

ANALISIS KLASTERING RISIKO PEKERJAAN MENGGUNAKAN METODE K-MEANS PADA DATA SIKA

Adetya Rielisa Putri, Dinna Yunika Hardiyanti

Sistem Informasi, Universitas Sriwijaya

Jl. Srijaya Negara, Bukit Lama, Kec. Ilir Barat I, Kota Palembang, 30139, Indonesia

adetya.rielisa31@gmail.com

ABSTRAK

Industri migas memiliki potensi bahaya tinggi sehingga memerlukan pengelolaan keselamatan berbasis data operasional yang akurat. Pada PT Pertamina Hulu Rokan Zona 4 Field Limau, data Sistem Izin Kerja Aman (SIKA) selama ini masih dimanfaatkan secara administratif dan belum dianalisis untuk mengidentifikasi pola izin kerja. Penelitian ini bertujuan menggali pola pengelompokan pekerjaan berdasarkan atribut lokasi, jenis pekerjaan, dan durasi izin kerja menggunakan data SIKA. Metode yang digunakan adalah klasterisasi *K-Means* terhadap 4.453 data izin kerja periode 2023–2025 yang telah melalui tahap pembersihan data, pengkodean *one-hot encoding*, dan normalisasi *Min-Max*. Penentuan jumlah kluster optimal dievaluasi menggunakan *Elbow Method*, *Silhouette*, *Davies-Bouldin Index*, *Calinski-Harabasz Index*. Hasil evaluasi menunjukkan konfigurasi optimal terdiri dari dua kluster dengan nilai *Silhouette Score* 0,9306 dan DBI 0,4437. Kluster pertama merepresentasikan pekerjaan rutin dengan durasi izin relatif pendek dan dominasi pekerjaan dingin, sedangkan kluster kedua menggambarkan pekerjaan teknis intensif dengan durasi izin lebih panjang dan berlokasi di area *Field*. Hasil penelitian ini memberikan gambaran objektif pola izin kerja yang dapat digunakan sebagai dasar awal pertimbangan dalam perencanaan pengendalian keselamatan kerja.

Kata kunci : *klastering, K3, K-Means, pola izin kerja, SIKA*

1. PENDAHULUAN

Risiko pekerjaan merupakan isu yang krusial di berbagai sektor industri karena banyak aktivitas kerja memiliki potensi menimbulkan kecelakaan maupun berdampak pada kesehatan pekerja. Industri minyak dan gas (migas) termasuk sektor dengan tingkat bahaya tinggi sehingga membutuhkan manajemen risiko yang terintegrasi dan konsisten [1]. Penerapan manajemen risiko yang efektif menjadi kunci untuk menjaga keselamatan pekerja sekaligus mempertahankan produktivitas serta keberlangsungan operasional perusahaan [2]. Meskipun upaya Keselamatan dan Kesehatan Kerja (K3) terus ditingkatkan, angka kecelakaan kerja di Indonesia masih tinggi. Kementerian Ketenagakerjaan mencatat 462.241 kasus kecelakaan kerja pada tahun 2024 [3], sementara SKK Migas melaporkan 56 kecelakaan di sektor hulu dan 48 di sektor hilir migas pada tahun 2022 [4]. Kondisi ini menunjukkan perlunya pendekatan yang lebih sistematis dan berbasis data untuk mendukung pengendalian risiko.

Di lingkungan operasional, PT Pertamina Hulu Rokan Zona 4 Field Limau telah menerapkan Sistem Izin Kerja Aman (SIKA) sebagai instrument pengendalian K3 untuk memastikan setiap aktivitas pekerjaan lapangan dilakukan sesuai prosedur keselamatan. Namun, dalam praktiknya, data SIKA selama ini masih dimanfaatkan sebagai dokumen administratif dan belum digunakan sebagai sumber analisis untuk mengidentifikasi pola izin kerja. Padahal, atribut dalam data SIKA seperti jenis pekerjaan, lokasi, dan durasi izin kerja memiliki potensi besar untuk dianalisis guna mendukung pengambilan keputusan K3. Permasalahan utama yang

dihadapi adalah belum adanya pemanfaatan data SIKA secara analitis untuk mengelompokkan pola izin kerja yang berpotensi memiliki tingkat risiko berbeda. Ketidakterhubungan antara pencatatan izin kerja dan proses analisis risiko menyebabkan perusahaan kehilangan peluang untuk mendeteksi pekerjaan berulang, pekerjaan berdurasi panjang, atau pekerjaan di lokasi tertentu yang berpotensi tinggi membutuhkan pengendalian K3. Kondisi ini membuat strategi K3 cenderung reaktif karena tidak memiliki dasar pengelompokan pola izin kerja yang dapat dianalisis secara sistematis.

Untuk menjawab permasalahan tersebut, diperlukan pendekatan analitis yang mampu mengolah data dalam skala besar dan mengekstraksi pola tersembunyi. Salah satu metode yang relevan adalah teknik klastering dalam data mining, yang memungkinkan pengelompokan objek berdasarkan kemiripan atribut tertentu [5]. Klastering, khususnya algoritma *K-Means*, telah terbukti efektif dalam berbagai penelitian [6]. Penelitian [7] berhasil mengelompokkan data kecelakaan kerja menjadi tiga kluster berdasarkan faktor seperti kelalaian pekerja, kondisi lantai, dan gangguan perhatian. Penelitian lain oleh [8] menunjukkan bahwa *K-Means* mampu menghasilkan kluster wilayah dengan tingkat kecelakaan lalu lintas rendah, sedang, dan tinggi sehingga membantu penentuan prioritas penanganan. Selain itu, beberapa studi juga menunjukkan efektivitas *K-Means* dalam mengidentifikasi pola risiko pada data operasional lainnya, termasuk pemetaan karakteristik pekerjaan maupun analisis kesehatan berbasis data. Namun, penerapan teknik klastering khususnya pada data izin kerja berbasis

SIKA di industri migas masih relatif terbatas.

Berdasarkan latar belakang dan permasalahan tersebut, penelitian ini bertujuan untuk mengidentifikasi pola pengelompokan izin kerja berdasarkan atribut jenis pekerjaan, Lokasi, dan durasi izin kerja menggunakan metode *K-Means* pada data SIKA PT Pertamina Hulu Rokan Zona 4 Field Limau. Rencana penyelesaian masalah dalam penelitian ini dilakukan melalui pengolahan dan analisis data SIKA menggunakan metode klasterisasi *K-Means* untuk menghasilkan gambaran awal mengenai kelompok pekerjaan dengan karakteristik serupa yang dapat dijadikan sebagai dasar pertimbangan dalam konteks pengelolaan K3, sekaligus mengisi kesenjangan penelitian sebelumnya serta mendukung penerapan strategi keselamatan yang lebih proaktif dan berbasis data.

2. TINJAUAN PUSTAKA

2.1. Sistem Izin Kerja Aman (SIKA)

PT Pertamina sebagai perusahaan energi nasional menempatkan aspek HSSE sebagai bagian penting dari operasionalnya, terutama pada aktivitas berisiko tinggi. Dalam unit kilang dan eksplorasi, Sistem Izin Kerja Aman (SIKA) diterapkan sebagai dokumen administratif wajib yang memuat informasi seperti jenis pekerjaan, lokasi, durasi, identifikasi potensi bahaya, jumlah tenaga kerja, APD, ventilasi, dan persyaratan keselamatan lainnya. SIKA digunakan pada pekerjaan panas, ruang terbatas, pekerjaan di ketinggian, penggalian, dan pemeliharaan peralatan kritikal. Prosesnya meliputi pengajuan izin, verifikasi, peninjauan risiko oleh HSSE, persetujuan manajemen, pelaksanaan pekerjaan, hingga evaluasi [9].

2.2. Keselamatan dan Kesehatan Kerja (K3)

K3 merupakan instrumen untuk memproteksi pekerja, perusahaan, dan lingkungan dari bahaya kecelakaan kerja [1]. ILO mencatat 2,78 juta pekerja meninggal tiap tahun akibat kecelakaan dan penyakit kerja, sehingga K3 menjadi fondasi penting lingkungan kerja produktif. Fokus utama K3 adalah pencegahan kecelakaan dan kerugian besar melalui penerapan prosedur keselamatan, audit, pelatihan, dan penggunaan APD [10].

2.3. Risiko Pekerjaan

Industri migas memiliki risiko tinggi sehingga memerlukan manajemen risiko yang terintegrasi [1]. Manajemen risiko dimulai dengan mengidentifikasi potensi bahaya (*hazard*) dan mengendalikan dampaknya [10]. Potensi bahaya ini mencakup risiko ledakan, kebakaran, ruang terbatas, kecelakaan, dan jatuh dari ketinggian. Upaya pengendalian dilakukan melalui prosedur keselamatan, audit berkala, pelatihan, dan penggunaan APD [11].

2.4. Data Mining

Data mining merupakan proses menemukan pola atau informasi bermakna dari data menggunakan

teknik tertentu untuk menghasilkan pengetahuan [12]. Data mining atau *Knowledge Discovery in Database (KDD)* berfokus pada pemanfaatan data historis untuk menemukan pola pada kumpulan data besar [11].

2.5. Klastering

Klastering merupakan metode data mining *unsupervised* yang mengelompokkan data berdasarkan kemiripan [7]. Tujuannya mengenali karakteristik serupa pada data [13]. Klastering terbagi menjadi *hierarchical* seperti dendrogram dan *non-hierarchical* seperti *K-Means*, *Fuzzy C-Means*, dan *K-Medoid* [14]. Dalam penelitian ini, klastering digunakan untuk mengenali pola izin kerja berdasarkan jenis pekerjaan, lokasi, dan durasi.

2.6. Algoritma K-Means

K-Means adalah metode clustering populer yang efisien dalam mengidentifikasi kemiripan objek berdasarkan jarak [15]. *K-Means* mudah diimplementasikan, efektif, efisien, akurat, dan hasilnya mudah diinterpretasikan [16]. Hasil penelitian [17] menunjukkan *K-Means* menghasilkan klaster yang lebih seimbang dan *silhouette score* lebih baik dibanding *K-Medoids* dan DBSCAN.

2.7. Penelitian Terdahulu

Sejumlah penelitian telah menunjukkan keberhasilan *K-Means* dalam mengidentifikasi pola risiko dan karakteristik pekerjaan dalam berbagai domain. Penelitian [7] mengelompokkan 672 kasus kecelakaan kerja menjadi tiga klaster berdasarkan faktor seperti kelalaian pekerja dan kondisi lingkungan. Penelitian lain oleh [18] menunjukkan bahwa *K-Means* dapat digunakan untuk mengidentifikasi pola risiko pada data kesehatan melalui pemetaan *dyssynchrony* berbasis MRI, membuktikan bahwa teknik klastering mampu mengungkap pola tersembunyi dalam data operasional maupun medis. Di sektor perkebunan, [11] menerapkan klastering untuk mengidentifikasi pola penanganan risiko dan penggunaan APD. Penelitian lain [19] juga menunjukkan bahwa *K-Means* efektif dalam mengidentifikasi kerentanan sosial pekerja informal. Di sektor transportasi, [8] serta [12] memanfaatkan *K-Means* untuk mengelompokkan data kecelakaan lalu lintas dan menghasilkan segmentasi wilayah berdasarkan tingkat kecelakaan. Pendekatan klastering juga digunakan oleh [13] dalam pengelompokan data rekam medis pasien COVID-19, serta oleh [15] untuk mengelompokkan provinsi berdasarkan risiko pandemi. Penelitian lain oleh [6] mengaplikasikan *K-Means* untuk mengelompokkan data lowongan pekerjaan menjadi empat klaster.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan data mining untuk menganalisis pola izin kerja pada data SIKA dengan tujuan mengidentifikasi kelompok pekerjaan berdasarkan kemiripan karakteristiknya.

Proses analisis dilakukan menggunakan algoritma *K-Means Clustering* yang sesuai untuk data berukuran besar dan bersifat campuran kategorikal–numerik. Secara keseluruhan, metode penelitian mengikuti tahapan *Knowledge Discovery in Database (KDD)*. Data mining sendiri memiliki enam tahapan utama untuk mengubah data mentah menjadi informasi yang bermanfaat yang terdiri dari *selection* data, *preprocessing*, *transformation*, *data mining*, *evaluation*, dan *knowledge* [12].

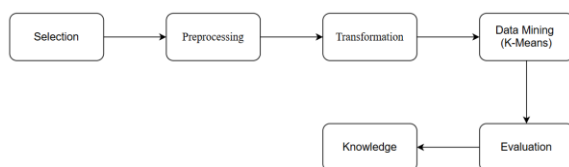
3.1. Data

Data penelitian diperoleh dari Sistem Izin Kerja Aman (SIKA) milik PT Pertamina Hulu Rokan Zona 4 Field Limau untuk periode tahun 2023–2025. Data ini berisi berbagai atribut pekerjaan seperti lokasi pekerjaan, jenis pekerjaan, durasi izin kerja, serta judul pekerjaan yang digunakan untuk proses interpretasi hasil kluster. Data yang digunakan bersifat sekunder dan merupakan data operasional yang dicatat secara rutin dalam sistem SIKA. Atribut yang relevan dipilih untuk dianalisis karena memiliki nilai informatif dalam mengidentifikasi pola pekerjaan, sedangkan atribut administratif seperti identitas pemohon izin tidak digunakan dalam pemodelan.

Jenis data terdiri dari gabungan data kategorikal dan numerik, sehingga diperlukan proses transformasi agar seluruh atribut dapat diproses menggunakan *K-Means*. Variabel kategorikal seperti jenis pekerjaan dan lokasi diubah menggunakan teknik *one-hot encoding*, sedangkan variabel numerik seperti durasi izin kerja dinormalisasi menggunakan *Min-Max normalization* dalam rentang [0,1]. Seluruh data kemudian dibersihkan dari duplikasi, *missing values*, serta penggabungan izin yang merupakan perpanjangan dari pekerjaan yang sama, sehingga setiap entri data mewakili satu aktivitas kerja yang valid.

3.2. Tahapan Penelitian (KDD)

Proses penelitian mengikuti alur KDD yang terdiri dari beberapa tahap berikut:



Gambar 1. Tahapan KDD

3.3. Selection

Pada tahap ini dilakukan pemilihan atribut yang relevan untuk analisis klustering, yaitu lokasi pekerjaan, jenis pekerjaan, dan durasi izin kerja. Atribut pendukung seperti judul pekerjaan dipertahankan hanya untuk interpretasi hasil kluster. Data yang tidak relevan terhadap tujuan penelitian dihilangkan.

3.4. Preprocessing

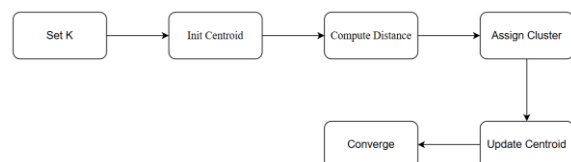
Tahap ini mencakup pembersihan data dari duplikasi, penyatuan perpanjangan izin yang masih terkait pekerjaan yang sama, serta standarisasi format data. Pada tahap ini juga dilakukan pengelompokan kategori lokasi menjadi dua klasifikasi utama, yaitu *Field* dan *Office*, untuk mengurangi kompleksitas fitur kategorikal dan mencegah peningkatan dimensi berlebihan.

3.5. Transformation

Seluruh data kategorikal ditransformasi menggunakan *one-hot encoding*, sementara durasi izin kerja dinormalisasi menggunakan *Min-Max normalization*. Hasil dari tahap ini berupa matriks fitur numerik yang seragam, sehingga sesuai untuk diproses menggunakan algoritma *K-Means*.

3.6. Data Mining (K-Means Clustering)

Tahap inti penelitian dilakukan dengan menerapkan algoritma *K-Means Clustering*. Dalam implementasinya, algoritma ini memiliki beberapa langkah utama [20], yaitu:



Gambar 2. Tahapan K-Means

- Menentukan jumlah kluster (*K*) menggunakan Elbow Method, kemudian memvalidasinya menggunakan *Silhouette Score*, *Davies–Bouldin Index* (DBI), dan *Calinski–Harabasz* (CH). *Silhouette* digunakan sebagai metrik utama, DBI sebagai penentu ketika nilai *Silhouette* berdekatan, dan CH sebagai indikator pendukung.
- Meminisialisasi *centroid* awal secara acak sesuai nilai *K*.
- Menghitung jarak setiap data terhadap *centroid* menggunakan Euclidean Distance dengan persamaan berikut:

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + \dots (n_i - n_j)^2} \quad (1)$$

Keterangan:

D_{ij} = Jarak antara *i* dan *j*

$X_i X_i$ = Koordinat *x* objek

$X_j X_j$ = Koordinat *x* pusat

$Y_i Y_i$ = Koordinat *y* objek

$Y_j Y_j$ = Koordinat *y* pusat

- Mengelompokkan data ke dalam kluster berdasarkan jarak terdekat.
- Memperbarui posisi *centroid* berdasarkan rata-rata anggota kluster dengan persamaan berikut:

$$C_i = \frac{x_1 + x_2 + x_3 + \dots + x_n}{\sum x} \quad (2)$$

Keterangan:

$x_1 x_1$ = nilai record data ke-1

$x_2 x_2$ = nilai record data ke-2

$\sum x \sum x$ = Jumlah record data

- f. Mengulangi proses hingga model mencapai konvergensi.

Hasil dari proses ini berupa pengelompokan pekerjaan berdasarkan kemiripan karakteristik izin kerja.

3.7. Evaluation

Evaluasi model dilakukan secara kuantitatif menggunakan *Silhouette Score* (semakin mendekati 1 semakin baik), DBI (semakin rendah semakin baik), serta CH Index sebagai indikator pendukung. Evaluasi kualitatif dilakukan dengan meninjau kesesuaian setiap kluster terhadap karakteristik pekerjaan yang tercatat dalam data SIKA, sebagai dasar pertimbangan dalam konteks K3.

3.8. Knowledge

Tahap akhir berupa interpretasi pola yang muncul dari hasil kluster. Setiap kluster dianalisis untuk melihat karakteristik pekerjaannya, seperti kecenderungan durasi izin, jenis pekerjaan dominan, atau lokasi tertentu. Hasil interpretasi kemudian digunakan sebagai dasar pertimbangan awal dalam

perencanaan peningkatan pengendalian K3 di lingkungan PT Pertamina Hulu Rokan Zona 4 Field Limau.

4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian yang diperoleh berdasarkan metode yang telah dijelaskan pada bagian Metode Penelitian. Hasil disajikan mulai dari deskripsi data akhir, penentuan jumlah kluster optimal, hasil klusterisasi, evaluasi, hingga interpretasi dan rekomendasi. Penyajian ini bertujuan untuk memberikan gambaran jelas mengenai karakteristik data serta pola pengelompokan yang ditemukan melalui algoritma K-Means.

4.1. Deskripsi Data Akhir

Data yang digunakan untuk klusterisasi K-Means terdiri dari 4.453 entri data dan 173 fitur numerik (hasil dari *one-hot encoding* dan normalisasi *Min-Max*). Atribut utama meliputi lokasi, jenis pekerjaan, dan durasi izin kerja (dalam hari). Semua variabel kategorikal telah dikonversi menjadi numerik melalui *one-hot encoding*, sedangkan variabel numerik dinormalisasi menggunakan *Min-Max* agar skala atribut seimbang. Dengan struktur ini, data siap digunakan untuk pemodelan K-Means. Berikut ini adalah contoh sampel data yang digunakan sebelum dilakukannya proses transformation:

Tabel 1. Data akhir sebelum ditransformation

Lokasi	Jenis Pekerjaan	Judul Pekerjaan	Durasi Izin (Hari)	Kategori Lokasi
AR-058	Dingin	Pemasangan Alat Produksi Smart Lift	14	Field
BEL-001	Dingin	Monitoring Lingkungan (Air, Udara, Emisi)	7	Field
BEL-001	Dingin	Pm 5 Genset 63412 Sn 7db01622	7	Field
....				
YARD TL27 LIMAU	Ketinggian	Pekerjaan Pengelasan Blink Plank Separator	7	Field
YARD TL27 LIMAU	Panas	Fabrikasi Pemotongan Header Spu Line 8 Inch & 6 Inch	7	Field
YARD TL27 LIMAU	Panas	Pemotongan Pipa Existing Untuk Pompa Centripugal, Genset, Dan Panel Di Sumur Sp Karangan 12	7	Field

4.2. Penentuan Jumlah Kluster

Berdasarkan evaluasi metrik *Silhouette*, DBI, CH, dan Inertia (Elbow), diperoleh nilai terbaik untuk K=2. Tabel 2 menampilkan ringkasan metrik K-Means untuk berbagai nilai K:

Tabel 2. Hasil Evaluasi Metrik K-Means untuk Berbagai Nilai K

K	Inertia (Elbow)	Silhouette	DBI	CH
2	1083234,5903	0,9306	0,4438	8824,7151
3	455160,2219	0,8823	0,3942	13568,8702
4	243443,8018	0,7930	0,4417	18198,8210
5	144399,4574	0,7908	0,4115	23768,6702
6	104838,4239	0,7908	0,4314	26519,9976

Dari Tabel 2 terlihat bahwa nilai *Silhouette* tertinggi terdapat pada K=2 (0,93), menunjukkan

pemisahan kluster paling tegas. Meskipun DBI terendah ada di K=3, peningkatan kualitas kluster tidak signifikan dibanding *Silhouette*. Nilai CH meningkat seiring bertambahnya K, tetapi tidak cukup untuk menyalip indikator utama, sehingga K=2 dipilih sebagai jumlah kluster optimal. Temuan ini menegaskan bahwa struktur data SIKA cenderung membentuk dua kelompok besar yang stabil secara statistik, sehingga K=2 menjadi konfigurasi yang paling representatif untuk menangkap pola izin kerja.

4.3. Hasil Klusterisasi K-Means

Proses klusterisasi dengan K=2 menghasilkan:

Tabel 3. Ukuran Kluster

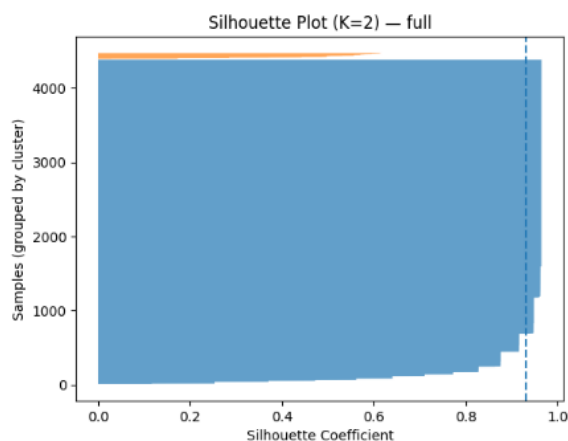
Cluster	Size
0	4376
1	77

Dari tabel 3 tersebut dapat dilihat bahwa distribusi sangat tidak seimbang dimana C0 mendominasi (pekerjaan rutin/seragam), sedangkan C1 adalah kelompok jarang tetapi berbeda jelas (rare-but-distinct). Meskipun jumlah anggota klaster tidak seimbang, hal ini mencerminkan pola izin kerja yang umum terjadi di lapangan, di mana sebagian besar aktivitas bersifat rutin dan hanya sebagian kecil memperlihatkan karakteristik izin yang berbeda atau lebih intensif.

Tabel 4. Top 5 fitur pembeda per klaster (rata-rata *centroid*)

Cluster	Feature	Mean_value
0	Durasi Izin (hari)	11.984232175503353
0	KatLok_Field	0.9718921389396706
0	Jenis_Dingin	0.8473491773308929
0	Jenis_Panas	0.09117915904936111
0	Lokasi_SP BELIMBING	0.06489945155392876
1	Durasi Izin (hari)	180.45454545454578
1	KatLok_Field	1.0
1	Jenis_Dingin	0.9999999999999998
1	Durasi_norm	0.43003565062388605
1	Lokasi_SP II	0.11688311688311713

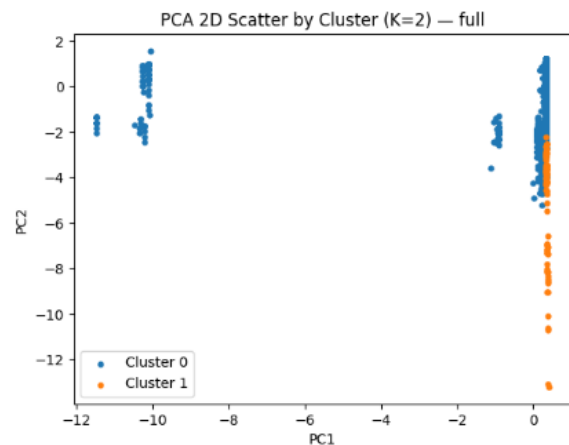
Berdasarkan Tabel 4, *Cluster 0* (C0/mayoritas) ditandai oleh durasi izin yang relatif pendek, dominasi kategori Field, serta jenis pekerjaan yang banyak berupa pekerjaan dingin, sehingga mencerminkan pola pekerjaan rutin dan stabil. Sementara itu, *Cluster 1* (C1/minoritas) memiliki durasi izin yang sangat panjang, seluruhnya berada di kategori Field, serta didominasi oleh pekerjaan teknis intensif. Pola ini menunjukkan karakter izin kerja yang lebih intensif dan dapat dijadikan dasar pertimbangan dalam konteks pengendalian K3.



Gambar 3. *Silhouette Plot*

Silhouette Plot pada gambar 3 menunjukkan bahwa hampir seluruh sampel memiliki koefisien *silhouette* positif (banyak yang tinggi). Garis putus-putus menandai rata-rata keseluruhan. Ini menegaskan bahwa klaster kompak & terpisah pada K=2 sehingga pola pengelompokan yang terbentuk cukup jelas dan

dapat digunakan untuk interpretasi lanjut terhadap karakteristik izin kerja.



Gambar 4. PCA 2D scatter

Visualisasi PCA pada gambar 4 memperlihatkan dua klaster jelas dimana C0 menyebar dominan di area luas, sedangkan C1 terkonsentrasi di zona tertentu. Pola ini mendukung interpretasi karakteristik klaster berdasarkan durasi dan jenis pekerjaan. Pola pemisahan ini menunjukkan bahwa pekerjaan berdurasi panjang dengan karakter teknis intensif benar-benar membentuk gugus tersendiri, yang secara praktis dapat dijadikan dasar pertimbangan sebagai kelompok pekerjaan yang memerlukan pengelolaan K3 lebih terstruktur.

4.4. Evaluation

Setelah proses klasterisasi dilakukan, langkah selanjutnya adalah menilai kualitas hasil pengelompokan untuk memastikan bahwa klaster yang terbentuk benar-benar mencerminkan pola izin kerja pada data SIKA. Evaluasi dilakukan secara kuantitatif dan kualitatif.

Tabel 5. Rata-rata *Silhouette* per klaster

Cluster	<i>Silhouette</i>
0	0.9387648424369408
1	0.46828995072447266

4.5. Kuantitatif

Hasil evaluasi menunjukkan bahwa nilai *silhouette overall* $\approx 0,93$, *DBI* $\approx 0,44$, *CH* ≈ 8.800 , dan *SSE* $\approx 1,08 \times 10^6$, konsisten dengan pemilihan $K = 2$ sebagai jumlah klaster optimal. Nilai rata-rata *silhouette* pada Tabel 5 juga menunjukkan bahwa kedua klaster memiliki koefisien positif. Klaster 0 memiliki nilai *silhouette* tinggi dan stabil karena anggotanya lebih banyak serta homogen. Sementara itu, Klaster 1 tetap memiliki nilai positif meskipun lebih rendah, yang menunjukkan bahwa pemisahan klaster tetap memadai bagi kelompok pekerjaan langka namun memiliki pola yang konsisten.

4.6. Kualitatif

Berdasarkan Tabel 4, C1 memiliki durasi izin sangat tinggi (± 180 hari), dan seluruhnya di area Field. Sementara itu, C0 berisi pekerjaan rutin dengan durasi pendek (± 12 hari) dan Sebagian besar merupakan pekerjaan kategori dingin. Temuan ini sejalan dengan *silhouette* plot (Gambar 3) & PCA 2D (Gambar 4) yang menunjukkan pemisahan pola izin kerja yang konsisten antara kedua klaster.

cluster	Jumlah Pekerjaan	freq	Lokasi (modul)	Jenis Pekerjaan (modul)	Rata-rata Durasi (hari)	Median Durasi (hari)
1	SONOLOG DYNAGRAPH	7	SP BELIMBING	Dingin	111.0	112.0
1	SONOLOG DYNAGRAPH	9	SP BELIMBING	Dingin	276.91	280.0
1	SERVICE PENAMBAHAN OLI PELUMAS ENGINE	3	SP BELIMBING	Dingin	235.67	238.0
1	SERVICE PENAMBAHAN OLI PELUMAS ENGINE	3	SP BELIMBING	Dingin	106.0	106.0
9	SAMPLING LINGKUNGAN OKTOBER	42	BEL-001	Dingin	7.0	7.0
9	SAMPLING LINGKUNGAN BULAN APRIL	58	BEL-001	Dingin	7.63	7.0
9	SAMPLING LINGKUNGAN APRIL	52	BEL-001	Dingin	7.0	7.0
9	R/UWPF	25	BEL-001	Dingin	10.86	7.0
9	R/UWPF	24	BEL-000	Dingin	10.5	7.0
1	PENGISIAN CHEMICAL DAN SERVICE PRODUCTS DAN M1	5	BOOSTER NIRU	Dingin	225.4	182.0

Gambar 5. Hasil akhir klaster

Pada gambar 5 menampilkan 10 contoh data dari masing-masing klaster. Pada Klaster 0, pekerjaan yang muncul cenderung periodik dan berulang, seperti Sampling Lingkungan Bulan April/Oktober dan R/UWPF, dengan durasi izin pendek dan lokasi dominan di Field (BEL-001, BEL-011, BEL-030). Pola ini menggambarkan karakter pekerjaan rutin dan pelaksanaan operasional standar.

Sebaliknya, Klaster 1 berisi pekerjaan dengan durasi izin jauh lebih panjang seperti Sonolog Dynagraph, Service Penambahan Oli Pelumas Engine, dan aktivitas chemical/produksi, yang seluruhnya berlangsung di lokasi Field seperti SP Belimbing dan Booster Niru. Pekerjaan dalam klaster ini bahwa pola izin kerja dalam klaster tersebut berbeda dari C0 sehingga dapat dijadikan dasar pertimbangan untuk interpretasi pengelolaan K3.

Dengan demikian, Evaluasi kuantitatif dan kualitatif sama-sama menunjukkan bahwa K-Means dengan dua klaster menghasilkan segmentasi yang stabil, terpisah dengan jelas, dan sesuai dengan karakteristik operasi di lapangan. Hasil ini mendukung pemanfaatan klaster sebagai sebagai dasar pertimbangan dalam pengelolaan K3, khususnya untuk merencanakan strategi pengawasan yang sesuai dengan variasi pola izin kerja yang lebih terarah.

4.7. Knowledge (Interpretasi dan Rekomendasi)

Tahap akhir dari proses analisis adalah menarik pengetahuan dari pola klaster yang terbentuk. Interpretasi ini digunakan untuk memahami karakteristik setiap kelompok pekerjaan serta menyusun rekomendasi yang relevan bagi penguatan pengendalian K3. Hasilnya didapatkan sebagai berikut:

a. Cluster 0 – Pekerjaan Rutin

Berisikan aktivitas administratif atau inspeksi ringan. Rekomendasi yang dapat dipertimbangkan adalah pengawasan rutin, pelatihan dasar, evaluasi izin kerja periodik.

b. Cluster 1 – Pekerjaan Teknis Intensif

Berisikan aktivitas teknis lapangan kompleks, durasi panjang. Rekomendasi yang dapat dipertimbangkan adalah analisis HIRADC untuk

pekerjaan kompleks, pengaturan durasi izin, inspeksi lapangan intensif, review personel dan shift.

K-Means dengan K=2 memberikan pemisahan klaster yang jelas dan konsisten, sehingga karakteristik pola izin kerja dapat digunakan untuk pengambilan keputusan berbasis data terkait pengendalian risiko dan perencanaan K3.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, metode K-Means klastering terbukti efektif dalam mengidentifikasi pola izin kerja pada data SIKA PT Pertamina Hulu Rokan Zona 4 Field Limau tahun 2023–2025, di mana atribut lokasi, jenis pekerjaan, dan durasi izin mampu merepresentasikan karakteristik pekerjaan yang berkaitan dengan potensi risiko operasional. Setelah melalui preprocessing menggunakan *one-hot encoding* dan normalisasi *Min-Max*, algoritma *K-Means* mampu mengolah data secara stabil dan menghasilkan nilai evaluasi yang baik, ditunjukkan oleh *Silhouette Score* tinggi (0,9306) dan *Davies-Bouldin Index* rendah (0,4437). Dua klaster yang terbentuk menggambarkan pola pekerjaan yang dapat dimanfaatkan untuk mendukung strategi pengawasan K3, yaitu klaster pekerjaan rutin berdurasi pendek dan klaster pekerjaan teknis intensif berdurasi panjang. Meski demikian, penelitian ini masih memiliki keterbatasan karena hanya menggunakan tiga atribut utama, sehingga disarankan untuk menambahkan variabel lain seperti tingkat bahaya area kerja, jumlah personel, atau riwayat insiden, serta menguji metode klastering lain seperti DBSCAN atau Hierarchical Clustering guna meningkatkan ketepatan pola dan konsistensi struktur klaster. Dengan demikian, penelitian ini memberikan dasar awal bagi penguatan manajemen K3 sekaligus membuka peluang pengembangan model analitik yang lebih komprehensif pada studi selanjutnya.

DAFTAR PUSTAKA

- [1] A. I. Maharani, A. H. Aziza, A. F. Lubis, and Y. T. Zaharani, "Manajemen risiko industri minyak bumi dan gas pada proses industri dan manajemen risiko," *Environment Conflict*, vol. 1, no. 1, Feb. 2024, doi: 10.61511/enviroinc.v1i1.2024.525.
- [2] F. Khan, I. Alsaduni, O. Khan, and M. Parvez, "Enhancing safety and risk management in biodiesel production: a machine learning approach using k-means clustering to address operational challenges and standardization issues," *Results in Engineering*, vol. 26, Jun. 2025, doi: 10.1016/j.rineng.2025.105175.
- [3] Kemnaker, "Kasus Kecelakaan Kerja Tahun 2024," Kementerian Ketenagakerjaan Republik Indonesia. Accessed: Sep. 08, 2025. [Online]. Available: <https://satudata.kemnaker.go.id/data/kumpulan-data/2447>

- [4] SKK Migas, *Statistik Minyak dan Gas Bumi*. 2023.
- [5] A. Rahmatika, "PENERAPAN METODE K-MEANS CLUSTERING UNTUK ANALISIS PROFIL LULUSAN TEKNIK INDUSTRI DI DUNIA KERJA," Universitas Islam Indonesia, Yogyakarta, 2024.
- [6] R. Mubarak, A. A. Syahputra, A. T. Permana, L. Sholiah, and Tarwoto, "Implementasi Data Mining untuk Clustering Lowongan Pekerjaan Menggunakan Metode Algoritma K-Means," *Jurnal JTik (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 9, no. 2, pp. 703–712, 2025.
- [7] P. Abela, R. Buaton, and M. Simanjuntak, "Penerapan Metode Clustering Pada Kasus Kecelakaan Kerja," *Merkurius : Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 2, no. 5, pp. 200–210, 2024, doi: 10.61132/merkurius.v2i4.
- [8] V. Irawan, A. Rizal, and I. Purnamasari, "Penerapan Algoritma K-Mean Clustering Pada Kecelakaan Lalu Lintas Berdasarkan Kabupaten/Kota Di Provinsi Jawa Tengah Tahun 2020," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 10, pp. 293–300, 2020, doi: 10.5281/zenodo.6820090.
- [9] Y. Budiman, N. B. Wirotomo, and M. W. Prasetyo, "Implementation as Prevention Effort from Fire Occurred in PT Pertamina (Persero) Refinery Unit VI Balongan," *Jurnal Syntax Admiration*, vol. 5, no. 5, May 2024.
- [10] B. T. Wizaksono and M. Waty, "ANALISIS RISIKO KERJA PROYEK PEMBANGUNAN IT MANDIRI BUMI SLEPI," *JMTS: Jurnal Mitra Teknik Sipil*, vol. 6, no. 4, pp. 885–898, Nov. 2023.
- [11] N. F. Khairani, R. Buaton, and M. P. U. Sitompul, "Pengelompokan Penanganan Resiko Pada Kegiatan Panen Berdasarkan Alat Pelindung Diri Yang digunakan (Studi Kasus: PT. Langkat Nusantara Kepong Kebun Padang Brahrang)," *Modem : Jurnal Informatika dan Sains Teknologi.*, vol. 2, no. 4, pp. 68–91, Sep. 2024, doi: 10.62951/modem.v2i4.232.
- [12] T. Kurniawan and M. Jajuli, "Clustering Data Kecelakaan Lalu Lintas di Kecamatan Cileungsi Menggunakan Metode K-Means," *Generation Journal*, vol. 6, no. 1, 2022.
- [13] K. Deswiasqa, E. Darmawan, and Sugiyarto, "Application of K-Means for Clustering Based on the Severity of COVID-19 in Indonesian Private Hospitals," *EKSAKTA: Journal of Sciences and Data Analysis*, vol. 3, no. 2, Sep. 2022, doi: 10.20885/eksakta.vol3.iss2.art5.
- [14] C. Syalom, U. Khaira, and D. Arsa, "CLUSTERING DAERAH RAWAN STUNTING DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS," *Djtechno: Jurnal Teknologi Informasi*, vol. 6, no. 1, pp. 389–401, Apr. 2025, doi: 10.46576/djtechno.
- [15] D. Abdullah, S. Susilo, A. S. Ahmar, R. Rusli, and R. Hidayat, "The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data," *Qual Quant*, vol. 56, no. 3, pp. 1283–1291, Jun. 2022, doi: 10.1007/s11135-021-01176-w.
- [16] Salsabila, M. Hafiyusholeh, H. Khaulasari, D. R. Candra, and D. Yuniarti, "IMPLEMENTASI K-MEANS CLUSTERING DALAM PEMETAAN WILAYAH RAWAN PENYAKIT LEPTOSPIROSIS DI JAWA TIMUR," *Jurnal Riset dan Aplikasi Matematika (JRAM)*, vol. 08, no. 02, pp. 189–200, 2024.
- [17] N. Dwitianti, S. A. Kumala, and S. D. Handayani, "Comparative Study of Earthquake Clustering in Indonesia Using K-Medoids, K-Means, DBSCAN, Fuzzy C-Means and K-AP Algorithms," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 6, pp. 768–778, Dec. 2024, doi: 10.29207/resti.v8i6.5514.
- [18] A. Gearhart *et al.*, "Identifying high-risk Fontan phenotypes using K-means clustering of cardiac magnetic resonance-based dyssynchrony metrics," *Journal of Cardiovascular Magnetic Resonance*, vol. 26, no. 2, Dec. 2024, doi: 10.1016/j.jocmr.2024.101060.
- [19] F. R. Dewantoro, "Analisis Risiko Kerentanan Pekerja Informal di Indonesia Tahun 2022," *Ecodemica: Jurnal Ekonomi, Manajemen dan Bisnis*, vol. 8, no. 1, 2024.
- [20] N. A. Maori and Evanita, "METODE ELBOW DALAM OPTIMASI JUMLAH CLUSTER PADA K-MEANS CLUSTERING," *Jurnal SIMETRIS*, vol. 14, no. 2, pp. 277–287, 2023