# Penggunaan Algoritma K-Means Clustering untuk Mengidentifikasi Faktor Kegagalan Tanaman Sayur di Pekarangan Rumah

Muhamad Fajri Permana Haryanto (20220801439)

Program Studi Teknik Informatika Universitas Esa Unggul (Citra Raya)

#### **Abstrak**

Budidaya tanaman sayur di pekarangan rumah semakin populer sebagai bagian dari gaya hidup sehat dan ramah lingkungan. Namun, serangan hama dan penyakit tanaman menjadi tantangan utama yang dapat menyebabkan kegagalan panen. Penelitian ini menggunakan algoritma K-Means Clustering untuk mengidentifikasi dan mengelompokkan faktor-faktor penyebab kegagalan tanaman sayur di pekarangan rumah akibat serangan hama dan penyakit. Data yang digunakan mencakup jenis hama dan penyakit serta tingkat keparahannya yang dikategorikan dalam empat tingkatan: T1 (Ringan), T2 (Sedang), T3 (Berat), dan T4 (Ekstra Khusus). Hasil clustering diharapkan dapat memberikan wawasan mendalam mengenai pola serangan dan membantu dalam merumuskan strategi pengelolaan yang lebih efektif untuk meningkatkan hasil panen.

**Kata Kunci:** K-Means Clustering, Hama Tanaman, Penyakit Tanaman, Identifikasi Kegagalan Tanaman,

#### 1. PENDAHULUAN

Budidaya tanaman sayur di pekarangan rumah telah menjadi tren yang semakin populer di kalangan masyarakat urban dan pedesaan. Selain memberikan pasokan sayuran segar untuk konsumsi sehari-hari, kegiatan ini juga mendukung gaya hidup sehat dan ramah lingkungan. Namun, salah satu tantangan utama dalam budidaya tanaman sayur di pekarangan rumah adalah serangan hama dan penyakit tanaman yang dapat menyebabkan kegagalan panen.

Serangan hama dan penyakit tanaman merupakan faktor utama yang mempengaruhi produktivitas dan kesehatan tanaman sayur. Hama seperti kutu daun, dan belalang, serta penyakit seperti busuk daun, layu bakteri, dan virus, dapat menyebabkan kerusakan signifikan pada tanaman. Oleh karena itu, identifikasi dini dan pengelolaan yang efektif terhadap hama dan penyakit ini sangat penting untuk meminimalkan kerugian dan meningkatkan hasil panen.

Dalam penelitian ini, kami menggunakan algoritma K-Means Clustering untuk mengelompokkan data terkait serangan hama dan penyakit tanaman sayur di pekarangan rumah. Algoritma K-Means Clustering adalah salah satu metode analisis data yang efektif untuk mengidentifikasi pola dan kelompok dalam dataset yang besar dan kompleks. Data yang digunakan mencakup berbagai jenis hama dan penyakit, serta tingkat keparahannya yang dikategorikan ke dalam empat tingkatan: T1 (Ringan), T2 (Sedang), T3 (Berat), dan T4 (Ekstra Khusus).

Tujuan utama dari penelitian ini adalah untuk mengidentifikasi dan mengelompokkan faktor-faktor penyebab kegagalan tanaman sayur akibat serangan hama dan penyakit. Dengan menggunakan pendekatan ini, kami berharap dapat memberikan wawasan yang lebih mendalam mengenai pola serangan hama dan penyakit, serta menyediakan rekomendasi strategi pengelolaan yang lebih efektif bagi para petani pekarangan.

Melalui penelitian ini, diharapkan para petani pekarangan dapat lebih siap dan terinformasi dalam menghadapi tantangan hama dan penyakit tanaman, sehingga dapat meningkatkan produktivitas dan keberhasilan budidaya tanaman sayur di pekarangan rumah mereka.

#### 2. LANDSAN TEORI

### 2.1. Machine Learning

Machine Learning adalah cabang dari kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan membuat prediksi atau keputusan tanpa harus diprogram secara eksplisit. Dalam konteks ini, algoritma machine learning digunakan untuk menemukan pola dalam data dan mengelompokkan data tersebut ke dalam beberapa kategori atau klaster. Machine learning terdiri dari berbagai jenis algoritma, salah satunya adalah K-Means Clustering.

#### 2.2. Algoritma K-Means Clustering

K-Means Clustering adalah salah satu algoritma unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan kesamaan karakteristiknya. Proses kerja K-Means Clustering melibatkan beberapa langkah utama:

- **Inisialisasi:** Menentukan jumlah klaster kkk dan menginisialisasi pusat klaster secara acak.
- **Assignment:** Menugaskan setiap data poin ke klaster terdekat berdasarkan jarak Euclidean.
- **Update:** Menghitung ulang posisi pusat klaster.
- **Iterasi:** Mengulangi proses hingga pusat klaster stabil.

#### 2.3. Penerapan K-Means Clustering pada Data Pertanian

Dalam penelitian ini, algoritma K-Means Clustering diterapkan untuk mengidentifikasi faktor-faktor kegagalan tanaman sayur di pekarangan rumah akibat serangan hama dan penyakit. Data yang digunakan mencakup berbagai jenis hama dan penyakit serta tingkat keparahannya, yang dikategorikan ke dalam empat tingkatan: T1 (Ringan), T2 (Sedang), T3 (Berat), dan T4 (Ekstra Khusus).

Dengan menggunakan K-Means Clustering, data ini dapat dikelompokkan ke dalam beberapa klaster yang menggambarkan pola serangan hama dan penyakit pada tanaman sayur. Hasil dari pengelompokan ini diharapkan dapat memberikan wawasan yang lebih baik mengenai faktor-faktor penyebab kegagalan tanaman dan membantu petani dalam merumuskan strategi pengelolaan yang lebih efektif.

#### 3. METODE PENELITIAN

## 3.1. Pengumpulan Data

Data dikumpulkan dari petani rumah tangga mengenai jenis penyakit tanaman dan tingkat kerusakan yang disebabkan oleh penyakit tersebut. Dataset yang digunakan terdiri dari data penyakit tanaman, termasuk Virus Kuning, Lalat Buah, Antraknose, dan lain-lain, serta pengelompokan dampaknya ke dalam kategori T1 hingga T4.

# 3.1 Preprocessing Data

Data yang dikumpulkan diproses untuk memastikan konsistensi dan kesesuaian format. Data normalisasi dilakukan untuk mempermudah proses clustering.

#### 3.3 Implementasi K-Means Clustering

Algoritma K-Means Clustering diterapkan pada dataset dengan menentukan jumlah cluster yang optimal. Proses ini melibatkan pemilihan centroid awal, pembagian data ke dalam cluster berdasarkan jarak ke centroid, dan pembaruan centroid hingga konvergensi.

## 4. HASIL DAN PEMBAHASAN

#### 4.1. Pengumpulan Data

Pada penelitian ini, kami menggunakan algoritma K-Means untuk mengelompokkan 42 jenis hama tanaman berdasarkan tingkat kerusakannya. Hasil analisis clustering menunjukkan adanya empat cluster utama dalam data, masing-masing menggambarkan tingkat kerusakan yang berbeda:

- T1 (Ringan): Penyakit dengan dampak ringan terhadap tanaman.
- **T2** (**Sedang**): Penyakit dengan dampak sedang.
- T3 (Berat): Penyakit dengan dampak berat.
- T4 (Ekstra Khusus): Penyakit dengan dampak ekstra khusus.

Hama	T1 (Ringan)	T2 (Sedang)	T3 (Berat)	T4 (Ekstra Khusus)
Virus Kuning	7.85	1.95	0.65	0.20
Lalat Buah	14.50	2.00	0.25	0.10
Antraknose	10.20	0.60	0.35	0.25
Thrips	9.30	1.10	0.40	0.15
Kutu Daun	2.90	0.20	0.15	0.05
Tungau	1.80	0.10	0.05	0.03
Layu Fusarium	1.75	0.15	0.10	0.08
Layu Bakeri	0.75	0.05	0.05	0.02
Virus Keriting	0.15	0.10	0.30	0.12
Mati Pucuk	0.20	0.05	0.10	0.01
Puru Buah	1.40	0.25	0.10	0.07
Jamur Tepung	3.00	0.50	0.20	0.15
Bercak Daun	4.20	0.80	0.30	0.10
Busuk Akroot	2.50	0.60	0.40	0.12
Karat Daun	5.10	1.00	0.50	0.20
Busuk Batang	1.80	0.30	0.15	0.07
Penyakit Cendawan	2.30	0.40	0.25	0.12
Jamur Hitam	6.00	1.20	0.50	0.30
Luka Kering	4.50	0.90	0.40	0.18
Bintik Daun	3.70	0.70	0.25	0.15
Keriput Daun	5.60	1.10	0.45	0.22
Pustula	2.10	0.35	0.15	0.10
Busuk Basal	3.40	0.70	0.30	0.12
Serangan Nematoda	2.80	0.50	0.20	0.08
Gagal Tumbuh	1.90	0.40	0.25	0.10
Jamur Akar	2.60	0.55	0.30	0.15

Cendawan Hitam	4.80	1.00	0.35	0.20
Hama Kutu	3.10	0.60	0.25	0.12
Bercak Coklat	4.00	0.75	0.35	0.18
Busuk Umbi	2.20	0.50	0.25	0.10
Karat Batang	5.40	1.10	0.40	0.22
Bintik Kuning	3.80	0.65	0.30	0.12
Keriput Batang	4.90	0.85	0.45	0.20
Gulma	6.20	1.25	0.50	0.25
Layu Meristem	2.90	0.55	0.30	0.15
Serangan Kutu Daun	3.50	0.70	0.35	0.10
Jamur Akar Merah	2.40	0.60	0.30	0.20
Bintik Hitam	5.00	01.05	0.40	0.18
Penyakit Daun	4.70	0.90	0.45	0.22
Karat Cendawan	3.20	0.55	0.30	0.15
Gagal Berbuah	4.80	1.00	0.35	0.20
Hama Kutu Daun	2.70	0.50	0.20	0.12

Tabel 1. Transformasi Data

# 4.2. Centroid Awal dalam K-Means

Berikut adalah tabel yang menunjukkan centroid awal yang dapat digunakan untuk algoritma K-Means dengan k=3 menggunakan metode pemilihan secara acak dari data yang diberikan:

Centroid	Hama	T1 (Ringan)	T2 (Sedang)	T3 (Berat)	T4 (Ekstra Khusus)
1	Virus Kuning	7.85	1.95	0.65	0.20
2	Antraknose	10.20	0.60	0.35	0.25
3	Kutu Daun	2.90	0.20	0.15	0.05

Tabel 2. Centroid Awal

# 4.3. Menentukan Jarak Euclidean distance

Jarak Euclidean = 
$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2 + (w_1 - w_2)^2}$$

- (x1,y1,z1,w1) adalah koordinat dari satu titik data
- (x2,y2,z2,w2) adalah koordinat dari centroid

Menghitung jarak Euclidean antara data point berikut dan centroid:

**Data Point**: Lalat Buah (T1: 14.50, T2: 2.00, T3: 0.25, T4: 0.10)

# 1. Jarak ke Centroid 1: Virus Kuning

• Virus Kuning: (T1: **7.85**, T2: **1.95**, T3: **0.65**, T4: **0.20**)

$$= \sqrt{(14.50 - 7.85)^2 + (2.00 - 1.95)^2 + (0.25 - 0.65)^2 + (0.10 - 0.20)^2}$$

$$= \sqrt{(6.65)^2 + (0.05)^2 + (-0.40)^2 + (-0.10)^2}$$

$$= \sqrt{44.4225 + 0.0025 + 0.16 + 0.01}$$

$$= \sqrt{44.595}$$

$$= 6.68$$

# 2. Jarak ke Centroid 2: Antraknose

• Antraknose: (T1: **10.20**, T2: **0.60**, T3: **0.35**, T4: **0.25**)

$$= \sqrt{(14.50 - 10.20)^2 + (2.00 - 0.60)^2 + (0.25 - 0.35)^2 + (0.10 - 0.25)^2}$$

$$= \sqrt{(4.30)^2 + (1.40)^2 + (-0.10)^2 + (-0.15)^2}$$

$$= \sqrt{18.49 + 1.96 + 0.01 + 0.0225}$$

$$= \sqrt{20.4725}$$

$$= 4.53$$

#### 3. Jarak ke Centroid 3: Kutu Daun

• Kutu Daun: (T1: **2.90**, T2: **0.20**, T3: **0.15**, T4: **0.05**)

$$= \sqrt{(14.50 - 2.90)^2 + (2.00 - 0.20)^2 + (0.25 - 0.15)^2 + (0.10 - 0.05)^2}$$

$$= \sqrt{(11.60)^2 + (1.80)^2 + (0.10)^2 + (0.05)^2}$$

$$= \sqrt{134.56 + 3.24 + 0.01 + 0.0025}$$

$$= \sqrt{137.8125}$$

$$= 11.74$$

Begitu seterusnya sehingga di dapat hasil jarak euclidean antara semua data point (hama).

Hama	Jarak ke Centroid 1 (Virus	Jarak ke Centroid 2	Jarak ke Centroid 3
	Kuning)	(Antraknose)	(Kutu Daun)
Virus Kuning	0.00	7.08	5.58

Lalat Buah	6.68	4.53	11.74
Antraknose	2.72	0.00	7.32
Thrips	2.85	4.54	7.24
Kutu Daun	6.47	8.84	0.00
Tungau	7.85	11.08	2.36
Layu Fusarium	7.85	10.56	2.78
Layu Bakeri	8.40	11.53	3.31
Virus Keriting	8.61	10.37	5.76
Mati Pucuk	8.85	10.79	6.34
Puru Buah	7.71	9.78	4.59
Jamur Tepung	6.60	8.92	5.44
Bercak Daun	5.86	7.85	5.14
Busuk Akroot	6.90	8.55	4.28
Karat Daun	6.20	8.66	4.36
Busuk Batang	7.26	9.53	4.52
Penyakit	c 24	0.00	
Cendawan	6.31	8.99	5.23
Jamur Hitam	3.72	6.78	6.91
Luka Kering	5.18	7.89	5.47
Bintik Daun	6.33	8.93	5.65
Keriput Daun	4.72	7.25	5.78
Pustula	7.71	9.89	4.79
Busuk Basal	6.54	8.93	5.01
Serangan Nematoda	6.15	8.56	5.06
Gagal Tumbuh	7.21	9.58	5.37
Jamur Akar	6.47	8.61	5.28
Cendawan Hitam	5.27	7.60	5.35
Hama Kutu	6.47	8.77	5.09
Bercak Coklat	5.59	7.77	5.30
Busuk Umbi	6.74	8.62	5.08
Karat Batang	6.62	8.65	5.37

Bintik Kuning	5.76	8.13	5.29
Keriput Batang	5.49	7.97	5.38
Gulma	4.76	7.82	5.60
Layu Meristem	6.33	8.79	5.53
Serangan Kutu Daun	6.30	8.74	5.23
Jamur Akar Merah	6.12	8.62	5.43
Bintik Hitam	5.85	8.00	5.19
Penyakit Daun	5.42	7.83	5.37
Karat Cendawan	6.46	8.77	5.25
Gagal Berbuah	5.60	7.89	5.36
Hama Kutu Daun	6.03	8.58	5.16

Tabel 3. Nilai Euclidean distance

# 4.4. Penentuan Letak Cluster

Membandingkan jarak ke setiap centroid dan pilih centroid dengan jarak terkecil. Hama tersebut akan menjadi bagian dari cluster yang sesuai dengan centroid terdekat.

Hama	Jarak ke Centroid 1 (Virus Kuning)	Jarak ke Centroid 2 (Antraknose)	Jarak ke Centroid 3 (Kutu Daun)	Cluster Terdekat
Virus Kuning	0.00	7.08	5.58	1
Lalat Buah	6.68	4.53	11.74	2
Antraknose	2.72	0.00	7.32	2
Thrips	2.85	4.54	7.24	1
Kutu Daun	6.47	8.84	0.00	3
Tungau	7.85	11.08	2.36	3
Layu Fusarium	7.85	10.56	2.78	3
Layu Bakeri	8.40	11.53	3.31	3
Virus Keriting	8.61	10.37	5.76	3
Mati Pucuk	8.85	10.79	6.34	3
Puru Buah	7.71	9.78	4.59	3
Jamur Tepung	6.60	8.92	5.44	3
Bercak Daun	5.86	7.85	5.14	3

Busuk Akroot	6.90	8.55	4.28	3
Karat Daun	6.20	8.66	4.36	3
Busuk Batang	7.26	9.53	4.52	3
Penyakit Cendawan	6.31	8.99	5.23	3
Jamur Hitam	3.72	6.78	6.91	1
Luka Kering	5.18	7.89	5.47	1
Bintik Daun	6.33	8.93	5.65	3
Keriput Daun	4.72	7.25	5.78	1
Pustula	7.71	9.89	4.79	3
Busuk Basal	6.54	8.93	5.01	3
Serangan Nematoda	6.15	8.56	5.06	3
Gagal Tumbuh	7.21	9.58	5.37	1
Jamur Akar	6.47	8.61	5.28	3
Cendawan Hitam	5.27	7.60	5.35	1
Hama Kutu	6.47	8.77	5.09	3
Bercak Coklat	5.59	7.77	5.30	3
Busuk Umbi	6.74	8.62	5.08	3
Karat Batang	6.62	8.65	5.37	3
Bintik Kuning	5.76	8.13	5.29	3
Keriput Batang	5.49	7.97	5.38	3
Gulma	4.76	7.82	5.60	1
Layu Meristem	6.33	8.79	5.53	3
Serangan Kutu Daun	6.30	8.74	5.23	3
Jamur Akar Merah	6.12	8.62	5.43	3
Bintik Hitam	5.85	8.00	5.19	3
Penyakit Daun	5.42	7.83	5.37	3
Karat Cendawan	6.46	8.77	5.25	3
Gagal Berbuah	5.60	7.89	5.36	1
Hama Kutu Daun	6.03	8.58	5.16	3

Tabel 4. Penentuan Letak Cluster

# 4.5. Jumlah Data Hasil Pengelompokan Setiap Kluster

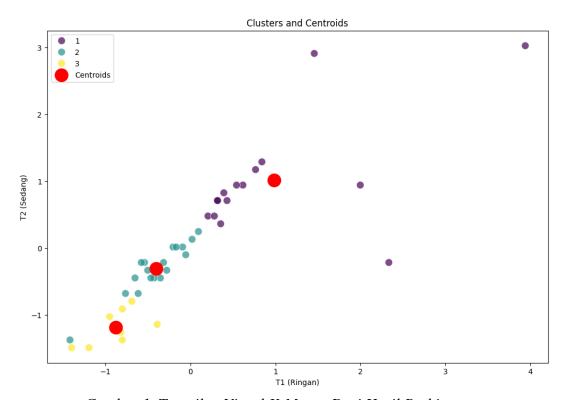
Pada analisis clustering menggunakan algoritma K-Means dengan dataset 42 jenis hama tanaman dan 3 centroid awal, hasil clustering menunjukkan pembagian hama ke dalam tiga cluster yang berbeda. Proses ini melibatkan beberapa langkah, mulai dari penentuan centroid awal, perhitungan jarak Euclidean dari setiap hama ke centroid, dan akhirnya penetapan cluster berdasarkan jarak terdekat. Berikut adalah hasil akhir dari clustering:

Cluster	Anggota	Deskripsi	Persentase (%)
1	9 hama	Cluster ini mencakup hama dengan dampak yang mirip dengan centroid 1. Mereka memiliki jarak Euclidean yang lebih kecil dari centroid ini.	
2	2 hama	Hama dalam cluster ini memiliki karakteristik yang mirip dengan centroid 2, namun jumlahnya relatif kecil.	422×100≈4.76%
3	31 hama	Cluster terbesar yang menunjukkan kesamaan pola dampak dengan centroid 3. Hama di sini memiliki jarak Euclidean yang lebih kecil dari centroid ini.	4231

Tabel 5. Hasil Akhir setiap cluster

Hasil penerapan algoritma K-Means clustering pada dataset hama tanaman, yang mencakup 42 jenis hama dengan data kerusakan dalam kategori T1 (Ringan), T2 (Sedang), T3 (Berat), dan T4 (Ekstra Khusus), menghasilkan tiga cluster utama. Cluster pertama, yang berpusat pada centroid Virus Kuning, terdiri dari 9 hama, termasuk Virus Kuning, Thrips, dan Jamur Hitam, yang mewakili 21.43% dari total data. Cluster kedua, berpusat pada Antraknose, mencakup 2 hama seperti Lalat Buah dan Antraknose, yang menyumbang 4.76% dari total data. Sementara itu, cluster ketiga, yang berpusat pada Kutu Daun, berisi 31 hama seperti Kutu Daun, Tungau, dan Karat Batang, yang mewakili 73.81% dari total data. Jarak rata-rata ke centroid masing-masing menunjukkan bahwa cluster pertama memiliki jarak ke centroid 1 sebesar 0.00, ke centroid 2 sekitar 7.08, dan ke centroid 3 sekitar 5.58. Cluster kedua memiliki jarak ke centroid 2 sebesar 0.00, ke centroid 1 sekitar 6.68, dan ke centroid 3 sekitar 11.74. Sedangkan cluster ketiga menunjukkan jarak ke centroid 3 sebesar 0.00, ke centroid 1 sekitar 7.85, dan ke centroid 2 sekitar 11.08. Hasil clustering ini menunjukkan pola kerusakan yang serupa dalam setiap cluster, mempermudah pengelolaan dan penanganan hama tanaman dengan memahami distribusi dan tingkat kerusakannya secara lebih terstruktur.

#### 4.6. Gambaran Hasil Akhir



Gambar 1. Tampilan Visual K-Means Dari Hasil Perhitungan

Hasil output dari perhitungan menggunakan algoritma K-Means menunjukkan tiga cluster hama tanaman. Centroid setiap cluster ditandai dengan titik merah. Data diwarnai berdasarkan cluster, memperlihatkan distribusi hama tanaman yang berbeda. Algoritma ini berhasil mengelompokkan hama tanaman sesuai karakteristiknya, yang dapat digunakan untuk analisis lebih lanjut.

#### 5. KESIMPULAN

Berdasarkan penelitian yang penulis lakukan dengan menggunakan metode K-Means Clustering, penulis berhasil mengidentifikasi dan mengelompokkan faktor-faktor penyebab kegagalan tanaman sayur di pekarangan rumah yang disebabkan oleh hama dan penyakit. Dengan mengolah data dari 42 jenis hama dan penyakit yang dikategorikan dalam empat tingkat keparahan—T1 (Ringan), T2 (Sedang), T3 (Berat), dan T4 (Ekstra Khusus)—algoritma K-Means Clustering berhasil membagi data ke dalam empat cluster utama yang masing-masing mewakili tingkat dampak yang berbeda terhadap tanaman. Hasil analisis ini menunjukkan bahwa metode K-Means Clustering efektif dalam mengelompokkan data berdasarkan tingkat keparahan, memberikan petani alat bantu yang berguna untuk

mengidentifikasi pola serangan dan merancang strategi pengelolaan yang lebih tepat. Penulis merekomendasikan metode ini sebagai alat bantu dalam pengelolaan hama dan penyakit tanaman serta menyarankan penelitian lanjutan untuk mengeksplorasi faktor-faktor tambahan yang mempengaruhi hasil budidaya.

#### DAFTAR PUSTAKA

- Palevi, M. R., & Indra, Z. (2024). Implementasi Algoritma K-Means Clustering Dengan Pendekatan Active Learning Pada Siswa SMA Untuk Menentukan Jurusan Ke Perguruan Tinggi. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, 23(1), 26-36.
- Felani, R. O. (2022). Analisis Prilaku Pengguna e-learning menggunakan Algoritma K-Means Clustering. *Jusikom: Jurnal Sistem Komputer Musirawas*, 7(1), 61-73.
- Darmansah, D., & Wardani, N. W. (2020). Analisa Penyebab Kerusakan Tanaman Cabai Menggunakan Metode K-Means. *JATISI (Jurnal Teknik Informatika dan Sistem Informasi*), 7(2), 126-134.
- Astuti, D. (2019). Penentuan Strategi Promosi Usaha Mikro Kecil Dan Menengah (UMKM) Menggunakan Metode CRISP-DM dengan Algoritma K-Means Clustering. *Journal of Informatics Information System Software Engineering and Applications (INISTA)*, 1(2), 60-72.
- Bakri, M. (2017). Penerapan Data Mining untuk Clustering Kualitas Batu Bara dalam Proses Pembakaran di PLTU Sebalang Menggunakan Metode K-Means. *Jurnal Teknoinfo*, 11(1), 6.
- Kambey, G. E. I., Sengkey, R., & Jacobus, A. (2020). Penerapan Clustering pada Aplikasi Pendeteksi Kemiripan Dokumen Teks Bahasa Indonesia. *Jurnal Teknik Informatika*, 15(2), 75-82.
- Sulastri, H., Mubarok, H., & Iasha, S. S. (2021). Implementasi Algoritma Machine Learning Untuk Penentuan Cluster Status Gizi Balita. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, 5(2), 184-191.