

## PENGELOMPOKAN DATA KONDISI MESIN *SCREW PRESS* MENGUNAKAN ALGORITMA *FUZZY C-MEANS*

M. Faiz Al Fiqri<sup>1)</sup>, Jasril<sup>2)</sup>, Suwanto Sanjaya<sup>3)</sup>, Lestari Handayani<sup>4)</sup>, Fitri Insani<sup>5)</sup>

<sup>1) 2) 3) 4) 5)</sup> Teknik Informatika, Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau

email: [12150113184@students.uin-suska.ac.id](mailto:12150113184@students.uin-suska.ac.id)<sup>1)</sup>, [jasril@uin-suska.ac.id](mailto:jasril@uin-suska.ac.id)<sup>2)</sup>, [suwantosanjaya@uin-suska.ac.id](mailto:suwantosanjaya@uin-suska.ac.id)<sup>3)</sup>, [lestari.handayani@uin-suska.ac.id](mailto:lestari.handayani@uin-suska.ac.id)<sup>4)</sup>, [fitri.insani@uin-suska.ac.id](mailto:fitri.insani@uin-suska.ac.id)<sup>5)</sup>

### INFO ARTIKEL

#### Riwayat Artikel:

Diterima Mei, 2025

Revisi Mei, 2025

Terbit Mei, 2025

### ABSTRAK

Kinerja mesin *screw press* sangat berpengaruh terhadap efisiensi dan kualitas produksi minyak kelapa sawit. Salah satu komponen penting dalam sistem ini adalah *Back Pressure Vessel (BPV)*, yang berfungsi menyalurkan uap ke berbagai stasiun proses. Penelitian ini bertujuan untuk mengelompokkan kondisi mesin berdasarkan atribut temperatur dan tekanan menggunakan algoritma *Fuzzy C-Means (FCM)*. Data yang digunakan berasal dari mesin *BPV* PT. XYZ pada periode April–Mei 2024, dengan total 23.002 entri. Tahapan penelitian meliputi seleksi data, prapemrosesan, normalisasi menggunakan *Min-Max Scaler*, klusterisasi dengan *FCM*, serta evaluasi menggunakan metode *Elbow* dan *Davies-Bouldin Index (DBI)*. Hasil awal menunjukkan adanya tiga kluster dengan distribusi kondisi mesin dari stabil hingga memerlukan perawatan. Metode *Elbow* mengindikasikan jumlah kluster optimal sebanyak empat, sedangkan *DBI* menunjukkan dua kluster terbaik dengan nilai 0,389. Temuan ini menunjukkan bahwa *FCM* mampu mengelompokkan kondisi mesin secara efektif dan dapat dijadikan dasar pengambilan keputusan perawatan. Pengembangan lebih lanjut disarankan dengan menambahkan atribut lain.

#### Kata Kunci :

*Fuzzy C-Means; Clustering; Back Pressure Vessel; Screw Press; Davies-Bouldin Index*

### ABSTRACT

The performance of the screw press machine significantly affects the efficiency and quality of palm oil production. One of the critical components in this system is the *Back Pressure Vessel (BPV)*, which distributes steam to various processing stations. This study aims to classify machine conditions based on temperature and pressure attributes using the *Fuzzy C-Means (FCM)* algorithm. The dataset was collected from the *BPV* machine at PT. XYZ during the period of April–May 2024, comprising 23,002 entries. The research stages include data selection, preprocessing, normalization using the *Min-Max Scaler*, clustering with *FCM*, and evaluation using the *Elbow Method* and the *Davies-Bouldin Index (DBI)*. Preliminary results identified three clusters ranging from stable to maintenance-required conditions. The *Elbow Method* suggested an optimal number of four clusters, while the *DBI* indicated two optimal clusters with the best score of 0.389. These findings demonstrate that *FCM* effectively groups machine conditions and can serve as a foundation for maintenance decision-making. Further development is recommended by incorporating additional attributes.

#### Keywords:

*Fuzzy C-Means; Clustering; Back Pressure Vessel; Screw Press; Davies-Bouldin Index*

### Penulis Korespondensi:

Jasril

Teknik Informatika, Sains dan  
Teknologi, Universitas Islam Negeri  
Sultan Syarif Kasim Riau

Email:

[jasril@uin-suska.ac.id](mailto:jasril@uin-suska.ac.id)

## 1. PENDAHULUAN

Pada era sekarang produksi minyak di dunia industri sangat membutuhkan kinerja dari mesin *press*. Mesin merupakan fasilitas yang memainkan peran krusial dalam kelancaran proses produksi. Dengan penggunaan dan pemeliharaan yang tepat, mesin tersebut dapat meningkatkan efisiensi dan efektivitas dalam setiap tahap produksi [1]. Dengan meningkatnya kapasitas produksi, pemanfaatan kinerja mesin produksi juga mengalami peningkatan. Namun, meskipun kapasitas penjualan semakin tinggi, kinerja mesin *press* di bagian produksi justru menunjukkan penurunan [2]. Dalam proses produksi, salah satu masalah yang sering dihadapi adalah *downtime* pada mesin *screw press*. Ketika mesin mengalami *downtime*, hal ini dapat mengganggu jalannya produksi dan bahkan memicu penghentian operasi di stasiun kerja berikutnya. Oleh karena itu, untuk mencegah terjadinya *downtime*, perusahaan perlu menerapkan sistem perawatan yang efektif guna meminimalkan risiko kerusakan pada mesin [3].

Mesin *screw press* juga sering mengalami masalah *oil losses*. *Oil losses* adalah kehilangan jumlah minyak yang seharusnya diperoleh dari proses produksi dapat menjadi masalah yang signifikan. Angka kehilangan minyak sawit mencerminkan jumlah minyak yang tidak berhasil dipanen selama proses pengolahan. Sebagian dari minyak yang terbuang ini mungkin akan digunakan sebagai bahan bakar untuk *boiler*, khususnya minyak yang berasal dari serat. Seperti yang terjadi pada PT.XYZ kehilangan minyak pada mesin *sterilizer* di pabrik XYZ mencapai 76,84%, sementara mesin *screw press* mencatat kehilangan sebesar 10,44%. Angka-angka ini jauh melebihi standar kehilangan minyak yang berlaku secara umum di pabrik kelapa sawit. Tingginya tingkat kehilangan minyak di pabrik XYZ disebabkan oleh kurang optimalnya kinerja *sterilizer* dan *screw press*. Selain itu, kurangnya perangkat pendukung seperti pembaca tekanan dan kesalahan dalam penentuan waktu perebusan juga berkontribusi pada masalah ini [4]. Selain itu mesin *screw press* adalah salah satu komponen dalam proses ini yang mengalami tingkat kehilangan minyak yang signifikan. Di PT. Y, kehilangan minyak tercatat sebesar 6,28 hingga 6,32 persen dari total produksi [5]. Untuk menjaga kinerja mesin ampas *press*, perusahaan perlu melakukan perawatan rutin dan kontrol kualitas TBS. Guna mendukung hal ini, dibutuhkan analisis data yang dapat mengelompokkan kondisi mesin secara sistematis. Salah satu pendekatan yang dapat digunakan adalah *clustering*, metode *unsupervised learning* dalam *data mining*.

*Clustering* adalah salah satu metode dalam *Data Mining* yang bersifat tanpa arahan (*unsupervised learning*). Terdapat dua jenis *clustering* data yang umum digunakan dalam proses pengelompokan, yaitu *clustering* data hierarki dan *clustering* data non-hierarki [6]. Dalam konteks *data mining*, *clustering* merupakan suatu teknik yang digunakan untuk menganalisis data. Tujuan utama dari *clustering* adalah mempartisi data menjadi kelompok-kelompok yang bermakna [7]. Proses *clustering* mengelompokkan data berdasarkan kemiripan antar elemen, sementara elemen yang berbeda ditempatkan di klaster terpisah. Metode ini banyak digunakan di berbagai bidang, seperti penelitian dan pemetaan wilayah. Salah satu metode non-hierarki yang populer adalah *Fuzzy C-Means*, karena mampu membentuk klaster secara fleksibel berdasarkan tingkat kemiripan data.

*Fuzzy C-Means* adalah algoritma yang digunakan untuk menentukan jumlah klaster optimal, dengan mengandalkan derajat keanggotaan untuk menilai keterkaitan data terhadap klaster. Metode ini memungkinkan pembentukan klaster yang fleksibel dan terukur [8]. *Fuzzy C-Means* merupakan metode pengelompokan data yang menggunakan derajat keanggotaan yang bervariasi antara 0 hingga 1 [9]. Pada metode *fuzzy c-means* semakin tinggi nilai derajat keanggotaan, semakin besar pula kemiripan suatu data dengan kelompok yang ada [10]. Jika kita mengasumsikan bahwa kita ingin mengelompokkan  $n$  objek data ke dalam  $c$  kelompok atau klaster.

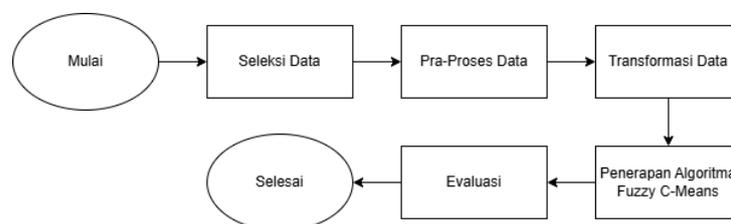
Pada penelitian ini, penulis menggunakan metode *fuzzy c-means* yang memiliki keunggulan dibandingkan dengan metode *clustering* lainnya yaitu kemampuannya dalam menempatkan pusat klaster dengan lebih tepat dibandingkan dengan metode *clustering* lainnya [11]. Pada penelitian lainnya dengan studi kasus *clustering* untuk *recredesiating* fasilitas kesehatan dengan penerapan metode *fuzzy c-means* (FCM) melalui aplikasi *matlab* menghasilkan dua klaster, yakni klaster pertama sebanyak 479 dan klaster kedua sebanyak 580. Pada tahap pelatihan klasterisasi menggunakan metode FCM, diperoleh hasil akurasi yang menunjukkan nilai *Partition Coefficient Index* (PCI) sebesar 0.50002 dan *Partition Entropy Index* (PEI) sebesar 0.99998. Hal ini menunjukkan bahwa tingkat akurasi nilai keanggotaan dalam klaster tersebut tergolong cukup baik [12].

Berdasarkan penjelasan-penjelasan dan masalah di atas maka, tujuan yang didapat dari penelitian ini berupa kluster terbaik yang akan didapatkan ketika menggunakan metode *fuzzy c-means* dengan menggunakan data asli dari PT.XYZ dari bulan April 2024 – Mei 2024.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Pada penelitian ini terdapat beberapa tahapan penelitian yang akan dilakukan. Tahapan-tahapan yang digunakan penulis dilakukan dengan menganalisis dan pengolahan data secara relevan. Tahapan pertama penulis melakukan seleksi data untuk mendapatkan data apa saja yang akan diolah, selanjutnya penulis melakukan pra-proses data dimana data akan dibersihkan agar bisa dilakukan transformasi data, lalu setelah dilakukan pra-proses data akan masuk ke dalam tahapan transformasi data, pada tahapan transformasi data akan dilakukan normalisasi data. Tahapan selanjutnya adalah klustering dengan menggunakan algoritma *fuzzy c-means*. Tahapan terakhir adalah evaluasi yang bertujuan untuk mencari berapa kluster terbaik saat menggunakan algoritma yang diterapkan. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian.

### 2.2 Seleksi Data

Tahapan ini penulis akan menyeleksi data yang akan diolah nantinya, agar data yang digunakan sesuai dengan kebutuhan penelitian ini, seperti variabel data dan jenis data yang akan digunakan nanti harus disesuaikan dengan kebutuhan penelitian.

### 2.3 Pra-Proses Data

Tahapan pra-proses ini dilakukan untuk membersihkan data yang tidak digunakan. Contohnya data akan dihapus jika adanya duplikat data, data kosong atau *null*, dan akan memperbaiki kesalahan ketikan pada data ataupun memeriksa dan memperbaiki data-data yang inkonsisten.

### 2.4 Transformasi Data

Tahapan transformasi ini bertujuan agar data yang digunakan nantinya sudah bisa diolah dengan algoritma *fuzzy c-means*. Pada tahapan ini akan dilakukan normalisasi data untuk kualitas data bisa tetap *valid* nantinya. Proses normalisasi adalah proses dimana data akan ditransformasi menjadi bentuk yang seragam, teratur, dan simpel untuk dianalisis, dengan maksud mengurangi pengulangan dan ketidakselarasan. Dalam konteks basis data relasional, proses normalisasi juga bertujuan mengurangi pengulangan data dan memperkuat integritas data. Pada penelitian ini akan menggunakan metode *MinMaxScaler* untuk tahapan normalisasi data. *MinMaxScaler* merupakan salah satu metode *scaling* dan memiliki rentang [0, 1] [13]. *MinMaxScaler* memiliki persamaan (1).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Dengan  $x'$  adalah data yang telah dinormalisasi,  $x$  adalah data yang asli,  $x_{max}$  adalah nilai tertinggi pada atribut,  $x_{min}$  adalah nilai terendah pada atribut,  $min_R$  adalah rentang minimum, dan  $max_R$  adalah rentang maksimum.

## 2.5 Penerapan Algoritma Fuzzy C-Means

*Fuzzy C-Means* merupakan metode yang digunakan untuk menentukan kluster secara optimal. Dalam metode ini, penentuan apakah suatu data termasuk ke dalam kluster tertentu didasarkan pada derajat keanggotaan. Berbeda dari pendekatan klustering konvensional yang bersifat tegas, *fuzzy c-means* memungkinkan pembagian data ke dalam beberapa kluster secara fleksibel dan terukur [8]. Langkah-langkah algoritma *fuzzy c-means* adalah:

- A. Hitung jumlah derajat keanggotaan

$$\mu_{ik} = \frac{1}{\sum_{j=1}^c \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (2)$$

Keterangan:

- $X_i$  = Data ke-i
- $C_k$  = Pusat kluster ke-k
- $C_k$  = Parameter *fuzziness*

- B. Lakukan perhitungan untuk mendapatkan pusat kluster ke-k dengan menggunakan persamaan (3)

$$V_{kj} = \frac{\sum_{i=1}^n (\mu_{ik})^w x_{ij}}{\sum_{i=1}^n (\mu_{ik})^w} \quad (3)$$

Keterangan:

- $V_{kj}$  = Nilai pusat (*centroid*) kluster ke-k pada atribut ke-j
- $\mu_{ik}$  = Derajat keanggotaan data ke-i terhadap kluster ke-k (nilai antara 0 dan 1)
- $x_{ij}$  = Nilai dari data ke-i pada atribut ke-j
- $w$  = Tingkat *fuzziness* (biasanya  $w = m$ , dengan nilai umum 2)
- $n$  = Jumlah total data
- $k$  = Indeks kluster
- $j$  = Indeks atribut

Fungsi objektif pada iterasi ke- $t$  dengan persamaan (4)

$$P_t = \sum_{i=1}^n \sum_{k=1}^c \left( \sum_{j=1}^m (X_{ij} - V_{kj})^2 (\mu_{ik})^w \right) \quad (4)$$

Keterangan:

- $P_t$  = Total fungsi objektif (semacam total error atau *loss function*)
- $c$  = Jumlah kluster
- $m$  = Jumlah atribut

- C. Selanjutnya, penting untuk memastikan kondisi penghentian. Terdapat beberapa kriteria yang digunakan untuk menentukan kapan proses perhitungan algoritma *Fuzzy C-Means* harus dihentikan, yaitu:

1. Jika  $(|P - P_{t-1}| < \epsilon)$  atau  $(t > \max Iter)$
2. Jika tidak  $t = t + 1$ , ulang kembali proses perhitungan proses kluster

## 2.6 Evaluasi

Pada penelitian ini akan dilakukan evaluasi menggunakan 2 metode, yaitu metode *elbow* dan metode *davies-bouldin index (DBI)*.

### A. Metode *Elbow*

Metode *Elbow* adalah teknik untuk menentukan jumlah kluster optimal dengan melihat grafik perubahan nilai *WCSS*. Titik optimal ditandai dengan sudut tajam (siku) pada grafik, yang menunjukkan penurunan *WCSS* paling signifikan saat jumlah kluster bertambah. Selisih *WCSS* antar kluster juga dapat dihitung secara numerik sebagai bahan analisis [14]. Dalam metode *elbow*, grafik digunakan untuk membandingkan variasi jumlah kluster berdasarkan nilai *WCSS*. Titik optimal ditandai dengan sudut tajam pada grafik, yang menunjukkan penurunan *WCSS* paling signifikan saat jumlah kluster bertambah. Selisih nilai *WCSS* antar kluster juga dapat dihitung secara numerik sebagai acuan evaluasi.

$$WCSS = \sum_{k=1}^k \sum_{x_i \in S_k} (X_i - C_k)^2 \quad (5)$$

Dimana  $C_k$  adalah atribut titik pusat kluster ke-  $i$ , sedangkan  $X_i$  adalah atribut data ke-  $i$ .

### B. Metode *Davies-Bouldin Index (DBI)*

*Davies-Bouldin Index (DBI)*, yang diperkenalkan oleh David L. Davies dan Donald W. Bouldin, adalah metode evaluasi kluster internal yang menilai kualitas kluster berdasarkan kedekatan dan keterpisahan antar data. Semakin rendah nilai *DBI*, maka semakin baik kualitas kluster yang terbentuk. Indeks ini tidak memerlukan data *eksternal*, karena penilaiannya hanya berdasarkan struktur kluster itu sendiri [15]. *Davies-Bouldin Index (DBI)* mengevaluasi kualitas kluster berdasarkan kedekatan data dalam kluster (kohesi) dan jarak antar kluster (separasi). Nilai *DBI* yang lebih kecil menunjukkan kluster yang lebih baik karena data dalam kluster lebih rapat dan antar kluster lebih terpisah [16].

Langkah-langkah dari *davies-bouldin index* sebagai berikut:

1. *Sum of Squares Within-Cluster (SSW)* digunakan untuk mengukur kohesi kluster, yaitu seberapa dekat data dalam kluster dengan pusat klusternya (*centroid*). Nilai *SSW* menunjukkan total deviasi atau penyimpangan data dari *centroid* di dalam kluster tersebut. Adapun rumus perhitungan *Sum of Squares Within-Cluster (SSW)* dapat dinyatakan sebagai berikut:

$$ssw = \frac{1}{m} \sum_{j=i}^{m_i} d(x_j, C_i) \quad (6)$$

Keterangan:

$m_i$  = Jumlah data dalam cluster ke- $i$

$C_i$  = *Centroid* cluster ke- $i$

$d(x_j, C_i)$  = Jarak setiap data ke *centroid*  $i$  yang dihitung menggunakan jarak *euclidean*

2. *Sum of Squares Between-Cluster (SSB)* mengukur tingkat separasi dengan melihat jarak antar pusat kluster (*centroid*) relatif terhadap pusat data keseluruhan. Nilai *SSB* yang lebih besar menunjukkan pemisahan kluster yang lebih baik. Adapun rumus yang digunakan untuk menghitung *Sum of Squares Between-Cluster (SSB)* adalah sebagai berikut:

$$SSB_{i,j} = d(C_i, C_j) \quad (7)$$

Dimana,  $d(C_i, C_j)$  adalah jarak antar cluster.

3. Selanjutnya, dihitung rasio antara nilai *Sum of Squares Within-Cluster (SSW)* dan *Sum of Squares Between-Cluster (SSB)* untuk menilai kualitas kluster. Kluster yang baik biasanya memiliki kohesi rendah (*SSW* kecil) dan separasi tinggi (*SSB* besar). Rasio ini digunakan untuk membandingkan antar kluster, khususnya antara kluster ke-*i* dan ke-*j*, dengan indeks *i*, *j*, dan *k* menunjukkan jumlah total kluster yang ada. Misalnya, jika ada tiga kluster, maka terdapat tiga indeks yang mewakili masing-masing kluster. Persamaan yang digunakan untuk menghitung rasio tersebut dinyatakan sebagai berikut:

$$R_{i,j} = \frac{ssw_i + ssw_j}{SSB_{i,j}} \quad (8)$$

Keterangan:

$ssw_i$  = *Sum of square within cluster pada centroid i.*

$SSB_{i,j}$  = *Sum of square between cluster data ke i dengan j pada cluster yang berbeda.*

4. Tahap akhir evaluasi adalah menghitung *Davies-Bouldin Index (DBI)* dengan mengambil rata-rata dari nilai rasio tertinggi antara setiap kluster dan kluster lainnya. Nilai *DBI* ini digunakan untuk menilai sejauh mana kluster memiliki kohesi internal yang baik dan pemisahan yang jelas antar kluster. Adapun rumus perhitungan *Davies-Bouldin Index* dapat dinyatakan sebagai berikut:

$$DBI = \frac{1}{k} \sum_i^k \max(R_{i,j}) \quad (9)$$

*Davies-Bouldin Index (DBI)* merupakan metrik evaluasi internal yang digunakan untuk mengukur kualitas klusterisasi. Parameter *k* menunjukkan jumlah kluster yang digunakan dalam proses pengelompokan data. Nilai *DBI* bersifat non-negatif ( $\geq 0$ ), di mana semakin kecil nilainya, maka semakin baik kualitas kluster yang terbentuk. Hal ini mengindikasikan bahwa kluster memiliki kohesi yang tinggi dan separasi yang jelas. Sebaliknya, nilai *DBI* yang besar menunjukkan bahwa hasil klusterisasi kurang optimal. Oleh karena itu, evaluasi menggunakan *DBI* sangat penting untuk menilai efektivitas algoritma dalam membentuk kelompok data yang representatif.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Seleksi Data

Penelitian ini akan mengolah data asli mesin *screw press* dengan kode device *BPV* dari bulan April – Mei 2024 yang diperoleh dari PT.XYZ, dengan total data sebanyak 23002 data. Berikut adalah data mesin yang akan digunakan dapat dilihat pada Tabel 1.

**Tabel 1.** Data Mesin *Screw Press* Kode *BPV*.

No	Kode Device	Temperature	Pressure
1	N00E-EP01-BPV-01	123	3,19
2	N00E-EP01-BPV-01	123	3
3	N00E-EP01-BPV-01	123	2,95
4	N00E-EP01-BPV-01	123	2,77
...	...	...	...
22997	N00E-EP04-BPV-01	0	2,78
22998	N00E-EP04-BPV-01	0	2,99
22999	N00E-EP05-BPV-01	100	1,62
23000	N00E-EP05-BPV-01	100	1,53
23001	N00E-EP06-BPV-01	100	2,12
23002	N00E-EP06-BPV-01	100	2,03

Setelah data diinput, selanjutnya data akan diambil 2 atributnya saja sebagai data latih yaitu *temperature* dan *pressure*, karena atribut berikut yang akan digunakan untuk diolah nantinya. Berikut hasil data setelah diambil data latih pada Tabel 2.

**Tabel 2.** Data Latih.

No	Temperature	Pressure
1	123	3,19
2	123	3
3	123	2,95
4	123	2,77
...	...	...
22997	0	2,78
22998	0	2,99
22999	100	1,62
23000	100	1,53
23001	100	2,12
23002	100	2,03

Setelah proses seleksi data, data yang digunakan dalam penelitian ini disajikan dalam Tabel 2. Penelitian ini berfokus pada dua atribut terpilih, karena dianggap paling relevan dan memiliki pengaruh signifikan terhadap tujuan analisis. Pemilihan atribut ini didasarkan pada pertimbangan integritas data, kontribusinya terhadap variabel target, dan kesesuaiannya dengan konteks masalah yang diteliti.

### 3.2 Pra-proses Data

Tahapan ini dilakukan untuk mencari dan membersihkan data yang tidak digunakan, contohnya data akan dihapus jika adanya duplikat data, data kosong atau *null*, dan akan memperbaiki kesalahan ketikan pada data ataupun memeriksa dan memperbaiki data-data yang inkonsisten. Hasil pengecekan data dapat dilihat pada Tabel 3.

**Tabel 3.** Data Setelah Dilakukan Pengecekan.

No	Atribut	Jumlah	Data Null	Tipe Data
1	Temperature	23002	Not-Null	Float64
2	Pressure	23002	Not-Null	Float64

Setelah dilakukan pengecekan data pada tahapan pra-proses, maka hasil yang didapat seperti yang tertera pada Tabel 3, tidak terdapat nilai *null* dalam data yang digunakan, dan memiliki tipe data yang seragam yaitu *float64*. Bisa disimpulkan bahwa data yang digunakan sudah siap untuk diolah ke tahapan selanjutnya.

### 3.3 Transformasi Data

Transformasi data dilakukan untuk mengatasi perbedaan skala antara atribut *temperature* (0–157) dan *pressure* (0–4,43). Tanpa normalisasi, atribut *temperature* akan mendominasi perhitungan jarak pada algoritma *Fuzzy C-Means*. Oleh karena itu, digunakan metode *Min-Max Scaler* untuk menyetarakan skala atribut ke rentang 0 hingga 1, sehingga kontribusi kedua atribut dalam pembentukan klaster menjadi seimbang. Data setelah transformasi dapat dilihat pada Tabel 4.

**Tabel 4.** Data Setelah Ditransformasi.

No	Temperature	Pressure
1	0,783	0,720
2	0,783	0,677
3	0,783	0,666
4	0,783	0,625
...	...	...
22997	0	0,627
22998	0	0,675
22999	0,637	0,365
23000	0,637	0,345
23001	0,637	0,478
23002	0,637	0,458

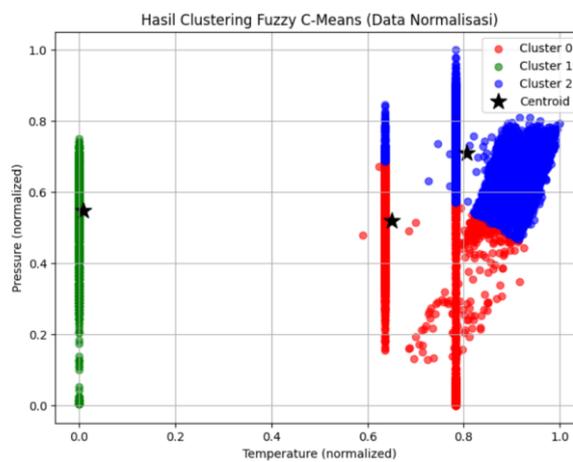
### 3.4 Algoritma *Fuzzy C-Means*

Proses klusterisasi menggunakan algoritma *Fuzzy C-Means* telah dilakukan terhadap total 23.002 data yang telah melalui tahap transformasi dan normalisasi. Pengujian dilakukan dengan jumlah cluster sebanyak 3. Proses *clustering* ini dilakukan dengan menggunakan rumus (2) – (4) dan juga menggunakan data pada Tabel 4.

Tabel 5. Pengujian dengan Jumlah 3 Cluster.

Kluster	Jumlah Data
<i>C0</i>	11603
<i>C1</i>	4473
<i>C2</i>	6926

Hasil data setelah dilakukan pengujian menggunakan algoritma *fuzzy c-means* maka didapatkan jumlah data tiap kluster seperti pada Tabel 5. Jumlah data pada *C0* sebanyak 11603 data, jumlah data pada *C1* sebanyak 4473 data, dan jumlah data pada *C2* sebanyak 6926 data. Tahap selanjutnya penulis akan menampilkan visualisasi dengan pusat kluster pada *temperature* dan *pressure* adalah ([0.65068467 0.51844509]) pada *C0*, ([0.00749823 0.54785929]) pada *C1*, dan ([0.80565425 0.70952575]) pada *C2*.



Gambar 2. Visualisasi *Clustering* dengan Pusat Cluster.

Setelah dilakukan visualisasi seperti pada Gambar 2., maka dapat disimpulkan bahwa ada kluster yang tergolong aman dikarenakan suhu yang tinggi tetapi tekanannya tetap stabil seperti *C1*, sedangkan cluster yang lain memiliki suhu yang tinggi dan tekanan yang tinggi. Mesin yang memiliki suhu tinggi dan tekanan yang tinggi tergolong ke dalam mesin yang butuh perawatan lebih lanjut seperti *C2*.

### 3.5 Evaluasi Menggunakan *Elbow* dan *Davies-Bouldin Index (DBI)*

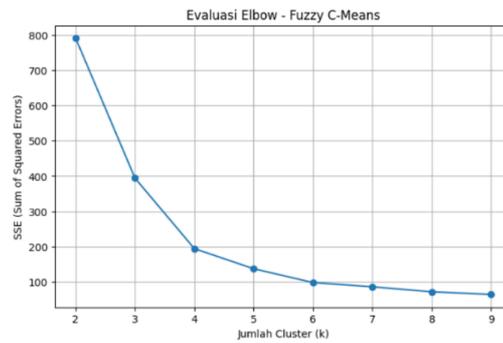
Tahap berikutnya adalah evaluasi hasil klusterisasi menggunakan *Elbow Method* dan *Davies-Bouldin Index (DBI)* untuk menilai kualitas dan menentukan jumlah kluster optimal. *DBI* mengukur pemisahan antar kluster dan kerapatan data dalam kluster, dengan nilai yang lebih kecil menunjukkan hasil yang lebih baik. *Elbow Method* melihat grafik fungsi objektif terhadap jumlah kluster, di mana titik siku menandai jumlah kluster optimal saat penambahan kluster tidak lagi menurunkan nilai fungsi secara signifikan.

#### A. *Elbow*

Evaluasi kualitas klusterisasi pertama dilakukan menggunakan metode *elbow*. Metode ini menggunakan rumus (5). Metode ini mengamati nilai *Within-Cluster Sum of Square (WCSS)* terhadap jumlah kluster yang ditentukan dan melihat grafik titik siku (*elbow*).

**Tabel 6.** Pengujian dengan Jumlah 3 Cluster.

Kluster	Nilai <i>WCSS</i>
2	790.2550872497438
3	393.90774838099287
4	194.3609335790021
5	137.38957620056715
6	98.30835112194879
7	86.26156896920403
8	72.10431511222495
9	64.78501850155637



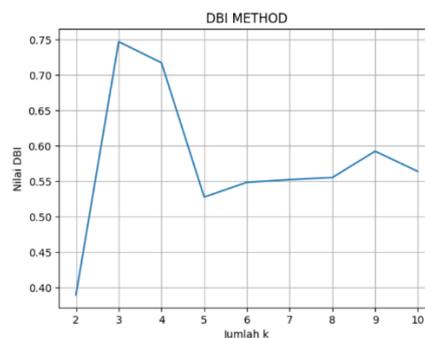
**Gambar 3.** Grafik Elbow.

Evaluasi jumlah kluster optimal dilakukan menggunakan metode *Elbow* seperti pada Tabel 6., dan Gambar 3., dengan mengamati nilai *Within-Cluster Sum of Square (WCSS)* terhadap jumlah kluster dari 2 hingga 9. Hasil pengujian menunjukkan bahwa nilai *WCSS* mengalami penurunan drastis dari kluster ke-2 hingga kluster ke-4, yaitu dari 790,26 menjadi 194,36. Setelah jumlah kluster melebihi angka 4, penurunan *WCSS* mulai melambat dan cenderung mendatar. Hal ini menunjukkan adanya titik siku (*elbow*) yang jelas pada jumlah kluster 4, sehingga dapat disimpulkan bahwa jumlah kluster optimal berdasarkan *Elbow Method* adalah sebanyak 4 kluster.

- B. Selain menggunakan *Elbow Method*, evaluasi kualitas klusterisasi juga dilakukan dengan menggunakan *Davies-Bouldin Index (DBI)*. Metode ini menilai seberapa baik kluster terbentuk berdasarkan tingkat kedekatan antar anggota dalam satu kluster dan keterpisahan antar kluster. Semakin kecil nilai DBI, maka semakin baik hasil klusterisasi.

**Tabel 7.** Nilai DBI.

Kluster	Nilai <i>WCSS</i>
2	0.3893238543791762
3	0.7468037921320506
4	0.7170579455321158
5	0.5274655011165283
6	0.5482702744824498
7	0.5520772107222089
8	0.5550781030972746
9	0.5921466046227795



**Gambar 4.** Grafik DBI.

Berdasarkan hasil perhitungan dan grafik yang tertera pada Tabel 7., dan Gambar 4., nilai *DBI* terendah diperoleh saat jumlah kluster (*k*) adalah 2, yaitu sebesar 0.389. Hal ini menunjukkan bahwa secara matematis, pembentukan kluster paling optimal terjadi ketika jumlah kluster adalah 2, karena kluster yang terbentuk memiliki tingkat kekompakan yang tinggi dan terpisah dengan baik satu sama lain. Namun demikian, pada saat jumlah kluster 3 dan 4, nilai *DBI* meningkat hingga 0.74 dan 0.71, menandakan penurunan kualitas pemisahan kluster.

### 3.6 Perbandingan Hasil Kluster Terbaik

Setelah dilakukan evaluasi, maka hasil kluster terbaik yang didapatkan dari metode *elbow* berjumlah 4 kluster sedangkan metode *DBI* berjumlah 2 kluster. Hasil pengujian 4 kluster dapat dilihat pada Tabel 8., dan untuk pengujian 2 kluster dapat dilihat pada Tabel 9.

Tabel 8. Pengujian 4-cluster.

Cluster	Jumlah Data	Interval Temperature	Interval Pressure
<i>C0</i>	11213	0.589172 - 0.925159	0.288939 - 0.742664
<i>C1</i>	5947	0.636943 - 0.917197	0 - 0.347630
<i>C2</i>	4473	0.636943 - 1	0.474041 - 1
<i>C3</i>	1369	0	0.004515 - 0.749436

Tabel 9. Pengujian 2-cluster.

Cluster	Jumlah Data	Interval Temperature	Interval Pressure
<i>C0</i>	11603	0.589172 - 1	0 - 1
<i>C1</i>	4473	0	0.004515 - 0.749436

## 4. KESIMPULAN

Penelitian ini membahas penerapan algoritma *Fuzzy C-Means* dalam pengelompokan data kondisi mesin *screw press* berdasarkan dua atribut penting, yaitu *temperature* dan *pressure*. Data yang digunakan merupakan data asli dari PT. XYZ pada periode April hingga Mei 2024, dengan jumlah total 23.002 data. Proses penelitian dimulai dari tahap seleksi data, pra-proses, transformasi dengan normalisasi *Min-Max Scaler*, hingga tahap klasterisasi menggunakan algoritma *FCM*. Hasil klasterisasi awal dengan jumlah cluster 3 menunjukkan bahwa data dapat dikelompokkan menjadi tiga kondisi utama mesin: mesin dalam kondisi aman, mesin dengan tekanan stabil namun *temperature* tinggi, dan mesin dengan tekanan serta *temperature* tinggi yang memerlukan perhatian lebih. Evaluasi menggunakan metode *Elbow* menunjukkan bahwa jumlah cluster optimal secara visual berada pada angka 4, sedangkan *Davies-Bouldin Index* menyarankan jumlah cluster optimal adalah 2 berdasarkan nilai *DBI* terendah (0.389). Hal ini menunjukkan adanya pertimbangan *trade-off* antara pemisahan cluster yang baik dan kejelasan visual klasterisasi. Penelitian ini menunjukkan bahwa algoritma *FCM* mampu mengelompokkan kondisi mesin dengan cukup baik dan dapat dijadikan acuan dalam pengambilan keputusan perawatan mesin secara lebih tepat dan preventif. Keterbatasan penelitian ini adalah hanya mempertimbangkan dua atribut (*temperature* dan *pressure*), sehingga untuk penelitian selanjutnya disarankan agar atribut tambahan seperti getaran atau usia mesin juga turut dimasukkan agar hasil klasterisasi lebih komprehensif.

## DAFTAR PUSTAKA

- [1] S. Yasir and A. Saputra, "Analisa Reliability Dan Availability Mesin Screw Press Kelapa Sawit (Studi Kasus di PT. Ujong Neubok Dalam)," *Unistek*, vol. 9, no. 2, pp. 83–94, Sep. 2022, [Online]. Available: <http://ejournal.unis.ac.id/index.php/UNISTEK>
- [2] T. Rahmanadi and F. Yuamita, "Analisis Kinerja Produksi Pada Mesin Screw Press di Pabrik Pengolahan Sawit: Studi Kasus PT. Agrindo Indah Persada," *JURITEK*, vol. 2, no. 2, pp. 73–2, Jun. 2022.
- [3] Demi Ramadian, Yoga Pratama, and Rizaldi Sardani, "Analisis Efektivitas Mesin Screw Press Menggunakan Metode Overall Equipment Effectiveness pada PT Bintara Tani Nusantara," *Go-Integratif: Jurnal Teknik Sistem dan Industri*, vol. 3, no. 01, pp. 36–46, May 2022, doi: 10.35261/gijtsi.v3i01.6530.
- [4] J. A. Saputri, N. Hidayat, M. Riduan, Y. P. Ayu, and A. Kahar, "Kajian Neraca Massa dan Penentuan Kehilangan Minyak (Oil losses) pada Bak Kondensat Unit Perebusan Alat Screw Pressing Pabrik Kelapa

- Sawit,” *Jurnal Chemurgy*, vol. 7, no. 1, pp. 33–39, Jun. 2023, [Online]. Available: <http://e-journals.unmul.ac.id/index.php/TK>
- [5] M. Basuki, I. Pamungkas, and T. Tamalika, “Identifikasi Penyebab Kehilangan Crude Palm Oil (CPO) di Pabrik Kelapa Sawit,” *Jurnal Optimalisasi*, vol. 09, no. 01, pp. 40–44, Apr. 2023.
- [6] N. Karolina, “Data Mining Pengelompokan Pasien Rawat Inap Peserta BPJS Menggunakan Metode Clustering (Studi Kasus: RSUD.Bangkalan),” *JOURNAL OF INFORMATION AND TECHNOLOGY UNIMOR*, vol. 1, no. 2, pp. 47–53, Sep. 2021, [Online]. Available: [www.kaputama.ac.id](http://www.kaputama.ac.id)
- [7] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, “Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means,” *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, vol. 10, no. 2, pp. 52–61, Oct. 2022, doi: 10.21063/jtif.2022.v10.2.52-61.
- [8] N. Nurdiana, A. Nilogiri, and M. Rahman, “Penerapan Algoritma Fuzzy C-Means dan Metode Elbow untuk Mengelompokkan Provinsi di Indonesia Berdasarkan Indeks Demokrasi Indonesia,” *Jurnal Smart Teknologi*, vol. 3, no. 5, pp. 544–551, Jul. 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>
- [9] H. S. Firdaus, A. L. Nugraha, B. Sasmito, M. Awaluddin, and C. A. Nanda, “Perbandingan Metode Fuzzy C-Means Dan K-Means Untuk Pemetaan Daerah Rawan Kriminalitas Di Kota Semarang,” *Jurnal Geodesi dan Geomatika*, vol. 04, no. 01, pp. 58–64, Nov. 2021.
- [10] R. Siringoringo, “Peningkatan Performa Cluster Fuzzy C-Means Pada Pengklasteran Sentimen Menggunakan Particle Swarm Optimization,” *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 6, no. 4, pp. 349–354, May 2018, doi: 10.25126/jtiik.2018561090.
- [11] Novianti Fahriza, Yasmin Yasirah R A, and Novitasari Dian C R, “Penerapan Algoritma Fuzzy C-Means (FCM) dalam Pengelompokan Provinsi di Indonesia berdasarkan Indikator Penyakit Menular Manusia,” *JUMANJI*, vol. 6, no. 1, pp. 23–33, May 2022, [Online]. Available: <https://pusdatin.kemkes.go.id>
- [12] V. Herlinda, D. Darwis, and Dartono, “Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means,” *Jurnal Teknologi dan Sistem Informasi (JTSI)*, vol. 2, no. 2, pp. 94–99, Jun. 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [13] W. Purba, G. A. Sembiring, A. Saputra, T. Turnip, B. Jua, and I. Manihuruk, “Penerapan Data Mining Untuk Pengolahan Data Rekam Medis Menggunakan Metode K-Means Clustering Pada Rumah Sakit Royal Prima Medan,” *Jurnal TEKINKOM*, vol. 6, no. 1, pp. 158–168, Jun. 2023, doi: 10.37600/tekinkom.v6i1.857.
- [14] Ahmad Rofiki Said, Deni Arifianto, and Habibatul Aziza Al Faruq, “Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Tanaman Pangan Dengan Algoritma Fuzzy C- Means Dan Metode Elbow,” Jun. 2020. Accessed: May 28, 2025. [Online]. Available: <http://repository.unmuhjember.ac.id/id/eprint/5069>
- [15] F. T. Permadi, A. Nilogiri, and U. A. Rosyidah, “Algoritma Fuzzy C-Means dan Metode Davies Bouldin Index (DBI) untuk Mengelompokkan Kabupaten/Kota di Jawa Timur Berdasarkan Sanitasi Layak, Air Minum Layak, dan Rumah Layak Huni,” *Jurnal Smart Teknologi*, vol. 4, no. 4, pp. 450–459, May 2023, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST>
- [16] M. Sholeh and K. Aeni, “Perbandingan Evaluasi Metode Davies Bouldin, Elbow Dan Silhouette Pada Model Clustering Dengan Menggunakan Algoritma K Means,” *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 8, no. 1, pp. 56–65, Jun. 2023, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Trave>