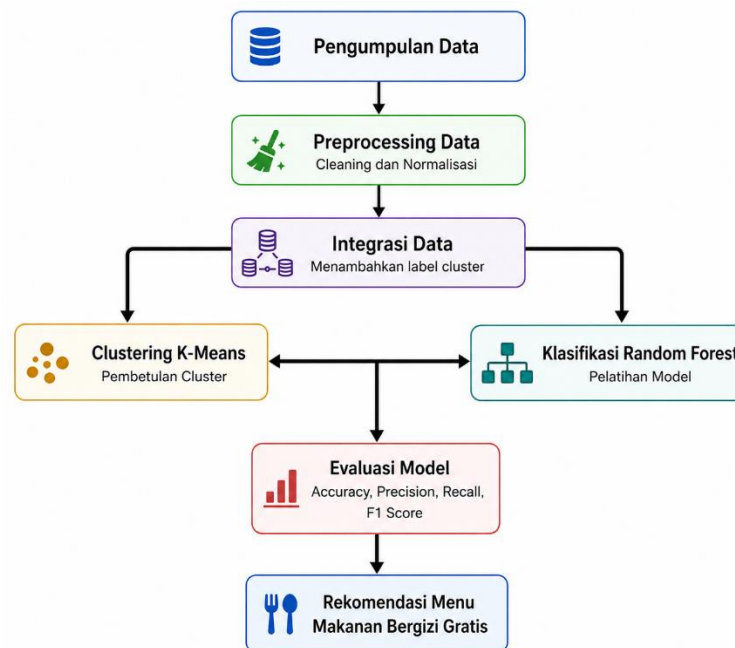
2.1. Data dan Bahan

Penelitian ini menggunakan data sekunder yang diperoleh dari salah satu mitra SPPG Umban Sari yang berkaitan dengan karakteristik penerima program makan bergizi gratis. Data yang digunakan meliputi data Angka Kecukupan Gizi (AKG) berdasarkan juknis Badan Gizi Nasional serta data komposisi pangan yang diperoleh dari dapur MBG. Variabel yang digunakan meliputi kelompok sasaran, energi, protein, lemak dan karbohidrat. Dengan jumlah data 1185 komposisi pangan Indonesia.

Selain itu, data menu makanan digunakan sebagai label target yang terdiri dari beberapa kategori. Data diperoleh dari sumber terkait seperti data kesehatan masyarakat atau hasil survei yang telah disesuaikan dengan kebutuhan penelitian.

2.2. Tahapan Penelitian

Untuk menghasilkan penelitian yang berkualitas dengan tujuan yang terarah, diperlukan langkah-langkah penelitian yang tersusun secara sistematis. Tahapan-tahapan yang perlu dilakukan dalam penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

1. Pengumpulan Data

Data penelitian diperoleh dari salah satu mitra SPPG Umban Sari. Data yang digunakan dalam penelitian ini adalah data AKG berdasarkan juknis Badan Gizi Nasional serta data komposisi pangan yang diperoleh dari dapur MBG tersebut. Adapun jumlah data komposisi pangan yang akan diproses menggunakan algoritma K-Means dan Random Forest pada penelitian ini adalah sebanyak 1185 data. Variabel data yang digunakan meliputi kelompok sasaran, energi, protein, lemak dan karbohidrat. Data diperoleh melalui proses pengumpulan secara langsung melalui mitra MBG maupun dari sumber pendukung yang relevan dengan penelitian. Selanjutnya, data yang telah diperoleh digunakan sebagai bahan dalam proses analisis untuk mendukung penerapan metode data mining, khususnya algoritma K-Means dan Random Forest dalam menghasilkan rekomendasi menu makanan bergizi berdasarkan usia penerima.

2. Preprocessing Data

Setelah proses identifikasi dan pengumpulan data dalam penelitian ini, tahapan selanjutnya adalah *preprocessing* data. *Preprocessing* data adalah serangkaian langkah yang dilakukan untuk membersihkan, mengubah, atau menyesuaikan data mentah sebelum data tersebut digunakan untuk pemodelan atau analisis lebih lanjut [13]. Tujuan utama dari *preprocessing* data adalah untuk mempersiapkan data sehingga dapat diolah dengan benar oleh algoritma analisis, termasuk di dalamnya algoritma *Machine Learning*. Selain itu, *preprocessing* data juga bertujuan untuk meningkatkan kualitas dan keakuratan hasil akhir dari analisis yang dilakukan [14].

3. *Clustering* Menggunakan K-Means

Pada tahap ini dilakukan pengelompokan data menggunakan algoritma K-Means. Algoritma K-means dalam *clustering* analysis merupakan salah satu algoritma yang populer digunakan dalam pengelompokan. K-means juga memberikan kemudahan dalam pengelompokan, karena penentuan jumlah *cluster* ditentukan terlebih dahulu sebelum melakukan analisis [15]. Algoritma ini bekerja dengan membagi data ke dalam beberapa *cluster* berdasarkan kemiripan karakteristik. Adapun langkah-langkah algoritma K-Means sebagai berikut :

- Menentukan jumlah *cluster* (k)
- Menginisialisasi *centroid*
- Menghitung jarak data ke *centroid*

$$d(x, y) = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (1)$$

Keterangan :

x : data

y : *centroid*

n : jumlah atribut

- Mengelompokkan data ke *cluster* terdekat
- Menghitung ulang *centroid* menggunakan rata-rata data dalam klaster
- Mengulangi proses hingga *centroid* stabil atau mencapai iterasi maksimum [16]

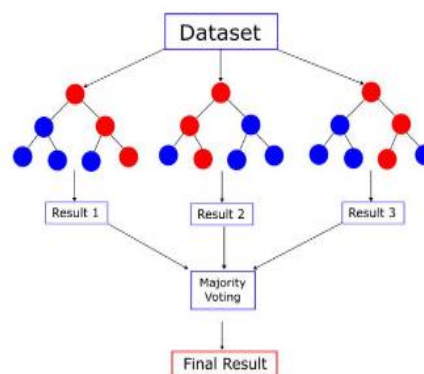
4. Integrasi Data (*Feature Engineering*)

Hasil dari proses *clustering* berupa label *cluster* kemudian ditambahkan sebagai fitur baru dalam dataset. Tahap ini dikenal sebagai *feature engineering*. Penambahan fitur ini bertujuan untuk memberikan informasi tambahan kepada model, meningkatkan performa klasifikasi, membantu model memahami pola data yang lebih kompleks. Integrasi antara *clustering* dan klasifikasi terbukti dapat meningkatkan akurasi model dalam berbagai studi data mining.

5. Klasifikasi Menggunakan Random Forest (lihat Gambar 2)

Tahap selanjutnya adalah membangun model klasifikasi menggunakan algoritma Random Forest. Model Random Forest merupakan salah satu teknik *Machine Learning* untuk mengatasi masalah klasifikasi maupun regresi. Cara kerja teknik ini adalah membuat beberapa pohon keputusan yang dibangun pada subset acak dari data pelatihan, dan prediksi akhir didasarkan pada agregat hasil dari pohon-pohon tersebut [17]. Adapun langkah-langkah dalam klasifikasi menggunakan Random Forest sebagai berikut :

- Membuat beberapa subset data (*bootstrap sampling*)
- Membangun decision tree dari setiap subset
- Menggabungkan hasil prediksi (*voting*)



Gambar 2. Random Forest

6. Evaluasi Model

Evaluasi dilakukan untuk mengukur performa model klasifikasi. Model Random Forest Regressor di evaluasi menggunakan beberapa metrik utama untuk menilai kinerjanya. Tujuannya adalah menarik kesimpulan dari hasil penerapan algoritma untuk memahami kinerja, keberhasilan, serta mengidentifikasi pola-pola atau karakteristik yang muncul selama proses learning [18]. Metode evaluasi yang digunakan adalah *confusion matrix* dengan metrik:

- a. *Accuracy*: tingkat ketepatan prediksi

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Keterangan : TP (*True Positive*)
 TN (*True Negative*)
 FP (*False Positive*)
 FN (*False Negative*)

- b. *Precision*: ketepatan prediksi positif
 c. *Recall*: kemampuan menemukan data positif
 d. *F1-Score*: keseimbangan antara precision dan recall
 Evaluasi ini penting untuk memastikan bahwa model yang dihasilkan memiliki performa yang baik dan dapat digunakan secara praktis

7. Rekomendasi Menu Makanan Bergizi

Tahap akhir adalah menghasilkan rekomendasi menu makanan berdasarkan hasil prediksi model. Output sistem berupa kategori menu yang sesuai dengan karakteristik usia dan kebutuhan gizi penerima.

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan dari penerapan algoritma K-Means dan Random Forest dalam memberikan rekomendasi menu makanan di tingkat PAUD, SD Kelas 1-3, SD Kelas 4-6, SMP dan SMA pada program Makan Bergizi Gratis (MBG). Pemilihan kelompok usia tersebut didasarkan pada sasaran utama program MBG adalah peserta didik dibangku sekolah mulai dari PAUD hingga Menengah Atas.

Setiap tahap penelitian dijelaskan secara terstruktur, mulai dari pengumpulan data, proses seleksi dan prapemrosesan, tahap transformasi, penerapan algoritma K-Means, integrasi data, penerapan algoritma Random Forest hingga evaluasi dan analisis hasil pengelompokan, untuk memperoleh pola rekomendasi menu yang dapat dijadikan landasan pertimbangan strategis bagi pihak dapur SPPG dalam mengambil keputusan menu makan harian berdasarkan kelompok usia.

3.1. Pengumpulan Data

Tahap awal pada penelitian ini adalah mengumpulkan data dari salah satu dapur MBG. Data yang digunakan dalam penelitian ini adalah data rujukan Rujukan Angka Kecukupan Gizi Juknis Bgn No. 401.1 Tahun 2025 terlihat pada Tabel 1 [19]. Data selanjutnya adalah tabel rekomendasi komposisi pangan Indonesia yang didapat dari dapur MBG tersebut. Variabel yang digunakan dalam penelitian meliputi Energi (Kkal), Protein (gram), Lemak (gram), Karbohidrat (gram). Data dikumpulkan dalam format Microsoft Excel dan diolah menggunakan bahasa pemrograman Python melalui Google Colab. Tabel 2 menampilkan data yang peneliti gunakan pada studi ini. Data tersebut digunakan sebagai dasar dalam proses analisis dan rekomendasi menu makanan bergizi yang sesuai dengan karakteristik penerima MBG.

Tabel 1.Data Rujukan Angka Kecukupan Gizi (AKG)

No.	Kelompok Sasaran	Pendistri busian MBG	Rujukan % AKG	Energi (kkal)		Protein (g)		Lemak (g)		KH (g)	
				Min	Max	Min	Max	Min	Max	Min	Max
1	Siswa TK/PAUD/LB	Pagi	20-25%	280	350	5	6,3	10	12,5	44	55
2	Siswa SD/MI/SLB Kelas 1-3	Pagi	20-25%	330	413	8	10	11	13,8	50	62,5
3	Siswa SD/MI/SLB Kelas 4-6	Siang	30-35%	585	683	15,8	18,4	19,5	22,8	87	101,5
4	Siswa SMP/MTS/SMPLB	Siang	30-35%	668	779	20,3	23,6	22,5	26,3	97,7	113,8
5	Siswa SMA/SMK/MK/MASMALB	Siang	30-35%	713	831	21	24,5	22,5	26,3	105	122,5

3.2. Preprocessing Data

Bagian *preprocessing* ini sangat penting untuk dilakukan. *Preprocessing* data dilakukan dengan beberapa tahap yang dimulai dari *data selection*. Pada tahapan ini dilakukan pemilihan atribut yang akan digunakan dalam pemrosesan algoritma K-Means dan Random Forest. Atribut yang digunakan dalam penelitian ini meliputi nama bahan makanan, energi, protein, lemak, dan karbohidrat. Pemilihan atribut

dilakukan karena variabel tersebut merupakan komponen utama dalam penentuan angka kecukupan gizi dan penyusunan menu makanan harian siswa.

Setelah tahapan selection menentukan atribut maka selanjutnya *cleaning data* yaitu membersihkan data yang kosong, mengubah data menjadi numerik, menghapus data tidak valid, menghapus duplikasi data, memastikan dataset siap diproses. Hasil dari data *cleaning data* dapat dilihat pada Tabel 3.

Tabel 2. Komposisi Pangan Menu Makanan

				Tabel Komposisi Pangan SPPG Umban Sari									
No	Nama Bahan	Sumber	Berat Std	% Bdd	Energi	Protein	Lemak	KH	Serat	Abu	Kalsium	Fosfor	Besi
				(%)	(Kal)	(g)	(g)	(g)	(g)	(g)	(mg)	(mg)	(mg)
1	Abon Ayam, hypermart	FATSECRET	100	100	711,0	30,0	19,0	0,0					
2	Abon ikan	KZGMI-2001	100	100	435,0	27,2	20,2	36,1		10,1			
3	Agar-Agar Plain	FATSECRET	100	100	10,0	0,0	0,0	3,0	0,0				
4	Akar tonjong, segar	KZGPI-1990	100	100	45,0	1,1	0,4	10,8	1,2	1,0	72,0	125,0	1,3
5	Ale,toge, segar	KZGPI-1990	100	100	37,0	4,4	0,5	3,8	1,7	0,7	50,0	248,0	2,0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1186	Wortel, kukus	KZGPI-1990	100	100	37,0	1,0	0,6	8,3	1,1	0,6	44,0	30,0	0,6
1187	Wortel, rebus	KZGPI-1990	100	80	28,0	0,7	0,5	6,3	0,8	0,4	46,0	28,0	0,6
1188	Wortel, segar	KZGPI-1990	100	80	36,0	1,0	0,6	7,9	1,0	0,6	45,0	74,0	1,0
1189	Yangko	KZGPI-1990	100	100	254,0	3,0	1,1	58,1	0,8	0,4	184,0	24,0	2,6
1190	Yoghurt, segar	DABM-1964	100	100	52,0	3,3	2,5	4,0	0,0	2,2	120,0	90,0	0,1

Tabel 3. *Cleaning data* Menu Makanan

Nama Bahan	Energi	Protein	Lemak	Karbohidrat
Abon Ayam, hypermart	711	30	19	0
Abon ikan	435	27,2	20,2	36,1
Agar-Agar Plain	10	0	0	3
Akar tonjong, segar	45	1,1	0,4	10,8
...	...	...	...	...
Wortel, kukus	37	1	0,6	8,3
Wortel, rebus	28	0,7	0,5	6,3
Wortel, segar	36	1	0,6	7,9
Yangko	254	3	1,1	58,1
Yoghurt, segar	52	3,3	2,5	4

Selanjutnya, tahap transformasi dilakukan agar data dapat diproses oleh algoritma *machine learning*. Kemudian dilakukan normalisasi menggunakan StandardScaler untuk menyamakan skala antarvariabel sebelum dilakukan proses *clustering* dengan metode K-Means dan klasifikasi dengan Random Forest.

Tabel 4. Hasil Normalisasi

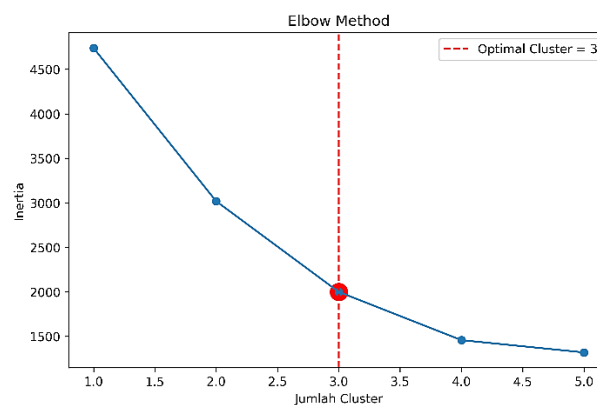
Energi	Protein	Lemak	Karbohidrat
3,073192	1,78471	0,888329	-0,95489
1,416925	1,53975	0,978902	0,494584
-1,13349	-0,83987	-0,54575	-0,83444
-0,92345	-0,74363	-0,51556	-0,52126
-0,97146	-0,45493	-0,50801	-0,80232
...	...	...	...
-0,97146	-0,75238	-0,50046	-0,62163
-1,02547	-0,77863	-0,50801	-0,70194
-0,97746	-0,75238	-0,50046	-0,6377
0,330749	-0,57741	-0,46273	1,377922
-0,88145	-0,55116	-0,35706	-0,79429

Setelah mendapatkan StandardScaler pada proses normalisasi pada Tabel 4, langkah penting yang harus dilakukan sebelum proses algoritma K-Means dalam menentukan jumlah *cluster* adalah menggunakan

metode elbow method. *Elbow Method* menentukan jumlah kluster optimal (K) dengan menganalisis perubahan nilai *Within-Cluster Sum of Squares* (WCSS) seiring peningkatan jumlah kluster. WCSS adalah jumlah total jarak kuadrat tiap titik data dengan *centroid* terdekat [20]. Metode elbow digunakan untuk menentukan jumlah *k-cluster* terbaik pada proses *clustering* melalui perhitungan SSE.[21] Cara kerja metode Elbow adalah dengan menghitung nilai WCSS untuk berbagai variasi jumlah kluster. Nilai WCSS yang semakin kecil mengindikasikan bahwa titik data di dalam kluster semakin dekat dengan *centroid* kluster, yang berarti kualitas *clustering* semakin baik. Proses elbow method pada Gambar 3 yang menunjukkan titik optimal. Hasil grafik menunjukkan bahwa titik elbow berada pada nilai $k = 3$, yang ditandai dengan mulai melandainya penurunan nilai WCSS setelah *cluster* ke-3.

3.3. Proses Algoritma K-Means Clustering

Setelah ditentukan jumlah kluster menggunakan metode Elbow Method, maka selanjutnya dilakukan proses algoritma *k-means clustering* dalam mengelompokkan menu makanan berdasarkan Angka Kandungan Gizi (AKG). Pengelompokan berdasarkan atribut yang telah ditentukan pada *preprocessing* data yaitu: energi, protein, lemak dan karbohidrat. Hasil dari algoritma K-Means terdapat 3 *cluster* dari jumlah data 1185 yang telah diproses, yang terlihat pada Tabel 5.



Gambar 3. Grafik *Elbow Method* dalam menentukan jumlah kluster

Tabel 5. Jumlah data yang telah dicluster

Kategori_Gizi	Jumlah_Data
Gizi Sedang	757
Gizi Tinggi	259
Gizi Rendah	169

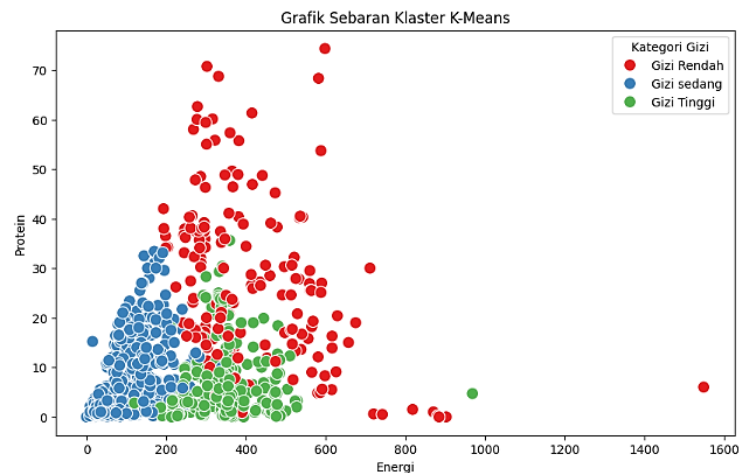
Dari Tabel 5 terlihat jumlah *cluster* yang dihasilkan tiga kluster dengan jumlah data yaitu, kluster 0 = 169 gizi Rendah, kluster 1 = 757 gizi sedang dan kluster 2 = 259 gizi tinggi. Setiap kluster memiliki pola yang berbeda terhadap atribut yang telah ditentukan yaitu : Energi, Protein, Lemak dan Karbohidrat. Berdasarkan hasil pengelompokan yang diperoleh maka dapat dilihat pada Tabel 6 yang menyajikan menu makanan pada setiap kluster.

Tabel 6. Hasil *cluster* menu makanan

Nama_Bahan	Energi	Protein	Lemak	Karbohidrat	Cluster	Kategori_Gizi
Abon Ayam, hypermart	711	30	19	0	0	Gizi Rendah
Abon ikan	435	27,2	20,2	36,1	0	Gizi Rendah
Agar-Agar Plain	10	0	0	3	1	Gizi Sedang
Akar tonjong, segar	45	1,1	0,4	10,8	1	Gizi Sedang
...	...	...	...	...	...	...
Wortel, kukus	37	1	0,6	8,3	1	Gizi Sedang
Wortel, rebus	28	0,7	0,5	6,3	1	Gizi Sedang
Wortel, segar	36	1	0,6	7,9	1	Gizi Sedang
Yangko	254	3	1,1	58,1	2	Gizi Tinggi
Yoghurt, segar	52	3,3	2,5	4	1	Gizi Sedang

Dari Tabel 6 maka didapatkan grafik sebaran kluster K-Means pada Gambar 4. Pada gambar terlihat hasil visualisasi berhasil membagi kedalam beberapa kelompok berdasarkan kemiripan kandungan gizinya.

Grafik sebaran klaster dapat menunjukkan bahwa metode K-Means dapat mengelompokkan data makanan secara efektif berdasarkan kandungan gizi.



Gambar 4. Grafik Sebaran Klaster K-Means

3.4. Integrasi Data

Integrasi pada penelitian ini diperlukan karena menggunakan dua sumber data yang berbeda yaitu : Data angka Kecukupan Gizi (AKG) dan data komposisi makanan. Sehingga data dapat digunakan bersama dalam sistem rekomendasi menu MBG harian. Dengan adanya proses integrasi data maka dapat menghubungkan data makanan dengan target AKG siswa dan membentuk rekomendasi menu MBG berdasarkan kelompok usia. Hasil integrasi digunakan sebagai dasar dalam penyusunan sistem rekomendasi menu makanan bergizi gratis. Tabel 7 menampilkan Integrasi Data.

Tabel 7. Integrasi Data

Kelompok	Energi	Protein
TK/PAUD/LB	350	6
SD Kelas 1-3	413	10
SD Kelas 4-6	683	18
SMP	779	23
SMA	831	24

3.5. Klasifikasi Random Forest

Setelah terbentuknya *cluster* pada proses K-Means selanjutnya algoritma Random Forest di implementasikan untuk merekomendasikan menu makanan berdasarkan kelompok usia penerima siswa disekolah. Sehingga hasil *cluster* dapat digunakan sebagai dasar dalam penyusunan rekomendasi menu makanan harian sesuai kebutuhan Angka Kecukupan Gizi (AKG) kelompok sasaran. Variabel yang digunakan sebagai fitur dalam proses klasifikasi meliputi energi, protein, lemak, dan karbohidrat, sedangkan variabel target menggunakan kategori gizi hasil *clustering* K-Means.

Langkah selanjutnya dataset dibagi menjadi data latih sebagai *training data*(*X-Train*) dan data uji (*X-Test*) menggunakan metode train-test split. Data latih digunakan untuk membangun model Random Forest, sedangkan data uji digunakan untuk mengevaluasi performa model. Dari *training* model random forest *n-estimator*=100 yang artinya jumlah decision tree sebanyak 100. Setelah dapat *training* model maka selanjutnya mengevaluasi model untuk menentukan akurasi dari model random fores dan didapatkan nilai akurasi dari algoritma Random Forest 0.98 atau 98%. Model Random Forest Regressor dievaluasi menggunakan beberapa metrik utama untuk menilai kinerjanya. Tujuannya adalah menarik kesimpulan dari hasil penerapan algoritma untuk memahami kinerja, keberhasilan, serta mengidentifikasi pola-pola atau karakteristik yang muncul selama proses learning[22]

Setelah model random forest berhasil dilatih proses berikutnya adalah melakukan prediksi kategori gizi terhadap data makanan. Hasil klasifikasi tersebut kemudian digunakan dalam proses penyusunan menu makanan harian berdasarkan kebutuhan Angka Kecukupan Gizi (AKG) setiap kelompok sasaran, seperti TK/PAUD, SD, SMP, dan SMA. Sistem rekomendasi menu harian memilih kombinasi makanan yang terdiri dari sumber karbohidrat, protein, sayur, dan buah dengan mempertimbangkan total energi dan protein agar mendekati nilai kebutuhan AKG yang telah ditentukan yang terdapat pada Tabel 8.

Tabel 8. Rekomendasi Menu Harian

Kelompok Siswa	Karbohidrat	Protein	Sayur	Buah	Total Energi	Total Protein
TK/PAUD/LB	Roti	Ikan patin segar	Labu Siam	Apel segar	531	22,8
SD Kelas 1-3	Kentang	Pempek telur	Kangkung, segar	Pisang segar	408	13,1
SD Kelas 4-6	Nasi	Ikan segar	Kangkung, segar	Pisang raja segar	366	26,6
SMP	Mie Sagu	Ayam	Wortel	Jeruk segar	621	35
SMA/SMK	Mie	Ikan Kembung	Brokoli	Apel malang segar	292	15,5

3.6. Pembahasan

Dari hasil penelitian diatas algoritma K-Means dan Random Forest dapat digunakan dalam menentukan rekomendasi menu makanan bergizi pada program Makan Bergizi Gratis (MBG). Proses penelitian dimulai dengan pengumpulan data Angka Kecukupan Gizi (AKG) berdasarkan juknis Badan Gizi Nasional serta data komposisi pangan yang diperoleh dari dapur MBG. Data yang telah didapat dilakukan proses *preprocessing* untuk memastikan kualitas data sebelum dilakukan proses algoritma K-Means dan Random Forest.

Pada tahap *preprocessing*, atribut yang dipilih sebagai data selection untuk kebutuhan pemrosesan data adalah energi, protein, lemak, dan karbohidrat. Pemilihan atribut tersebut dilakukan karena setiap hari makanan yang diberikan kepada peserta didik harus diberi label dengan keterangan jumlah energi, protein, lemak, dan karbohidrat agar diketahui oleh orang tua dan guru. Dan juga merupakan komponen utama dalam menentukan kualitas gizi suatu makanan serta digunakan sebagai dasar dalam penyusunan menu harian siswa. Dan dilanjutkan dengan proses *cleaning data* agar data siap digunakan pada proses K-Means *clustering*.

Sebelum dilakukan penentuan jumlah *cluster* maka dilakukan penentuan jumlah *cluster* menggunakan metode elbow method agar dapat menentukan jumlah klaster terbaik. Berdasarkan grafik elbow method, diperoleh titik optimal pada nilai $k = 3$, yang menunjukkan bahwa jumlah cluster terbaik pada penelitian ini adalah 3 *cluster*.

Dari total data 1185, hasil *cluster* dengan algoritma K-Means dapat membagi data menjadi 3 *cluster* yaitu : kategori gizi rendah sebanyak 169, kategori gizi sedang 757, dan kategori gizi tinggi 259. Hasil tersebut menunjukkan bahwa sebagian besar data makanan berada pada kategori gizi sedang. Perbedaan antarcluster menunjukkan bahwa algoritma K-Means berhasil mengelompokkan makanan berdasarkan kemiripan kandungan energi, protein, lemak, dan karbohidrat. Dan juga visualisasi grafik sebaran *cluster* menunjukkan bahwa data makanan berhasil terbagi ke dalam beberapa kelompok yang memiliki karakteristik gizi berbeda.

Setelah didapatkan *cluster* dari kategori gizi, maka diterapkan algoritma Random Forest sebagai kalsifikasi kategori gizi makanan berdasarkan variabel yang ditentukan. Dataset dibagi menjadi x-train dan x-test, sehingga model Random Forest yang dibangun menggunakan parameter n-estimator sebanyak 100 pohon keputusan. Hasil pengujian menunjukkan bahwa algoritma Random Forest memperoleh nilai akurasi sebesar 98%, yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dalam memprediksi kategori gizi makanan berdasarkan hasil *clustering* K-Means.

Hasil klasifikasi kemudian digunakan dalam sistem rekomendasi menu harian. Sistem berhasil menghasilkan rekomendasi menu makanan yang terdiri dari sumber karbohidrat, protein, sayur, dan buah sesuai kebutuhan energi dan protein setiap kelompok sasaran, berdasarkan data Angka Kecukupan Gizi (AKG) dan juknis dari Badan Gizi Nasional.

Secara keseluruhan, kombinasi metode K-Means dan Random Forest terbukti efektif dalam membangun sistem rekomendasi menu makanan berbasis data. Algoritma K-Means dan Random Forest telah banyak diterapkan dalam berbagai penelitian sebelumnya, salah satunya pada penelitian yang dilakukan oleh Sarah Astuti dkk [23] yang mampu mengintegrasikan pengelompokan K-Means dan klasifikasi Random Forest efektif untuk memprediksi potensi daerah penangkapan ikan menggunakan data cuaca dan oseanografi di perairan Batam dengan akurasi 89,2%, presisi 87,5%. Hal ini sejalan dengan penelitian [24] dan [25] yang menunjukkan efektivitas algoritma K-Means *Clustering* dan Random Forest dalam mengelompokkan objek dengan hasil akurasi 95%, walaupun dengan objek penelitian yang berbeda.

4. KESIMPULAN

Setelah melakukan penelitian, kita dapat menyimpulkan bahwa menggabungkan algoritma K-Means dan Random Forest sangat efektif dalam membangun sistem rekomendasi menu makanan bergizi untuk program Makan Bergizi Gratis. Proses pengolahan data yang mencakup pemilihan data, pembersihan data,

transformasi, dan normalisasi berhasil menghasilkan dataset yang siap digunakan untuk proses pengelompokan dan klasifikasi.

Dengan menggunakan metode Elbow Method, kita menemukan bahwa jumlah kelompok yang optimal dalam penelitian ini adalah 3 kelompok. Kemudian, algoritma K-Means berhasil mengelompokkan 1185 data makanan menjadi tiga kategori gizi, yaitu gizi rendah sebanyak 169 data, gizi sedang sebanyak 757 data, dan gizi tinggi sebanyak 259 data berdasarkan variabel energi, protein, lemak, dan karbohidrat. Grafik sebaran kelompok menunjukkan bahwa metode K-Means dapat melakukan segmentasi makanan secara efektif berdasarkan kemiripan kandungan gizinya.

Kita juga berhasil mengintegrasikan data Angka Kecukupan Gizi dan data komposisi pangan untuk menyusun sistem rekomendasi menu makanan harian sesuai dengan kelompok sasaran siswa, seperti TK/PAUD, SD, SMP, dan SMA. Selanjutnya, algoritma Random Forest digunakan untuk melakukan klasifikasi kategori gizi makanan berdasarkan hasil pengelompokan K-Means dan menghasilkan nilai akurasi sebesar 98%, yang menunjukkan bahwa model memiliki performa klasifikasi yang sangat baik.

Hasil akhir penelitian menunjukkan bahwa sistem dapat menghasilkan rekomendasi menu makanan harian yang terdiri dari sumber karbohidrat, protein, sayur, dan buah sesuai dengan kebutuhan energi dan protein masing-masing kelompok siswa. Dengan demikian, penelitian ini membuktikan bahwa menggabungkan metode K-Means dan Random Forest dapat mendukung proses pengambilan keputusan dalam menyusun menu makanan bergizi yang lebih tepat, terstruktur, dan sesuai dengan kebutuhan gizi peserta didik pada program Makan Bergizi Gratis.

REFERENSI

- [1] Merlinda Ajeng and Yusuf Yusmar, “Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjau dari Perspektif Sosiologi Pendidikan,” *Ranah Res. J. Multidiscip. Res. Dev.*, vol. 7, no. 2, pp. 1364–1373, 2025.
- [2] Nurul Azzahra, Akbar Dwi Dharmawan, Aras Fikar Mardatilah, Mohamad Ilham Habibi, Virliana Amanda Putri Aryani, and Suhardi, “Pelaksanaan Uji Coba Program Makan Bergizi Gratis di SMP Negeri 4 Tangerang,” *J. Pengabd. Masy. dan Ris. Pendidik.*, vol. 3, no. 4, pp. 5036–5044, 2025, doi: 10.31004/jerkin.v3i4.1335.
- [3] D. A. M. C. Dhirya and I. G. K. Warmika, “Efektivitas Program Makan Gratis Pada Status Gizi Siswa Sekolah Dasar: Tinjauan Sistematis,” *MANUJU MALAHAYATI Nurs. J.*, vol. 7, no. 1, pp. 27–48, 2025.
- [4] L. Wikarsa, L. Sitanayah, V. E. Manginsela, and M. Kim, “Hybrid Bert-Knn for Emotion Detection in Social Media Responses To Indonesia’S Free Meal Program: Performance and Challenge,” *ICIC Express Lett. Part B Appl.*, vol. 17, no. 5, pp. 525–532, 2026, doi: 10.24507/icicelb.17.05.525.
- [5] D. D. L. H. Dida Akhmad Gurnida, Nanan Nur’aeny and I. R. Fransisca Sri Susilaningih, Dewi Marhaeni Diah Herawati, “Korelasi antara tingkat kecukupan gizi dengan indeks massa tubuh siswa sekolah dasar kelas 4, 5 dan 6,” *Padjadjaran J. Dent. Res. Student*, vol. 4, no. April, pp. 43–50, 2020, doi: 10.24198/pjdrs.v3i2.25763.
- [6] N. Hendrastuty, “Penerapan Data Mining Menggunakan Algoritma K-Means *Clustering* Dalam Evaluasi Hasil Pembelajaran Siswa,” *J. Ilm. Inform. dan Ilmu Komput.*, vol. 3, no. 1, pp. 46–56, 2024, doi: 10.58602/jima-ilkom.v3i1.26.
- [7] N. Sayuti Hanapiah, N. Suarna, and W. Prihartono, “Meningkatkan Penanganan Stunting Pada Anak Melalui Klasifikasi Pemberian Makanan Tambahan Berdasarkan Usia Dengan Metode K-Means Di Desa Cintarasa,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 6, pp. 3484–3493, 2023, doi: 10.36040/jati.v7i6.8213.
- [8] F. Rahman, T. A. Siswa, and R. Rudiman, “Penerapan Metode GA-RU Pada Algoritma Random Forest Untuk Mengatasi Class Imbalance Data Beasiswa KIP-Kuliah,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 4, pp. 2345–2357, 2025, doi: 10.47065/bits.v6i4.6757.
- [9] T. Nugroho, U. Enri, and I. Maulana, “*Clustering* Menu Makanan Berdasarkan Kebutuhan Kalori Harian Menggunakan Algoritma K-Means,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 4, pp. 4411–4417, 2024, doi: 10.36040/jati.v8i4.9958.
- [10] N. Fitroha, A. E. Suandi, and M. F. Nurcahyadi, “Perancangan Sistem Rekomendasi Resep Bergizi Untuk Masyarakat Menggunakan K-Means *Clustering*,” *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 10, no. 1, pp. 184–189, 2026.
- [11] F. I. Rahman, L. Lukman, and H. Hildayati, “Sistem Klasifikasi Kerusakan Jalan Metode *Machine Learning* dengan Algoritma K-Means dan Random Forest,” *Arus J. Sains dan Teknol.*, vol. 3, no. 1, pp. 116–126, 2025, doi: 10.57250/ajst.v3i1.1212.
- [12] S. Sulistiyani, “Student Characteristics, Acceptability, and Suitability of Portion Standards with Recommended Dietary Allowance in School Meals at Al Furqan Primary School, Jember Regency,” *Amerta Nutr.*, vol. 8, no. 3SP, pp. 285–294, 2024, doi: 10.20473/amnt.v8i3SP.2024.285-294.
- [13] Y. Elda, S. Defit, Y. Yunus, and R. Syaljumairi, “Klasterisasi Penempatan Siswa yang Optimal untuk

- Meningkatkan Nilai Rata-Rata Kelas Menggunakan K-Means,” *J. Inf. dan Teknol.*, vol. 3, pp. 103–108, 2021, doi: 10.37034/jidt.v3i3.130.
- [14] N. N. Saragih and R. Kurniawan, “(Journal of Computer Engineering, System and Science),” *J. Comput. Eng. Syst. Sci.*, vol. 10, no. 1, pp. 299–311, 2025.
- [15] A.- Akramunnisa and F. Fajriani, “K-Means *Clustering* Analysis pada PersebaranTingkat Pengangguran Kabupaten/Kota di Sulawesi Selatan,” *J. Varian*, vol. 3, no. 2, pp. 103–112, 2020, doi: 10.30812/varian.v3i2.652.
- [16] Putri Azhiah, Jasmir, and Purnama Benni, “Klasifikasi Kelayakan Penerima Beasiswa Menggunakan Naive Bayes dengan Optimasi Atribut Berbasis K-Means,” *Inst. Ris. dan Publ. Indones.*, vol. 5, no. October, pp. 1450–1462, 2025.
- [17] H. Tantyoko, D. K. Sari, and A. R. Wijaya, “Prediksi Potensial Gempa Bumi Indonesia Menggunakan Metode Random Forest Dan Feature Selection,” *IDEALIS Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 83–89, 2023, doi: 10.36080/ideal.v6i2.3036.
- [18] A. Fauzi, R. Supriyadi, and N. Maulidah, “Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest,” *J. Infortech*, vol. 2, no. 1, pp. 96–101, 2020, doi: 10.51717/simkom.v10i1.703.
- [19] K. K. B. G. NASIONAL, R. INDONESIA, and N. 401. . T. 2025, “Kepala Badan Gizi Nasional Republik Indonesia.” pp. 1–282, 2025.
- [20] G. Ngurah, D. Kajeng, I. M. A. Wirawan, I. K. Dwi, A. Putra, and R. N. Putera, “Integrasi Algoritma K-Means untuk *Clustering* Data Kuesioner Kepribadian Mahasiswa,” *Kumpul. Artik. Mhs. Pendidik. Tek. Inform.*, vol. 15, no. Volume 15, Nomor 2, Tahun 2026, pp. 247–258, 2026.
- [21] N. A. Maori and E. Evanita, “Metode Elbow dalam Optimasi Jumlah *Cluster* pada K-Means *Clustering*,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.
- [22] Suci Amaliah, M. Nusrang, and A. Aswi, “Penerapan Metode Random Forest Untuk Klasifikasi Varian Minuman Kopi di Kedai Kopi Konijiwa Bantaeng,” *VARIANSI J. Stat. Its Appl. Teach. Res.*, vol. 4, no. 3, pp. 121–127, 2022, doi: 10.35580/variensiunm31.
- [23] S. Astiti and A. W. Aranski, “Prediction of Potential Fishing Zones Using K- Means *Clustering* and Random Forest in Batam Waters,” *JUITA J. Inform.*, vol. 14, no. 1, pp. 221–230, 2026.
- [24] Fadil Danu Rahman, M. I. Z. Mulki, and A. Taryana, “*Clustering* Dan Klasifikasi Data Cuaca Cilacap Dengan Menggunakan Metode K-Means Dan Random Forest,” *J. SINTA Sist. Inf. dan Teknol. Komputasi*, vol. 1, no. 2, pp. 90–97, 2024, doi: 10.61124/sinta.v1i2.15.
- [25] D. R. Forest, W. H. Sasoko, E. W. Pujiarto, R. Haris, and A. Y. Kania, “R. Haris, W. Haryo, E. W. Pujiarto, A. Yuza, K. Kusriani, dan K. Kusnawi, ‘Prediksi banjir di DKI Jakarta dengan menggunakan algoritma K-Means dan random forest,’ Jurnal Informatika dan Teknologi Komputer (J-ICOM), vol. 5, no. 1, pp. 43–49, 2024.,” vol. 05, no. 01, pp. 43–49, 2024, [Online]. Available: <https://ejurnalunsam.id/index.php/jicom/>