

## PENERAPAN BIG DATA ANALYST TERHADAP PENGIRIMAN BARANG CACAT MENGUNAKAN METODE K-MEANS

Ardhana Febriansyah<sup>✉</sup>, Fithry Tahel

Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia

Email: [ardhana.aizu@gmail.com](mailto:ardhana.aizu@gmail.com)

### ABSTRACT

*PT ID Express, a delivery service company, faces the challenge of high levels of damage to goods during the shipping process. This damage ranges from minor scratches or dents to severe damage such as breakage or destruction, which directly impacts customer satisfaction and the company's reputation. To address this issue, a data-driven approach capable of comprehensively identifying patterns and causal factors for damage is required. This study aims to analyze the types of damage to goods based on shipping data using the K-Means Clustering method. This method is used to group damaged goods data into several clusters based on the level of similarity in their characteristics. The results show that the damage data can be grouped into two main clusters: minor damage and major damage, each consisting of five dominant types of damage. Through this clustering, companies can gain a better understanding of frequently occurring damage patterns and can design more effective preventive measures. This research is expected to serve as a reference in the application of big data analysis in the logistics sector to improve service quality and reduce the risk of damage to goods during the distribution process.*

**Keywords:** *Big Data Analyze, Shipping, Defective Goods, K-Means, Clustering.*

### ABSTRAK

*PT ID Express sebagai perusahaan jasa pengiriman menghadapi tantangan berupa tingginya tingkat kerusakan barang selama proses pengiriman. Kerusakan tersebut mencakup kategori ringan seperti goresan atau penyok kecil hingga kerusakan berat seperti pecah atau hancur, yang berdampak langsung terhadap kepuasan pelanggan dan reputasi perusahaan. Untuk mengatasi permasalahan ini, diperlukan pendekatan berbasis data yang mampu mengidentifikasi pola serta faktor penyebab kerusakan secara komprehensif. Penelitian ini bertujuan untuk menganalisis jenis-jenis kerusakan barang berdasarkan data pengiriman dengan menerapkan metode K-Means Clustering. Metode ini digunakan untuk mengelompokkan data kerusakan barang ke dalam beberapa cluster berdasarkan tingkat kesamaan karakteristiknya. Hasil penelitian menunjukkan bahwa data kerusakan dapat dikelompokkan menjadi dua cluster utama, yaitu kerusakan ringan dan kerusakan berat, yang masing-masing terdiri dari lima jenis kerusakan dominan. Melalui hasil pengelompokan ini, perusahaan dapat memperoleh pemahaman yang lebih baik mengenai pola kerusakan yang sering terjadi serta dapat merancang langkah preventif yang lebih efektif. Penelitian ini diharapkan dapat menjadi referensi dalam penerapan analisis big data di bidang logistik guna meningkatkan kualitas layanan dan mengurangi risiko kerusakan barang selama proses distribusi.*

**Kata Kunci:** *Big Data Analisis, Pengiriman, Barang Cacat, K-Means, Clustering.*

### PENDAHULUAN

Perkembangan teknologi informasi yang pesat telah mendorong dunia pendidikan dan sektor lainnya untuk beradaptasi serta memanfaatkan berbagai alat digital (Desi dkk., 2025). Kemajuan teknologi, khususnya di bidang komputer, telah memberikan kemudahan dalam pengolahan data perhitungan yang kini dapat dikerjakan dengan fasilitas komputer yang semakin canggih dari waktu ke waktu (Astuti & Tahel., 2022). Selain itu, perkembangan teknologi informasi dan komunikasi (TIK) dalam satu dekade terakhir juga menawarkan solusi yang lebih efektif untuk menyelesaikan berbagai permasalahan (Syahputrid kk., 2025). Komputer sendiri merupakan mesin elektronik

yang bekerja berdasarkan instruksi yang tersimpan dalam memori, mampu menerima data, memproses sesuai aturan tertentu, menghasilkan output, dan menyimpannya untuk penggunaan di masa mendatang (Kesuma & Tahel., 2023).

PT ID Express merupakan salah satu perusahaan logistik terkemuka di Indonesia yang menyediakan layanan pengiriman barang secara domestik dan internasional. Perusahaan ini dikenal karena efisiensi dan kecepatan pengiriman yang menjadi keunggulan kompetitif di pasar. Seiring meningkatnya permintaan layanan e-commerce, perusahaan terus berinovasi untuk menjaga kualitas pelayanan dan memastikan kondisi barang tetap baik selama proses pengiriman.

Namun, permasalahan sering muncul berupa barang rusak atau cacat yang tidak hanya merugikan pelanggan, tetapi juga berdampak negatif terhadap citra perusahaan. Berdasarkan data internal, jenis kerusakan yang terjadi bervariasi, mulai dari kerusakan ringan seperti goresan hingga kerusakan berat seperti barang pecah. Kurangnya pemahaman terhadap pola kerusakan membuat perusahaan kesulitan menentukan langkah preventif yang tepat.

Untuk mengatasi masalah tersebut, diperlukan pendekatan berbasis data yang mampu menganalisis pola kerusakan barang secara menyeluruh. Salah satu solusi yang dapat diterapkan adalah penggunaan Big Data Analisis, yang berfungsi untuk mengolah data dalam jumlah besar agar dapat mengidentifikasi pola yang berguna untuk pencegahan kerusakan di masa depan. Big data sendiri merupakan kumpulan data yang sangat besar dan kompleks, baik terstruktur maupun tidak, yang tidak dapat diolah dengan metode tradisional (Nurisyifa & Agustin., 2024). Ciri utamanya meliputi Volume (skala besar), Velocity (kecepatan tinggi), dan Variety (beragam format) (Azruna & Sinaga., 2024), serta berfokus pada penyimpanan dan pengelolaan data masif untuk menemukan wawasan, tren, dan pola tersembunyi (Ramadhana & Puspasari., 2023).

Selain itu, konsep data mining digunakan untuk menganalisis data dari berbagai perspektif guna menemukan pola dan hubungan bermakna (Purwadia & Rosnelly., 2023). Teknik ini memanfaatkan metode statistik, machine learning, dan algoritma khusus untuk mengekstraksi informasi berharga dari data mentah (Nabila & Haryanto., 2024). Tujuannya adalah memprediksi tren masa depan dan membantu pengambilan keputusan yang lebih akurat (Karim dkk., 2021).

Metode yang digunakan dalam penelitian ini adalah K-Means Clustering, yaitu teknik yang mampu mengelompokkan data berdasarkan kemiripan karakteristik tertentu (Filki., 2022). Algoritma ini bekerja dengan membagi data ke dalam beberapa kelompok (cluster) berdasarkan kedekatan terhadap titik pusat (centroid) yang diperbarui secara iteratif (Akbar & ujiati., 2024). Berdasarkan penelitian terdahulu, penerapan metode K-Means telah membantu dalam pengelompokan data penerima bantuan biaya pendidikan sehingga proses analisis data menjadi lebih efisien (Andayania & Nababan., 2024). Metode ini juga mempermudah proses identifikasi dan alokasi bantuan sosial secara lebih tepat sasaran dengan hasil klasifikasi yang baik (Saria & Sembiring., 2024). Selain itu, penerapan algoritma K-Means mampu mengelompokkan data siswa baru berdasarkan

aktivitas dan nilai-nilai yang diperoleh selama proses pembelajaran (Halawa., 2025).

## **KAJIAN LITERATUR**

Big Data merujuk pada kumpulan data yang sangat besar, baik dalam hal volume, kecepatan, dan varietas, yang sehingga terlalu kompleks untuk diolah dan dianalisis dengan metode tradisional. Konsep ini tidak hanya sekadar tentang ukurannya yang "besar", tetapi lebih pada karakteristik utamanya yang sering disebut sebagai 3V (Volume, Velocity, dan Variety). *Volume* menggambarkan skala data yang massive, mulai dari terabyte hingga zettabyte, yang berasal dari berbagai sumber seperti media sosial, sensor, atau transaksi. *Velocity* merujuk pada kecepatan data yang dihasilkan dan harus diproses, sering kali dalam waktu nyata atau hampir nyata. Sementara *Variety* menekankan pada beragamnya format data, mulai dari yang terstruktur (seperti database) hingga yang tidak terstruktur (seperti teks, video, dan gambar). Dengan memanfaatkan teknologi dan algoritma khusus, Big Data dapat dianalisis untuk mengungkap pola, tren, dan asosiasi yang sebelumnya tersembunyi, terutama yang berkaitan dengan perilaku dan interaksi manusia. Pada akhirnya, tujuan dari Big Data adalah untuk menghasilkan wawasan (insight) yang dapat mendukung pengambilan keputusan yang lebih tepat, inovatif, dan strategis di berbagai bidang, mulai dari bisnis hingga kesehatan (Mayasari & Agussalim., 2023).

K-Means adalah salah satu algoritma *unsupervised learning* (pembelajaran tanpa pengawasan) yang populer untuk melakukan pengelompokan (*clustering*). Tujuan utamanya adalah untuk mengelompokkan sejumlah data ke dalam K buah kluster (kelompok), di mana setiap data dalam satu kluster memiliki karakteristik yang serupa dan berbeda dengan data di kluster lainnya. Cara kerjanya berpusat pada konsep centroid, yang merupakan titik pusat rata-rata dari semua data dalam sebuah kluster. Algoritma ini dimulai dengan memilih secara acak K buah centroid awal, kemudian melakukan proses iteratif yang terdiri dari dua langkah utama. Pertama, setiap titik data ditugaskan ke centroid terdekat berdasarkan jarak (biasanya jarak Euclidean), sehingga membentuk K kluster. Kedua, centroid diperbarui dengan menghitung rata-rata dari semua titik data yang tergabung dalam kluster tersebut. Kedua langkah ini akan terus berulang hingga tidak ada lagi perubahan penugasan data ke kluster yang berbeda, yang menandakan bahwa centroid telah stabil dan kluster optimal telah terbentuk. Kelebihan utama K-Means adalah sederhana, cepat, dan efisien untuk data

dalam jumlah besar. Namun, kelemahannya adalah peneliti harus menentukan nilai  $K$  dari awal, dan hasilnya bisa sensitif terhadap pemilihan centroid awal serta kurang efektif untuk data yang bentuk klusternya tidak bulat (spherical). Metode ini banyak diterapkan dalam segmentasi pelanggan, kompresi gambar, analisis genetika, dan eksplorasi data lainnya.. Berikut merupakan proses Algoritma metode *K-Means* :

1. Menentukan banyak klaster yang diinginkan.
2. Menentukan *centorid* awal (pusat *cluster*), biasanya dipilih secara acak.
3. Menghitung jarak terdekat dari setiap objek pengamatan dengan *centroid* awal yang sudah ditentukan dengan rumus jarak *euclidean*.

$$d(x_i, x_j) = \sqrt{\sum_{l=1}^n (x_{il} - x_{jl})^2} \dots \dots \dots (1)$$

$d(x_i, x_j)$  = Jarak antara objek ke  $i$  dengan objek ke  $j$ .

$x_{il}$  = Nilai objek ke  $i$  pada variabel  $k$

$x_{jl}$  = Nilai objek ke  $j$  pada variabel  $k$

$n$  = Banyaknya variabel.

4. Menentukan jarak terdekat, antara objek dengan *centroid*.
5. Menentukan *centroid* baru dengan menghitung rata-rata masing-masing *cluster* menggunakan rumus.

$$C_{kl} = \frac{x_{1l} + x_{2l} + \dots + x_{pl}}{p} \dots \dots \dots (2)$$

Keterangan:

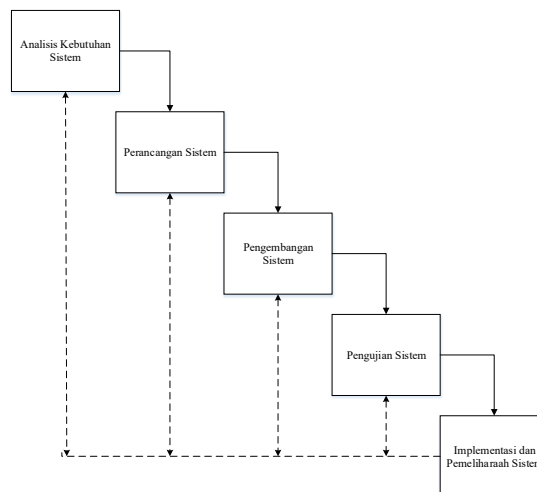
$C_{kl}$  = Nilai *centroid* ke- $k$  pada variabel ke- $l$

$p$  = Banyaknya data

6. Hitung pada setiap objek menggunakan *cluster* baru. Jika objek tidak berpindah *cluster*, maka proses *clustering* selesai. Atau ulangi langkah 3 sampai pusat *cluster* tidak berubah (Hasymi dkk., 2021).

## METODE PENELITIAN

Metode pengembangan sistem merupakan tahapan terstruktur yang digunakan untuk merancang, membangun, dan memelihara sistem informasi atau perangkat lunak. Proses ini mencakup analisis kebutuhan, perancangan, pengembangan, pengujian, implementasi, dan pemeliharaan sistem. Salah satu model yang sering digunakan adalah model Waterfall, yang menekankan urutan langkah yang sistematis. Tujuan utama metode ini adalah menghasilkan sistem yang efisien, andal, sesuai kebutuhan pengguna, serta mampu beradaptasi dengan perkembangan di masa mendatang.



**Gambar 1.** Diagram Waterfall

Keterangan Gambar:

1. Analisis Kebutuhan Sistem

Tahap ini bertujuan mengidentifikasi kebutuhan PT ID Express terkait pengelolaan data kerusakan barang. Informasi diperoleh melalui wawancara dengan staf pengiriman untuk memahami jenis kerusakan, dampak terhadap operasional, serta kebutuhan solusi yang efektif dalam mengurangi kerusakan barang.

2. Perancangan Sistem

Pada tahap ini dilakukan perancangan arsitektur dan alur kerja sistem berdasarkan hasil analisis kebutuhan. Sistem dirancang untuk mengumpulkan dan menganalisis data kerusakan menggunakan metode K-Means Clustering, serta menyediakan antarmuka yang mudah digunakan untuk input data dan pemantauan hasil analisis.

3. Pengembangan Sistem

Tahap ini melibatkan pembuatan perangkat lunak menggunakan teknologi seperti PHP dan integrasi dengan database untuk menyimpan serta memproses data kerusakan barang. Sistem akan mengelompokkan data menjadi dua kategori utama, yaitu kerusakan ringan dan berat.

4. Pengujian Sistem

Pengujian dilakukan untuk memastikan seluruh fitur sistem berfungsi dengan baik dan hasil analisis K-Means Clustering akurat. Proses ini menggunakan data uji yang merepresentasikan berbagai jenis kerusakan barang.

5. Implementasi dan Pemeliharaan Sistem

Tahap akhir meliputi penerapan sistem di PT ID Express serta pemeliharaan berkala untuk memastikan sistem tetap optimal dan dapat menyesuaikan dengan kebutuhan atau perubahan di masa mendatang.

## HASIL DAN PEMBAHASAN

Adapun langkah-langkah kasus K-means dalam analisis terhadap pengiriman barang cacat dapat dijelaskan sebagai berikut:

### 1. Atribut

Atribut merupakan karakteristik atau ciri-ciri yang digunakan untuk menggambarkan suatu objek atau entitas dalam suatu system. Adapun atribut yang digunakan dalam penelitian ini dapat dijelaskan pada tabel berikut.

**Tabel 1.** Atribut

No	Atribut	Keterangan
A1	Goresan dan Lecet (Surface Scratches)	Barang mengalami kerusakan ringan pada permukaannya, seperti goresan atau lecet, yang tidak memengaruhi fungsi.
A2	Kerusakan Kemasan (Packaging Damage)	Kerusakan hanya terjadi pada kemasan luar, seperti kotak penyok atau sobek, tanpa memengaruhi barang di dalamnya.
A3	Deformasi Kecil (Minor Deformation)	Barang mengalami perubahan bentuk kecil, seperti sedikit bengkok atau terlipat, tetapi masih dapat dikembalikan ke bentuk asalnya.
A4	Kotor atau Berdebu (Dirt or Dust)	Barang hanya terlihat kotor akibat kontaminasi debu atau kotoran selama pengiriman, tetapi tetap dalam kondisi baik.
A5	Stiker atau Label Rusak (Damaged Labels)	Stiker atau label pada kemasan terkelupas atau rusak, tetapi tidak memengaruhi isi barang.
A6	Barang Pecah atau Retak (Broken or Cracked Items)	Barang mengalami kerusakan fisik, seperti retak atau pecah, sehingga tidak dapat digunakan.
A7	Barang Hancur Total (Destroyed Items)	Barang tidak dapat diperbaiki karena tingkat kerusakannya terlalu parah, seperti hancur atau terbelah.
A8	Kerusakan Fungsi (Functional Damage)	Barang terlihat utuh secara fisik tetapi tidak berfungsi, seperti perangkat elektronik yang mati.
A9	Barang Hilang Komponen Utama (Missing Main Components)	Bagian penting dari barang, seperti aksesoris utama atau komponen vital, hilang sehingga barang tidak dapat digunakan sebagaimana mestinya.
A10	Barang Rusak Akibat Air (Severe Water Damage)	Barang rusak secara permanen karena terpapar air, menyebabkan material membengkak, berkarat, atau rusak struktur dalamnya.

### 2. Cluster

*Cluster* merupakan pengelompokan data atau objek berdasarkan kesamaan karakteristik atau fitur tertentu. Adapun *cluster* yang digunakan sebagai berikut :

**Tabel 2.** Cluster

No	Cluster
1	Kerusakan Ringan
2	Kerusakan Berat

### 3. Dataset

Adapun sekumpulan dataset yang digunakan dalam penelitian ini dapat dari perusahaan yaitu barang yang mengalami kecacatan pada saat proses pengiriman. Adapun datasetnya dapat dijelaskan sebagai berikut :

**Tabel 3.** Dataset

T	A1	A2	A3	A4	A5	...	A10
1	100	75	42	96	89	...	88
2	28	97	80	77	84	...	79
3	62	73	60	1	78	...	39
4	86	79	39	59	88	...	97
5	75	40	100	77	50	...	11
6	16	72	94	91	83	...	87
7	6	72	90	4	53	...	78
8	94	80	74	66	86	...	72
9	38	47	54	81	61	...	74
10	60	65	75	83	72	...	61
...	...	...	...	...	...	...	...
125	61	52	19	87	27	...	7
126	78	82	30	68	66	...	83
127	93	65	82	63	65	...	80
128	34	75	70	83	95	...	98
129	85	77	1	33	71	...	78
130	100	92	84	66	71	...	99
131	45	62	61	29	0	...	67
132	96	73	75	75	64	...	62

Dalam metode *K-Means*, pemilihan pusat *cluster* awal (*initial centroid*) merupakan tahap penting karena akan memengaruhi hasil akhir proses pengelompokan data. Pada penelitian ini, data ke-88 dan ke-89 dipilih sebagai *centroid* awal untuk membentuk dua *cluster*. Dengan pemilihan data 88 dan 89 dianggap tepat untuk dijadikan pusat *cluster* awal sebelum algoritma melakukan proses iterasi penghitungan jarak dan pembaruan *centroid* selanjutnya.

**Tabel 4.** Cluster Awal

T	A1	A2	A3	A4	A5	...	A10
88	70	30	68	57	61	...	95
89	75	76	68	66	61	...	79

Setelah menetapkan nilai *k* dan pusat *cluster* awal, langkah berikutnya adalah menghitung jarak antara setiap data dengan pusat *cluster*:

$$C1.88 = \sqrt{(100 - 70)^2 + \dots + (88 - 95)^2}$$

$$C1.88 = \sqrt{(30)^2 + \dots + (-7)^2}$$

$$C1.88 = \sqrt{900 + \dots + 49}$$

$$C1.88 = 91.110$$

$$C2.89 = \sqrt{(28 - 75)^2 + \dots + (79 - 79)^2}$$

$$C2.89 = \sqrt{(-47)^2 + \dots + (-9)^2}$$

$$C2.89 = \sqrt{2209 + \dots + 81}$$

$$C2.89 = 107.238$$

dan seterusnya.

Bandingkan setiap *cluster*, cari nilai *cluster* yang paling terkecil, “C1=91,110”, dan “C2=107,238” sehingga dapat disimpulkan dari 2 *cluster* tersebut yang memiliki nilai *cluster* terkecil adalah *cluster* C1 sehingga untuk iterasi pertama dan untuk masuk kedalam *cluster* C1. Setelah perhitungan pada semua data selesai, hasilnya dapat dilihat pada tabel di bawah ini.

**Tabel 5.** Iterasi 1

No	C1	C2	Cluster
1	91.110	107.238	C1
2	130.996	92.190	C2
3	141.393	111.897	C2
4	99.318	85.539	C2
5	140.485	105.513	C2
6	137.394	113.965	C2
7	133.776	113.697	C2
8	110.368	53.907	C2
9	83.409	73.661	C2
10	91.940	67.469	C2
...	...	...	...
125	135.654	116.709	C2
126	105.128	70.986	C2
127	95.258	73.790	C2
128	120.304	114.569	C2
129	123.296	85.691	C2
130	111.274	73.383	C2
131	93.038	104.005	C1
132	88.544	58.489	C2

Untuk meng-*update* nilai titik pusat *cluster*, dapat dilakukan dengan menghitung rata-rata setiap *cluster* yang muncul. Adapun contoh perhitungan ya dapat dijelaskan sebagai berikut :

$$C_{1.A1} = 76,574$$

$$C_{2.A1} = 65,024$$

Adapun hasil *cluster* yang baru dapat dilihat sebagai berikut:

**Tabel 6.** Cluster Iterasi 1

C	A1	A2	...	A10
C1	76.574	69.128	...	72.383
C2	60.024	68.282	...	72.894

Lakukan proses iterasi secara berulang sampai setiap *cluster* yang di *update* memiliki kesamaan.

Setelah pengelompokan *cluster* didapat pada iterasi ke-9 ya dapat dilihat sebagai berikut.

**Tabel 7.** Iterasi 9

No.	C1	C2	Cluster
1	76.754	96.836	C1
2	97.260	61.511	C2
3	105.055	98.527	C2
4	51.438	75.283	C1
5	118.590	92.920	C2
6	116.376	77.675	C2
7	115.807	93.752	C2
8	74.700	72.028	C2
9	53.364	45.890	C2
10	50.650	31.478	C2
...	...	...	...
125	90.615	107.625	C1
126	39.912	66.618	C1
127	59.007	47.763	C2
128	106.822	82.472	C2
129	65.512	90.301	C1
130	70.802	62.096	C2
131	88.051	108.137	C1
132	48.245	50.710	C1

**Tabel 8.** Jumlah Clustering

No	Cluster	Jumlah	Persentase
1	C1	47	35,61%
2	C2	85	64,39%

Sehingga dapat disimpulkan jumlah *cluster* 1 (47) dengan nilai persentase sebesar 35.61% menyatakan “Kerusakan Ringan”, untuk *cluster* 2 (85) dengan nilai persentase sebesar 64.39% menyatakan “Kerusakan Berat”. Dari ke-2 *cluster* tersebut menyatakan “Kerusakan Berat” lebih besar dari *cluster* lainnya. Sehingga 2 tahun terakhir jenis kerusakan atau cacat masuk dalam kategori “Kerusakan Berat” terhadap pengiriman barang pada PT ID Express.

## KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa penerapan Big Data Analisis di PT ID Express telah berhasil meningkatkan kemampuan perusahaan dalam mengidentifikasi pola kerusakan barang secara lebih sistematis dan menyeluruh. Dengan memanfaatkan metode K-Means Clustering, data kerusakan berhasil dikelompokkan ke dalam dua kluster utama, yaitu kerusakan ringan dan kerusakan berat. Klasifikasi ini terbukti sangat memudahkan analisis lebih lanjut untuk menentukan faktor penyebab dan langkah perbaikan yang tepat sesuai dengan tingkat kerusakan masing-masing kategori. Secara keseluruhan, hasil analisis mengungkap bahwa kerusakan ringan merupakan insiden yang lebih dominan. Variasi penyebabnya, seperti pengemasan

yang kurang tepat dan penanganan yang tidak hati-hati, menyoroti area perbaikan yang kritis dan dapat segera ditangani untuk meningkatkan kualitas layanan pengiriman secara signifikan.

#### DAFTAR PUSTAKA

- Akbar, M. H. (2024). Penentuan Persediaan Stok Barang Menggunakan K-Means Clustering (Studi Kasus: CV. Papaqi (Alsyifa Grafika)). *Jurnal Info Digit (JID)*, 2(3A), 1761-1771.
- Andayani, S., & Nababan, L. (2024). Algoritma K-Means Clustering Untuk Program Kartu Indonesia Pintar Bagi Siswa SMP Alwasliyah. *Jurnal Info Digit (JID)*, 2(3A), 1688-1703.
- Astuti, E., & Tahel, F. (2022). Pelatihan Pengenalan Microsoft Office Excel Pada Sekolah Menengah Atas Swasta Dharmawangsa Medan. *Jurnal Pengabdian Masyarakat Berkarya*, 1(04), 87-94.
- Azrun, R. P., & Sinaga, M. D. (2024). Analisa Pola Belanja Konsumen Pada PT. Megah Jaya Plasindo Menggunakan Metode Apriori. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(3), 1147-1165.
- Desi, E., Aliyah, S., Elhias, M. A., Anggraini, R., Tahel, F., & Ginting, E. (2025). Pemanfaatan teknologi penggunaan Canva sebagai sarana media pembelajaran interaktif di SMK Sinar Husni. *Jurnal Pemberdayaan Masyarakat*, 2(1), 6-12.
- Filki, Y. (2022). Algoritma K-Means Clustering dalam Memprediksi Penerima Bantuan Langsung Tunai (BLT) Dana Desa. *Jurnal Informatika Ekonomi Bisnis*, 166-171.
- Halawa, L. M. (2025). Klasifikasi data dengan kategori Siswa dan siswi BARU dengan algoritma K-MEANS CLUSTERING. *Jurnal Interaksi Sains Dan Teknologi*, 1(1), 11-22.
- Hasymi, M. A., Faisol, A., & Ariwibisono, F. X. (2021). Sistem Informasi Geografis Pemetaan Warga Kurang Mampu Di Kelurahan Karang Besuki Menggunakan Metode K-Means Clustering. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(1), 284-290.
- Karim, A., Adeni, A., Fitri, F., Fitri, A. N., Hilmi, M., Fabriar, S. R., & Rachmawati, F. (2021). Pemetaan untuk Strategi Dakwah di Kota Semarang Menggunakan Pendekatan Data Mining (Mapping for Da'wah Strategy in Semarang City Using Data Mining Approach). *Jurnal Dakwah Risalah*, 32(1), 40-55.
- Kesuma, G., & Tahel, F. (2023). Perancangan Aplikasi Android Pembuatan Game Tebak Gambar Komponen-Komponen Komputer. *Jurnal Sistem Informasi dan Teknologi Informasi*, 2(1), 75-84.
- Mayasari, E., & Agussalim. (2023). Literature review: Big data dan data analys pada perusahaan. *Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer*, Jurnal Ilmiah Sistem Informasi dan Ilmu Komputer Vol. 3 No. 3 November 2023 p-ISSN: 2827-8135 e-ISSN: 2827-7953, Hal 171-187.
- Nabila, D., & Haryanto, E. V. (2024). Penerapan Metode Regresi Linear Berganda Dalam Data Mining Prediksi Penjualan Produk Pada PT Cipta Mebelindo Berbasis Web. *Jurnal Info Digit (JID)*, 2(3A), 1940-1953.
- Nurisyifa, A. F., & Agustin, F. (2024). Penerapan Data Mining Menggunakan Metode C. 45 Dalam Cari Pola Minat Prangko Pada PT. Pos Indonesia Berbasis Web. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(2), 474-486.
- Purwadi, B., & Rosnelly, R. (2023). Data Mining Pemilihan Jasa Ekspedisi Di Kota Medan Menggunakan Algoritma k-NN (K-Nearest Neighbor). *Jurnal Info Digit (JID)*, 1(2), 671-687.
- Ramadhan, H., & Puspasari, R. (2023). Rancang Bangun Aplikasi Prediksi Pola Penjualan Sparepart Mobil Menggunakan Metode C4. 5 Pada Pt Jaya Diesel. *Jurnal Info Digit (JID)*, 1(3), 932-944.
- Sari, R., & Sembiring, N. S. B. (2024). Optimalisasi Penentuan Penerima Bantuan Bedah Rumah Layak Huni Menggunakan K-Means Clustering di Desa Klambir Lima. *Jurnal Rekayasa Sistem (JUREKSI)*, 2(3 A), 1234-1249.
- Siringoringo, R., & Jamaluddin, J. (2019). Initializing the fuzzy c-means cluster center with particle swarm optimization for sentiment clustering. In *Journal of Physics: Conference Series* (Vol. 1361, No. 1, p. 012002). IOP Publishing.
- Syahputri, N., Andriani, N., Tahel, F., & Nurhayati, N. (2025). Analisis Perancangan Sistem Monitoring Praktik Kerja Lapangan Siswa Smk Jurusan Teknik Komputer Dan Jaringan (TKJ). *JIMR: Journal Of International Multidisciplinary Research*, 4(01), 109-117.