

# Implementasi Algoritma *K-Nearest Neighbor* untuk Prediksi Kerusakan Truk *Mixer* dengan Pendekatan CRISP-DM

Oleh:

**Meisya Azzahra Rachman**

Dosen Pembimbing:

**Tedjo Sukmono, ST., MT.**

Progam Studi Teknik Industri

Universitas Muhammadiyah Sidoarjo

Juni 2023

# Latar Belakang

- Alat penunjang berupa truk mixer yang digunakan dalam kegiatan pengecoran beton pada PT. Varia Usaha Beton kerap mengalami *breakdown*.
- Target operasional truk *mixer* per bulannya sebesar 90%, berbeda jauh dengan *reability* yang diperoleh di lapangan yang hanya diperoleh sebesar 60%.
- Permasalahan berupa *breakdown* selain berpengaruh pada penurunan produksi, juga berpengaruh pada persediaan suku cadang dan berakibat pada hal-hal lainnya seperti: biaya-biaya yang berhubungan dengan persediaan suku cadang dan lamanya waktu tunggu pengiriman suku cadang di luar dari jadwal.

# Rumusan Masalah dan Tujuan

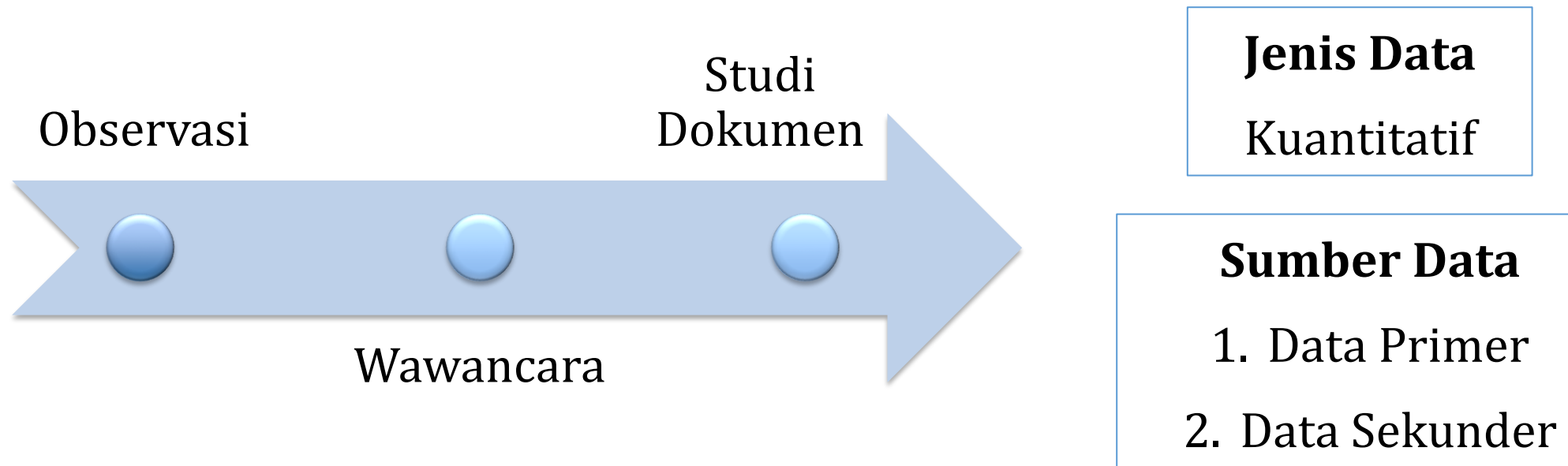
## Rumusan Masalah

Seberapa besar akurasi dan *error* yang dihasilkan dalam deteksi kerusakan truk *mixer* dengan menggunakan algoritma *K-Nearest Neighbor*.

## Tujuan

Untuk mengetahui nilai *k* yang menghasilkan akurasi tertinggi dalam deteksi kerusakan truk *mixer*.

# Metode



# Pembahasan

Penyelesain dengan pendekatan CRISP-DM meliputi beberapa fase diantaranya yaitu:

## ***Business Understanding***

Pada tahapan ini dilakukan penetapan masalah yang akan diteliti yaitu mengenai kerusakan truk *mixer* di PT. Varia Usaha Beton di masa mendatang. Data yang digunakan adalah data kerusakan yang diperoleh dari data laporan *preventive maintenance* harian periode Januari 2018 – Desember 2022.

## ***Data Understanding***

Pada tahapan ini dilakukan pengumpulan data.

Yang dimana, data yang diperoleh terdiri dari 4 kriteria yaitu jenis kerusakan sebagai *id* kerusakan, tingkat risiko, frekuensi kerusakan, dan bagian kerusakan. Data yang diperoleh sebanyak 60 data. Jenis kerusakan dilakukan pengelompokkan kerusakan berdasarkan bagian kerusakannya, untuk tingkat risiko dilakukan pembobotan yaitu 1 ditandai dengan ringan, 2 ditandai dengan sedang, 3 ditandai dengan tinggi. Pembobotan ini ditentukan dan dibuat oleh teknisi dan kepala gudang perawatan truk *mixer* di PT. Varia Usaha Beton. Sedangkan untuk frekuensi diambil keseluruhan selama 5 tahun (2018-2022). Untuk bagian kerusakan terdiri dari 5 bagian yaitu chasis *unit*, chasis molen, listrik, mesin, dan kabin. Variabel penilaian dapat dilihat pada tabel 1.

# Pembahasan

**Tabel 1. Variabel Penilaian**



Jenis Variabel	Pilihan	No	Konversi Nilai
Jenis Kerusakan	1 (Chasis unit)	1	Handrem
		2	Tempat aki
		3	Slebor
		...	...
		...	...
		16	Tabung angin
		17	Ban
		18	Seal gardan
		19	Tangki air
		20	Chasis molen
		21	Corong
		...	...
		...	...
		32	Dinding molen
		33	Keran air tandon
		34	Tuas molen
	35	Whipper	
	36	Lampu	
	...	...	
	...	...	
	38	Relay starter	
	39	Terminal accu	
	40	Pompa hidrolis molen	
	41	Selang solar	
42	Minyak booster kopling		
...	...		
...	...		
54	Selang kompresor bensin		
...	...		
55	Kabel sling transmisi		
56	Tensioner fan belt		
57	Handle pintu kabin		
58	Pintu kabin		
59	Gagang spion		
60	Kaca kabin		
Tingkat Resiko	Ringan	1	
	Sedang	2	
	Tinggi	3	
Bagian Kerusakan	Chasis unit		
	Chasis molen		
	Listrik		
	Mesin		

# Pembahasan

## **Data Preparation**

Tahapan ini adalah tahapan untuk membangun *dataset*. Sebelum membangun sebuah *dataset* dilakukan pembersihan data, pembersihan dilakukan pada data yang tidak memiliki nilai (null) dan data-data yang sama atau diambil secara berulang pada saat proses *scrapping*. Data-data yang tidak dibersihkan nantinya akan mempengaruhi performansi dari sistem data *mining* karena berkurangnya jumlah dan komplektisitasnya. Pada tahapan ini data masih sebanyak 60, artinya pada tahap awal sampai pada tahap ini jumlahnya sama tidak ada yang dibuang. *Dataset final* dapat dilihat pada tabel 3.

## **Modelling**

Pada tahapan ini dilakukan pemodelan pada *dataset final* dengan menggunakan algoritma *K-Nearest Neighbor* (K-NN). Adapun pembagian dataset yaitu 60:40, 70:30, 80:20, dan 90:10. Data yang akan dibagi sesuai dengan variansi data yang dijelaskan diatas kemudian akan dikelompokkan menjadi data latih dan data training dengan dilakukan pengacakan pada sistem agar terbagi secara menyeluruh. Penentuan nilai k menggunakan google colaboratory. Nilai k yang digunakan adalah  $k = 20$ . Selanjutnya perhitungan jarak dengan menggunakan *euclidean distance*.

# Pembahasan

**Tabel 2.** Data Kerusakan Truk *Mixer*

Jenis Kerusakan	Tingkat Risiko	2018	2019	2020	2021	2022	Bagian Kerusakan
		Frekuensi Kerusakan	Frekuensi Kerusakan	Frekuensi Kerusakan	Frekuensi Kerusakan	Frekuensi Kerusakan	
1	1	24	21	23	20	21	Chasis unit
2	1	5	1	4	3	2	Chasis unit
3	1	5	7	5	5	6	Chasis unit
4	2	10	12	9	9	11	Chasis unit
5	2	27	24	30	28	26	Chasis unit
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
56	2	2	3	2	1	4	Mesin
57	2	3	2	1	3	1	Kabin
58	2	3	4	3	5	4	Kabin
59	2	3	1	1	2	1	Kabin
60	2	5	6	5	6	6	Kabin

**Tabel 3.** *Dataset Final*

Jenis Kerusakan	Tingkat Resiko	Frekuensi Kerusakan	Bagian Kerusakan
1	1	109	Chasis unit
2	1	15	Chasis unit
3	1	28	Chasis unit
4	2	51	Chasis unit
5	2	135	Chasis unit
...	...	...	...
...	...	...	...
56	2	12	Mesin
57	2	10	Kabin
58	2	19	Kabin
59	2	8	Kabin
60	2	28	Kabin



# Pembahasan

Data latih dan data uji yang digunakan pada pembagian data 90% data latih dan 10% data uji dapat dilihat pada tabel 4 dan tabel 5.

**Tabel 4.** Data Latih

No	Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan
1	34	2	44	Chasis molen
2	49	2	87	Mesin
3	13	2	76	Chasis unit
4	58	2	19	Kabin
5	47	2	32	Mesin
...	...	...	...	...
...	...	...	...	...
50	43	2	8	Mesin
51	15	2	123	Chasis unit
52	29	2	83	Chasis molen
53	52	2	15	Mesin
54	39	2	40	Listrik

**Tabel 5.** Data Uji

No	Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan
1	1	1	109	Chasis unit
2	6	2	102	Chasis unit
3	37	2	60	Chasis molen
4	46	2	37	Mesin
5	14	1	55	Chasis unit
6	55	2	27	Mesin

# Pembahasan

Selanjut perhitungan jarak menggunakan *euclidean distance*, perhitungan manual berikut adalah contoh perhitungan pada data latih 90% dan data uji 10%. Nilai yang diambil yaitu baris ke satu dari data latih dan data uji. Berikut perhitungan manual *euclidean distance*:

$$d = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2}$$
$$d = \sqrt{(2 - 1)^2 + (44 - 109)^2}$$
$$d = 65,0077$$

Selanjutnya jika perhitungan *euclidean distance* selesai. Dilakuakn pengurutan jarak data dari yang terkecil hingga terbesar, hasil jarak data dapat dilihat pada tabel 6.

**Tabel 6.** Hasil Jarak Data Setelah Diurutkan

Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan	Euclidean Distance	Sortir
31	2	110	Chasis molen	1,4142	1
44	2	113	Mesin	4,1231	2
28	2	99	Chasis molen	10,0499	3
32	1	122	Chasis molen	13,0000	4
15	2	123	Chasis unit	14,0357	5
...	...	...	...	...	...
...	...	...	...	...	...
17	2	256	Chasis unit	147,0034	54

# Pembahasan

Selanjutnya adalah pengujian dengan menggunakan 90% data latih dan 10% data uji, untuk melihat hasil prediksi kerusakan dengan perbandingan kelas yang ada. Hasil perbandingan dapat dilihat pada tabel 7.

**Tabel 7.** Hasil Uji Coba

No	Jenis Kerusakan	Tingkat Risiko	Frekuensi Kerusakan	Bagian Kerusakan (Asli)	Bagian Kerusakan (Prediksi)	Hasil Pengujian
1	1	1	109	Chasis unit	Chasis unit	Benar
2	6	2	102	Chasis unit	Chasis unit	Benar
3	37	2	60	Chasis molen	Chasis unit	Salah
4	46	2	37	Mesin	Mesin	Benar
5	14	1	55	Chasis unit	Chasis unit	Benar
6	55	2	27	Mesin	Mesin	Benar

# Pembahasan

## Evaluation

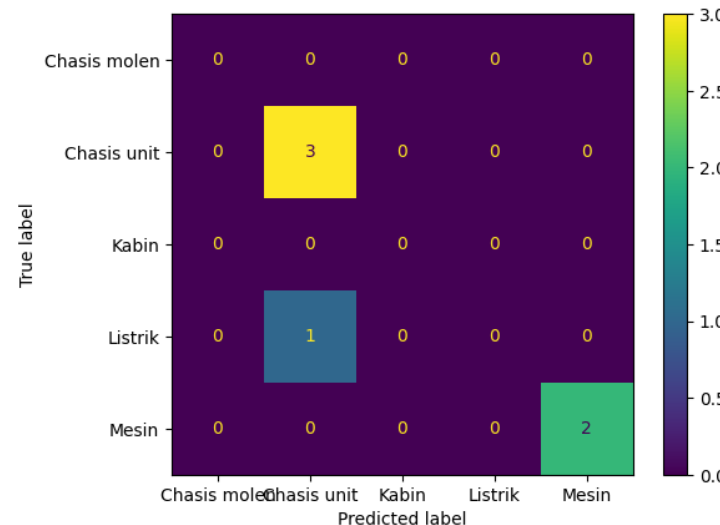
Tahapan ini adalah tahapan untuk mengevaluasi hasil perhitungan algoritma K-NN. Evaluasi yang dilakukan dapat dilihat dari nilai akurasi yang dihasilkan dengan menggunakan *confusion matrix*.

$$\text{Akurasi} = \frac{\text{Jumlah data terprediksi benar}}{\text{Jumlah seluruh data}} \times 100\%$$

Pada tabel 8 dan gambar 1, adalah contoh confusion matrix dan perhitungan akurasi dari pembagian 90% data latih dan 10% data uji.

**Tabel 8.** Hasil Uji Coba Pada Data Latih 90% dan Data Uji 10%

Jenis Kerusakan	Tingkat Resiko	Frekuensi Kerusakan	Bagian Kerusakan (Asli)	Bagian Kerusakan (Prediksi)	Hasil Pengujian
1	1	109	Chasis unit	Chasis unit	Benar
6	2	102	Chasis unit	Chasis unit	Benar
37	2	60	Chasis molen	Chasis unit	Salah
46	2	37	Mesin	Mesin	Benar
14	1	55	Chasis unit	Chasis unit	Benar
55	2	27	Mesin	Mesin	Benar



**Gambar 1.** Confusion Matrix

Pada gambar 1 dan tabel 8, diperoleh kelas terbanyak pada hasil prediksi yaitu chasis *unit*. Maka dapat disimpulkan, bahwa kemungkinan kerusakan di masa mendatang akan terjadi pada bagian chasis *unit* dengan akurasi sebesar 83%, hasil akurasi dibuktikan dengan perbedaan hasil yang diprediksi dengan data asli.

# Hasil

Setelah dilakukan perhitungan akurasi secara menyeluruh. Dilakukan rekapitulasi jumlah kerusakan truk *mxer* berdasarkan bagian kerusakan periode Januari 2018-Desember 2022, dapat dilihat pada tabel 9.

**Tabel 9.** Rekapitulasi Jumlah Kerusakan Truk *Mixer* (Data Asli)

Bagian Kerusakan	Total Kerusakan				
	2018	2019	2020	2021	2022
Chasis unit	337	322	365	322	348
Chasis molen	271	273	252	274	265
Listrik	112	104	107	104	110
Mesin	163	184	163	143	159
Kabin	14	13	10	16	12

Berdasarkan tabel 9, data asli perusahaan menunjukkan bahwa total kerusakan yang banyak terjadi yaitu pada bagian kerusakan chasis *unit*. Hal ini dapat menjadi acuan dalam membandingkan hasil prediksi yang dilakukan pada setiap variasi *dataset*. Dalam hal ini data latih 90% dan data uji 10% diprediksi benar dan sesuai dengan tingkat akurasi 83%. Kemudian pada data latih 80% dan data uji 20% diprediksi untuk kerusakan di periode selanjutnya adalah chasis *unit* dan/ atau mesin dengan akurasi 75%. Kemudian pada data latih 70% dan data uji 30% diprediksi untuk kerusakan di periode selanjutnya adalah chasis molen dengan akurasi 39%. Sedangkan pada data latih 60% dan data uji 40% diprediksi untuk kerusakan di periode selanjutnya adalah chasis molen dengan akurasi 33%.

**Gambar 1.** *Confusion Matrix*

# Manfaat Penelitian

Hasil dari penelitian ini yaitu dapat menjadi rekomendasi perencanaan perawatan kepada perusahaan maupun usaha lainnya untuk mendeteksi kerusakan truk *mixer* atau kendaraan lainnya, agar dapat: mengevaluasi interval waktu perawatan kendaraan, untuk meningkatkan *availability*, merencanakan penyediaan kebutuhan dan sumber daya manusia dalam penjadwalan waktu perawatan agar penanganan kerusakan dapat teratasi dengan cepat dan tepat.

# Referensi

- [1] C. Dhaniel, M. Pamadi, dan A. Savitri, "Pengelolaan Proyek Pembangunan Perumahan The Icon Menggunakan Manajemen Konstruksi di Masa Pandemi Covid-19," vol. 5, hal. 548–557, 2022.
- [2] A. M. Liang dan Koespiadi, "Pengaruh Mutu Material Beton Terhadap Efisiensi Biaya Pembangunan Gedung Bertingkat," vol. 3, hal. 1–8, 2019.
- [3] M. H. A. Sarwandy dan N. Royan, "Produktivitas Alat Berat Excavator Backhoe Pada Proyek Perumahan Al Zafa Tegal Binangun Kota Palembang," hal. 121–125.
- [4] K. Lorosae, A. I. Sembiring, dan S. Debataraja, "Analisa Produktivitas Alat Berat Pada Pekerjaan Beton Ready Mix Studi Kasus Bangunan Pelimpah Bendungan Lau Simeme," vol. 11, no. 1, hal. 95–111, 2023.
- [5] A. A. W. P. R, F. Rozi, dan F. Sukmana, "Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor," *JIPi (Jurnal Ilm. Penelit. dan Pembelajaran Inform.,* vol. 6, no. 1, hal. 155–160, 2021, doi: 10.29100/jipi.v6i1.1910.
- [6] T. Jaelani, M. Yamin, dan C. P. Mahandari, "Machine Learning untuk Prediksi Produksi Gula Nasional," *JMPM (Jurnal Mater. dan Proses Manufaktur),* vol. 6, no. 1, hal. 31–36, 2022, doi: 10.18196/jmpm.v6i1.14897.
- [7] F. M. Subqi dan D. Anggraini, "Data Mining untuk Pemeliharaan Prediktif Mesin Produksi Berdasarkan Database Kerusakan Mesin Menggunakan Naïve Bayes Classifier," *J. Ilm. Komputasi,* vol. 20, no. 2, hal. 143–154, 2021, doi: 10.32409/jikstik.20.2.368.
- [8] M. N. Maskuri, Harliana, K. Sukerti, dan R. M. H. Bhakti, "Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke," *J. Ilm. Intech Inf. Technol. J. UMUS,* vol. 4, no. 1, hal. 130–140, 2022.
- [9] M. A. Hasanah, S. Soim, dan A. S. Handayani, "Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir," *J. Appl. Informatics Comput.,* vol. 5, no. 2, hal. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [10] S. Saepudin, M. Muslih, dan Sihabudin, "Pemilihan Jurusan Dengan Metode K-Nearest Neighbor Untuk Calon Siswa Baru," *J. Rekayasa Teknol. Nusa Putra,* vol. 5, no. 2, hal. 15–19, 2019.
- [11] I. G. Gusti, M. Nasrun, dan R. A. Nugrahaeni, "Rekomendasi Sistem Pemilihan Mobil Menggunakan K-Nearest Neighbor (KNN) Collaborative Filtering," *TEKTRIKA - J. Penelit. dan Pengemb. Telekomun. Kendali, Komputer, Elektr. dan Elektron.,* vol. 4, no. 1, hal. 26, 2019, doi: 10.25124/tektrika.v4i1.1846.
- [12] A. N. Utomo dan N. Lestari, "Deteksi Kerusakan Jalan Raya Menggunakan Algoritma K-NN (K-Nearest Neighbor)," vol. 10, no. 1, hal. 59–66, 2021.

# Referensi

- [13] Cumel, D. Zamri, Rahmadden, dan Syamsurizal, "Perbandingan Metode Data Mining untuk Prediksi Banjir dengan Algoritma Naive Bayes dan KNN," *SENTIMAS Semin. Nas. Penelit. dan ...*, hal. 40–48, 2022, [Daring]. Tersedia pada: <https://journal.irpi.or.id/index.php/sentimas/article/view/353%0Ahttps://journal.irpi.or.id/index.php/sentimas/article/download/353/132>
- [14] M. Y. R. Rangkuti, M. V. Alfansyuri, dan W. Gunawan, "Penerapan Algoritma K-Nearest Neighbor (KNN) dalam Memprediksi dan Menghitung Tingkat Akurasi Data Cuaca di Indonesia," *Hexagon*, vol. 2, no. 2, hal. 11–16, 2021, doi: 10.36761/hexagon.v2i2.1082.
- [15] B. Y. A. Pratama dan H. A. Yuniarto, "Perancangan Proses Implementasi Machine Learning dalam Maintenance Management untuk Mencegah Derating," *J@ti Undip J. Tek. Ind.*, vol. 16, no. 2, hal. 134–142, 2021, doi: 10.14710/jati.16.2.134-142.
- [16] A. D. A. Putra dan S. Juanita, "Analisis Sentimen pada Ulasan Pengguna Aplikasi Bibit dan Bareksa dengan Algoritma KNN," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 2, hal. 636–646, 2021, doi: 10.35957/jatisi.v8i2.962.
- [17] F. Hasanah, T. Suprpti, N. Rahaningsih, dan I. Ali, "Implementasi Algoritma K-Nearest Neighbor dalam Menentukan Buku Berdasarkan Peminatan," *J. Account. Inf. Syst.*, vol. 5, no. 1, hal. 102–111, 2022, doi: 10.32627/aims.v5i1.467.
- [18] P. Putra, A. M. H. Pardede, dan S. Syahputra, "Analisis Metode K-Nearest Neighbour (KNN) Dalam Klasifikasi Data Iris Bunga," *J. Tek. Inform. Kaputama*, vol. 6, no. 1, hal. 297–305, 2022.
- [19] Ariyadi, "Klasifikasi Spesies Lebah Berbasis Data Citra dengan Metode Support Vector Machine," *J. Inov. Penelit.*, vol. 1, no. 6, hal. 1065–1070, 2020.
- [20] A. A. D. Halim dan S. Anraeni, "Analisis Klasifikasi Dataset Citra Penyakit Pneumonia menggunakan Metode K-Nearest Neighbor (KNN)," *Indones. J. Data Sci.*, vol. 2, no. 1, hal. 01–12, 2021, doi: 10.33096/ijodas.v2i1.23.
- [21] M. Wahyudi, R. Buatun, dan H. Sembiring, "Diagnosa Gejala Kecanduan Game Online dengan Metode K-Nearest Neighbor," *Semin. Nas. Inform.*, vol. 6, no. 3, hal. 106–117, 2022.
- [22] M. F. Rilwanu, H. Taufikurachman, dan F. Huwaidi, "Penerapan Algoritma K-Nearest Neighbor untuk Mendeteksi Diabetes Berbasis Web Application," vol. 3, no. 1, hal. 145–152, 2022.
- [23] M. A. Wiratama dan W. M. Pradnya, "Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes," vol. 11, hal. 1–12, 2022.
- [24] M. Fansyuri, "Analisa Algoritma Klasifikasi K-Nearest Neighbor Dalam Menentukan Akurasi Terhadap Kepuasan Pelanggan (Studi Kasus PT. Trigatra Komunikatama)," *Humanika J. Ilmu Sos. Pendidikan, dan Hum.*, vol. 3, no. 1, hal. 29–33, 2020.



