BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Penerapan Algoritma *K-Means Clustering* pada Kinerja Mesin *Screw press*

Fikri Kurnia Rahman*, Jasril, Suwanto Sanjaya, Lestari Handayani, Fitri Insani

Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim, Pekanbaru, Indonesia
Email: [1]12150114585@students.uin-suska.ac.id, [2,*]jasril@uin-suska.ac.id, , [3]suwantosanjaya@uin-suska.ac.id
[4]lestari.handayani@uin-suska.ac.id, [5]fitri.insani@uin-suska.ac.id
Corresponding Author: jasril@uin-suska.ac.id

Abstrak-*Screw press* adalah salah satu mesin yang berperan sebagai proses pemisahan minyak dari tangki yang berisi Tandan Buah Segar (TBS). Bagian dari mesin *screw press* terdiri dari screw ganda yang berfungsi sebagai alat mengeluarkan cairan minyak dari alat penekan, dan diberikan tekanan balik yang berasal dari hydraulic double cone. Ampas buah yang diaduk ditekan sehingga minyak yang terkandung dalam ampas buah keluar akibat tekanan dari mesin press. Pemeliharaan dan perbaikan mesin merupakan kegiatan yang penting dalam menunjang kegiatan produktif suatu sektor. Oleh karena itu, diperlukan pembelajaran untuk menemukan pola kondisi mesin pada pabrik. Salah satu cara untuk menemukan pola kondisi mesin yaitu dengan menggunakan teknik *Clustering*. *Clustering* adalah teknik pengelompokan data berdasarkan parameter tertentu sehingga membentuk kelas-kelas objek yang mempunyai karakteristik yang sama. Pada penelitian ini data diperoleh dari PT. XYZ periode April 2024 – Mei 2024 dengan jumlah data sebanyak 23002. Analisis dilakukan dengan menerapkan algoritma *K-Means Clustering* dengan melakukan pengujian terhadap 3-15 *cluster*. Berdasarkan evaluasi hasil *clustering* dengan menggunakan algoritma *K-Means Clustering* terhadapa metode Davies Bouldin Index (DBI) diperoleh hasil *cluster* yang paling optimal pada 3 *cluster* dingan nilai DBI terkecil 0,386. Sedangkan dengan menggunakan metode *Elbow* diperoleh hasil *cluster* paling optimal dilihat dengan titik siku pada 4 *cluster* dengan nilai *Sum of Square Error* (WCSS) 270. Oleh karena itu, dapat disimmpulkan bahwa hasil *cluster* dengan menerapkan algoritma *K-Means Clustering* sudah relavan dalam menentukan pola kondisi mesin dan diharapkan dapat membantu dalam menentukan pola kondisi mesin *screw press*.

Kata Kunci: *Screw press*; *K-Means*; *Clustering*; *Davies-Bouldin Index*; *Elbow*

Abstract- The *screw press* is one of the machines used in the process of separating oil from tanks containing Fresh Fruit Bunches (FFB). The machine consists of a twin-screw system that functions to extract oil from the pressing unit, with back pressure applied by a hydraulic double cone. The mixed fruit residue is compreWCSSd, causing the oil contained within the residue to be released due to the pressure exerted by the press machine. Maintenance and repair of machinery are eWCSSntial activities to support productive operations in any sector. Therefore, it is necessary to conduct analysis to identify patterns in machine conditions within the factory. One effective approach to discovering machine condition patterns is through *clustering* techniques. *Clustering* is a method of grouping data based on certain parameters to form *cluster*s of objects that share similar characteristics. In this study, data were collected from PT. XYZ for the period of April 2024 to May 2024, with a total of 23,002 records. The analysis was conducted using the *K-Means Clustering* algorithm, with testing carried out on 3 to 15 *cluster*s. Based on the evaluation using the *Davies-Bouldin Index* (DBI), the most optimal *clustering* result was obtained with 3 *cluster*s, achieving the lowest DBI value of 0.386. Meanwhile, using the *Elbow* Method, the optimal number of *cluster*s was determined to be 4, as indicated by the *Elbow* point on the WCSS graph, with a *Sum of Square Error* (WCSS) value of 270. Therefore, it can be concluded that the *clustering* results using the *K-Means Clustering* algorithm are relevant for identifying machine condition patterns and are expected to assist in monitoring and managing the condition of the *screw press* machine.

Keywords: *Screw press*; *K-Means*; *Clustering*; *Davies-Bouldin Index*; *Elbow*

1. PENDAHULUAN

Pengolahan minyak kelapa sawit merupakan proses pengubahan bahan Tandan Buah Segar (TBS) menjadi minyak kelapa sawit mentah (Crude Palm Oil atau CPO) dan inti dari kelapa sawit (Palm Kernel). Pengolahan minyak kelapa sawit sangat bergantung pada kinerja mesin yang digunakan. Jika terjadi kerusakan pada salah satu mesin, maka proses pengolahan minyak kelapa sawit dapat terhenti sehingga menimbulkan kerugian besar bagi pabrik[1]. Salah satu mesin yang terlibat dalam produksi minyak kelapa sawit adalah mesin *screw press*. *Screw press* merupakan alat yang berperan penting dalam proses pemisahan minyak kelapa sawit dari tangki yang berisi TBS. Mesin ini terdiri dari *screw* ganda yang berfungsi sebagai alat pengeluaran cairan minyak dari alat pengepres. Proses ini dilengkapi dengan tekanan balik yang dihasilkan oleh *hydraulic double cone*, dimana ampas buah diaduk dan ditekan, sehingga minyak dalam ampas buah dapat keluar dikarenakan tekanan dari mesin press[2].

Kerusakan pada stasiun press menyebabkan kesulitan produksi karena jumlah mesin yang digunakan berkurang. Oleh karena itu, pemeliharaan dan perawatan mesin menjadi aktivitas yang sangat penting dalam kelancaran dan keberlangsungan produktivitas pada pabrik. Tanpa dilakukannya pemeliharaan uang teratur, resiko terjadinya kerusakan pada mesin akan meningkat yang dapat berdampak pada operasional dalam pengolahan minyak kelapa sawit dalam mencapai target prouksi[3]. Dampak negatif terhadap pemeliharaan dan perawatan mesin yang tidak teratur menyebabkan kerusakan mesin yang fatal, kegagalan dalam pemenuhan target produksi, berkurangnya waktu produksi, dan mahalnya biaya perbaikan dari mesin yang sudah mengalami kerusakan. Sebaliknya jika mesin dipelihara dan dirawat dengan baik dapat mencegah terjadinya kerusakan pada mesin dan dapat memperpanjang masa pakai sehingga target target produksi dapat terpenuhi[4].

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Pada mesin *Screw Press* terdapat beberapa komponen penting, salah satunya adalah *Back Pressure Vessel* (BPV). BPV merupakan suatu alat yang bekerja sebagai bejana tekan yang berfungsi untuk menyimpan dan menyalurkan uap. Peran ini sangatlah penting pada proses pengolahan minyak kelapa sawit, karena kestabilan tekanan uap menjadi faktor utama dalam keberhasilan proses pemisahan dan pemurnian minyak kelapa sawit. BPV menyalurkan uap ke berbagai stasiun penting, diantaranya stasiun perebusan, stasiun pengepresan, stasiun pemisahan dan stasiun pengolahan benih. Uap yang disalurkan oleh BPV umumnya menyalurkan uap rendah dengan tekanan maksimal 3,5 kg/cm³[5]. Dengan demikian, BPV berperan sebagai penjaga keberlangsungan pasokan uap tiap proses dalam Pabrik Kelapa Sawit (PKS). Uap tersebut berasal dari *boiler* atau bejana tekan yang bertugas untuk memanaskan air dan menghasilkan uap, kemudian diubah menjadi listrik oleh turbin, lalu uap yang tersisa akan di alirkan ke mesin BPV. Selanjutnya, uap diarahkan ke berbagai stasiun yang membutuhkannya. Oleh karena itu, uap yang dihasilkan *boiler* harus memenuhi kebutuhan uap seluruh proses yang dibutuhkan untuk pengolahan PKS[6].

Sementara itu, dalam konteks pengolahan data industri, Data mining merupakan proses penggalian wawasan yang tersembunyi dari sekumpulan data besar dan kompleks. Data mining memiliki tujuan untuk mengetahui pola hubungan, atau informasi yang tidak terlihat dalam data sehingga memberikan pengetahuan yang sangat berguna dalam pengambilan keputusan[7]. Metode yang paling sering dilakukan pada kegiatan data mining adalah *clustering*, khususnya dengan *K-Means*. Metode ini dilakukan dengan melakukan pengelompokan data menjadi beberapa kelompok (*cluster*) berbeda berdasarkan karakteristik data yang serupa[8]. *Clustering* adalah kegiatan pengelompokkan dan pengamatan dalam beberapa *cluster* objek berdasarkan tingkat kesamaan yang tinggi satu sama lainnya, namun berbeda dengan *cluster* lain. Tujuan dari *clustering* adalah memngelompokkan data menjadi *cluster-cluster* yang seragam (homogen), sehingga kemiripan antar data dalam satu *cluster* maksimal, sementara kemiripan antar data dari *cluster* berbeda diminimalkan[9]. Berdasarkan hasil penelitian terkait penerapan algoritma *K-Means* pada *clustering* nilai akhir karya akhir mahasiswa, pengujian dilakukan pada berbagai ukuran data (100 hingga 500 data) dan variasi jumlah *cluster* (2, 3, 4, dan 5). Hasilnya menunjukkan bahwa algoritma ini mampu mengelompokkan data secara optimal berdasarkan kesamaan karakteristik. Dengan demikian, *K-Means* sangat tepat digunakan untuk analisis data numerik, khususnya dalam pengelompokan nilai akhir mahasiswa.[10]. Penelitian ini menerapkan algoritma *K-Means Clustering* terhadap kondisi mesin *screw press* karena dinilai mampu dalam melakukan pengelompokkan data berdasarkan kemiripan karakteristik pada data. Metode *K-Means* digunakan dalam mengidentifikasi pola kondisi mesin *Screw Press* berdasarkan kinerja mesin BPV, sehingga setiap kondisi berdasarkan parameter tertentu dapat dikelompokkankedalam suatu kelompok yang sama. Hal ini juga didukung dengan penelitian-penelitian terkait yang mengatakan bahwa algoritma tersebut lebih unggul dan sangat tepat digunakan untuk analisis pola kondisi mesin industri berdasarkan parameter tertentu.

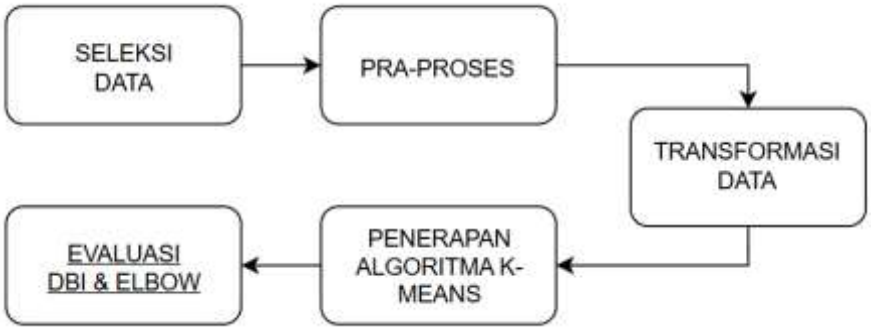
Penelitian lainnya mengatakan bahwa algoritma *K-Means* memiliki kinerja lebih cepat. Hal ini terlihat dari waktu eksekusi (*elapsed time*) hanya memerlukan 0.433755 detik, lebih singkat dibandingkan FCM yang memerlukan waktu 0.781679 detik[11]. Hasil pengklasteran dengan metode *K-Means* menunjukkan bahwa produksi tertinggi terjadi pada bulan September, dengan total 8.474.600 item di *Cluster* 1 dan 25.106.750 item di *Cluster* 2. Sebaliknya, bulan Oktober mencatat produksi terendah, yaitu 7.420.733 item di *Cluster* 1 dan 2.939.250 item di *Cluster* 2.[12] Hasil analisis menunjukkan bahwa algoritma *K-Means* berhasil mengelompokkan 99 item sebagai produk laris dan 23 item sebagai kurang laku. Informasi ini dapat digunakan pemilik usaha untuk menetapkan prioritas pembelian ulang berdasarkan tingkat penjualan.[13]. Penelitian lain membuktikan bahwa algoritma *K-Means* efektif dalam menganalisis data calon penerima PKH. Proses klasterisasi mempermudah pengelompokan tingkat kelayakan secara objektif dan sistematis, sehingga penyaluran bantuan menjadi lebih tepat sasaran.[14]. Penelitian lain menunjukkan bahwa penerapan algoritma K-Means dengan evaluasi Davies-Bouldin Index (DBI) efektif untuk menentukan jumlah cluster optimal. Klasterisasi berhasil mengelompokkan provinsi berdasarkan jumlah industri desa ke dalam tiga kategori: rendah, sedang, dan tinggi.[15].

Oleh karena itu penulis mengusulkan sebuah penilitan dengan judul "Penerapan Algoritma *K-Means Clustering* pada kondisi mesin *Screw press*". Dengan data yang diperoleh merupakan data real-time dari bulan April 2024 – Mei 2024. Diharapkan penelitian ini dapat membantu pihak PT.XYZ dalam menemukan pola kondisi mesin *Screw press.*

2. METODOLOGI PENELITIAN

Metodologi penelitian pada data mining adalah proses yang meliputi tahapan-tahapan yang tersusun untuk menganalisis dan mengolah data dengan relavan. Pada tahap ini penulis proses data mining dimulai dari seleksi data dengan menggunakan data yang ingin diolah saja, selanjutnya melakukan pra-proses data dimana data dibersihkan dan dipersiapkan untuk tahap berikutnya, lalu setelah dilakukan pra-proses maka data akan berubah yang mana pada tahap ini disebut dengan transformasi data, selanjutnya menerapkan algoritma *K-Means* pada atribut data yang telah dilakukan tranfsormasi data untuk mengidentifikasi karakteristik serupa yang tersembunyi dalam data, dan langkah terakhir yaitu mencari jumlah klaster terbaik dengan melakukan berbagai teknik evaluasi untuk memastikan keakuratan dari algoritma yang sudah diterapkan.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT



Gambar 1. Alur Metodologi penelitian

2.1 Seleksi data

Seleksi adalah langkah awal yang penting dilakukan sebelum memasuki tahap eksplorasi informasi tersembunyi dari sekumpulan data set. Data ynag diseleksi harus relavan dan konsisten dengan tujuan analisis, karena akan digunakan dalam proses penggalian data untuk mengidentifikasi pola, hubungan, atau informasi tersembunyi yang bernilai[16]. Seleksi data merupakan fase awal dari proses *Knowledge Discovery in Database* (KDD) yang diambil dari kumpulan data yang besar. Tahap ini bertujuan sebagai tahap dipastikannya data agar siap untuk dianalisis dengan hasil yang lebih akurat[17].

2.2 Pra-proses data

Setelah memilih data, selanjutnya data akan dibersihkan dan dicek guna untuk meningkatkan kualitas data yang akan diolah. Proses pembersihan meliputi penanganan nilai kosong (*missing values*), penghapusan baris data yang tidak relevan, serta eliminasi noise atau *outlier*. Langkah awal ini dapat melibatkan penerapan metode khusus yang kompleks atau penggunaan algoritma data mining tertentu sebagai pengecekan data untuk siap dilakukan analisis dengan hasil yang lebih relavan terhadap penelitian. Salah satu langkah dalam pra-proses data yaitu data *cleaning*[18]. Data *cleaning* dilakukan untuk memastikan data yang dipilih dan diintegrasikan sesuai untuk diolah. Data *cleaning* perbaikan data yang rusak, penghapusan kolom yang tidak diperlukan, dan pengecekan kolom kosong (*missing values*)[19].

2.3 Transformasi data

Data kemudian ditransformasi menggunakan berbagai metode, terutama ketika dataset yang ingin digunakan berskala besar. Tahap ini penting dilakukan sebagai peningkatan hasil yang diperoleh agar lebih bagus dan mempermudah dalam perhitungan. Agar data yang disediakan siap untuk diolah, maka transformasi data perlu dilakukan. Salah satu tahapan dalam transformasi data adalah dengan melakukan normalisasi untuk melindungi kualitas data agar tetap stabil saat digunakan[20].

Normalisasi merupakan proses yang berfungsi sebagai pengubahan skala nilai data agar berada dalam rentang tertentu. Tahap ini penting dilakukan untuk memastikan bahwa setiap fitur atau atribut mempunyai kontribusi yang setara pada pemodelan. Selain itu, normalisasi membantu menjaga validitas model, terutama ketika terdapat fitur dengan simpangan baku yang sangat kecil, sehingga mencegah dominasi atribut tertentu dalam proses analisis atau pembelajaran mesin. Hal ini disebabkan oleh fakta bahwa data sering kali memiliki rentang nilai yang sangat bervariasi antar variabel, yang dapat memengaruhi kinerja algoritma dalam analisis data[21]. Dalam penelitian ini, metode normalisasi yang digunakan adalah *MinMax Scaler*. *MinMax Scaler* merupakan salah satu teknik praproses yang berfungsi untuk mengubah nilai fitur dalam suatu dataset agar berada dalam rentang 0 hingga 1. Dengan cara ini, setiap variabel memiliki skala yang sebanding, sehingga membantu meningkatkan akurasi dan stabilitas dalam proses pemodelan[22]. Berikut adalah rumus dalam menggunakan MinMaxScaler.

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

Dimana, x' merupakan atribut data setelah dilakukan normalisasi, min(x) merupakan nilai minimum dalam kumpulan data, dan max(x) adalah nilai maksmimum dalam kumpulan data.

2.4 Penerapan Algoritma *K-Means Clustering*

Algoritma *clustering* yang biasa digunakan untuk pengelompokkan data berdasarkan ciri-ciri dari atribut data yang serupa merupakan Algoritma *K-Means Clustering*. Algoritma ini merupakan algoritma yang bekerja sebagai penerima parameter *k* sebagai jumlah klaster yang diinginkan, lalu membagi *n* objek ke dalam *k* klaster. Tujuan dari algoritma *K-Means* merupakan sebagai wadah untuk meningkatkan kemiripan antar data dalam satu *cluster* dan meminimalkan kesamaan dengan *cluster* lainnya. Dengan demikian, setiap *cluster* yang terbentuk berisi data-data yang memiliki

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

karakteristik serupa antara satu dan lainnya, namun berbeda secara signifikan dari data pada *cluster* lainnya[23]. Keunggulan dari algoritma ini terletak pada kemudahannya dalam implementasi serta kompleksitas waktu dan ruang yang relatif rendah. Hal ini menjadikan algoritma ini efisien secara komputasi. Sealin itu, algoritma *K-Means* mampu menghasilkan kualitas yang sangat bagus dan memuaskan, terutama ketika *cluster* terbentuk bersifat *compact*, berbentuk *hypersperical*, dan memiliki pemisahan fitur yang jelas antar *cluster*[24].

Penerapan algoritma *K-Means* pada umumnya dilakukan sebagai berikut[25]:

a. Menentukan jumlah *cluster*.
b. Membagi data ke beberapa kelompok secara acak dan lakukan inisialisasi centroid (pusat *cluster*) dengan cara acak. Pemilihan ini akan mempengaruhi hasil akhir, sehingga proses inisialisasi diulang beberapakli untuk memproleh hasil yang optimal.
c. Hitung ulang centroid dari dalam kelompok masing-masing.
d. Tetapkan setiap titik data ke pusat *cluster*, kedekatan antara data dan pusat *cluster* ditentukan dengan menghitung jarak diantara keduanya. Setiap data akan didekatkan kepada centroidnya. Untuk menghitung seluruh data dari setiap titik hingga centroidnya adalah dengan menerapkan rumus Euclidean Distance:

$$d(ij) = \sqrt{\sum_{i=1}^{n} \left(x_{ki} - x_{kj}\right)^2} \quad (2)$$

Dimana, d(ij) adalah jarak data (x) ke pusat *cluster* (j) dan x_{ki} adalah data ke (i) pada atribut data (k).

e. Hitung kembali centroid dengan data saat ini, centroid baru merupakan rata-rata seluruh data yang ada pada setiuap *cluster*.
f. Ulangi langkah diatas menggunakan *cluster* dan centroid yang baru. Jika pusat *cluster* sudah tidak berubah maka proses *clustering* selesai

2.5 Evaluasi hasil *Clustering*

Proses menentukan jumlah *cluster* terbaik ditentukan dengan melakukan teknik evaluasi model dengan menggunakan berbagai metode evaluasi. Penelitian ini menggunakan evaluasi hasil *clustering* dengan metode *Davies-Bouldin Index* dan metode *Elbow*.

a. *Davies-Bouldin Index* (DBI)

Metode ini bertujuan untuk mengevaluasi hasil kualitas hasil *clustering* dengan mempertimbangkan dua komponen utama, yaitu kohesi dan separasi. Kohesi mengukur sejauh mana data dalam suatu *cluster* saling berdekatan terhadap pusat *cluster* (centroid), sedangkan separasi mengukur jarak antar centroid dari *cluster* yang tidak serupa. Semakin rendah nilai DBI, maka semakin bagus hasil *clustering* yang diperoleh, karena mengindikasikan bahwa *cluster-cluster* terbentuk dengan baik dan kompak di dalam dan terpisah dari klaster lain. Perhitungan DBI dimulai dengan menghitung *Sum of Square Within Cluster* (SSW) dan *Sum of Square Between Cluster* (SSB). Nilai SSW digunakan untuk mengukur jumlah karakteristik dalam setiap *cluster* terhadap centroid, sedangkan SSB digunakan untuk mengukur jarak antar centroid dari *cluster* yang berbeda. Berikut adalah rumus yang digunakan dalam menghitung SSW[26]:

$$ssw = \frac{1}{m}\sum_{j=i}^{m_i} d(x_i, C_i) \quad (3)$$

Dengan m adalah jumlah data dalam *cluster* (i), x adalah atribut data pada *cluster*, c adalah centroid, dan d(x,c) adalah jarak data (x) pada centroid.

Rumus yang digunakan untuk menghitung *Sum of Square Between-Cluster*(SSB):

$$SSB_{i,j} = d\left(C_i, C_j\right) \quad (4)$$

Dimana, d(c,c) adalah jarak centroid dengan centroid lain.

Selanjutnya, menentukan nilai rasio yang dimiliki oleh tiap-tiap *cluster*:

$$R_{i,j} = \frac{ssw_i + ssw_j}{SSB_{i,j}} \quad (5)$$

Rumus DBI adalah:

$$DBI = \frac{1}{k}\Sigma_i^k \max\left(R_{i,j}\right) \quad (6)$$

Dimana, k adalah jumlah *cluster* yang terbentuk.

b. *Elbow*

Metode ini adalah salah satu teknik yang umum digunakan untuk mencari jumlah *cluster* terbaik pada bidang analisis *clustering*. *Elbow* bekerja dengan mengevaluasi nilai *Within-Cluster Sum of Squares* (WCSS) untuk setiap jumlah *cluster* (k), lalu menyajikannya dalam bentuk grafik. Grafik tersebut menggambarkan hubungan antara jumlah klaster dengan nilai WCSS yang dihasilkan. Seiring bertambahnya jumlah klaster, nilai WCSS akan menurun, namun pada titik tertentu penurunannya mulai melambat. Titik tersebut membentuk pola seperti "siku" (*Elbow*), dan dianggap sebagai jumlah klaster yang optimal karena menunjukkan keseimbangan antara kompleksitas model dan performa pemisahan data. Metode ini bertujuan untuk memilih nilai *cluster* yang relatif kecil namun masih

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

mempertahankan nilai WCSS yang kecil, sehingga efisien namun tetap akurat dalam pemodelan. Adapun rumus yang digunakan dalam menghitung WCSS adalah:

$$WCSS = \sum_{k=1}^{k} \sum_{x_i \in Ck} (x_i - \varphi k)^2 \quad (7)$$

Dimana, C_k adalah *cluster* yang yang dibentuk, k adalah total *cluster*, x_i adalah data x pada fitur i, dan φk adalah rata-rata *cluster* pada nilai k (k=1,2,...,k)[27].

3. HASIL DAN PEMBAHASAN

3.1 Seleksi data

Penelitian ini menggunakan data yang diperoleh secara langsung dari perusahaan manufaktur PT. XYZ, yang bergerak dalam bidang pengolahan kelapa sawit. Fokus utama terletak pada mesin *screw press*, yang merupakan salah satu komponen vital dalam proses ekstraksi minyak sawit. Dataset yang digunakan berasal dari mesin *screw press* dengan "KODE_DEVICE" BPV EP01-09, yang merupakan identifikasi unik untuk mesin tersebut dalam sistem pemantauan perusahaan. Jumlah total data yang dikumpulkan sebanyak 23.002 data, yang mencakup berbagai parameter operasional mesin seperti kode mesin, tanggal, tekanan, suhu.

Dataset ini memiliki cakupan waktu pengumpulan data yang konsisten, sehingga memungkinkan analisis tren dan pola yang representatif terhadap performa mesin dalam jangka waktu tertentu. Data-data tersebut selanjutnya dianalisis untuk menemukan pola-pola yang berkaitan dengan kondisi operasional mesin, baik dalam kondisi normal maupun saat mendekati kegagalan atau penurunan performa. Dengan adanya analisis ini, diharapkan dapat diidentifikasi indikator-indikator awal dari kerusakan mesin sehingga dapat diterapkan sistem pemeliharaan prediktif (predictive maintenance). Berikut adalah visualisasi datset yang digunakan:

Tabel 1. Data mentah PT.XYZ

NO	KODE_DEVICE	TANGGAL	TEMPERATURE	PRESSURE
1	N00E-EP01-BPV-01	18/04/2024	123	3,19
2	N00E-EP01-BPV-01	18/04/2024	123	3
3	N00E-EP01-BPV-01	18/04/2024	123	2,95
…	…	…	…	…
782	N00E-EP04-BPV-01	18/04/2024	0	2,23
…	…	…	…	…
5531	N00E-EP05-BPV-01	18/04/2024	100	2,55
…				
6834	N00E-EP01-BPV-01	19/04/2024	123	4,43
…	…	…	…	…
22650	N00E-EP09-BPV-01	25/04/2024	123	0
…				
23002	N00E-EP06-BPV-01	04/05/2024	100	2,03

Selanjutnya, dari keseluruhan dataset yang tersedia, dilakukan proses seleksi atribut untuk membentuk data latih yang akan digunakan dalam tahap analisis dan pemodelan. Berdasarkan Tabel 1. Dilakukan seleksi atribut yang dibutuhkan pada penelitian ini. Atribut yang dipilih adalah ['TEMPERATURE'] dan ['PRESSURE'], karena kedua parameter ini dianggap paling relevan dan berpengaruh langsung terhadap kinerja serta kondisi mesin *screw press*. Pemilihan atribut ini juga didasarkan pada literatur teknis dan wawasan praktis dari tim teknisi di PT. XYZ, yang menyatakan bahwa fluktuasi suhu dan tekanan sering menjadi indikator utama dalam mendeteksi adanya anomali atau potensi kerusakan mesin. Berikut adalah tampilan data latih yang diperoleh:

Tabel 2. Dataset setelah diambil data latih

NO	TEMPERATURE	PRESSURE
1	123	3,19
2	123	3
3	123	2,95
…	…	…
782	0	2,23
…	…	…
5531	157	3,51
…		
6834	123	4,43

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

…	…	…
22650	123	0
…	…	…
23002	100	2,03

Data latih yang telah diperoleh dengan atribut berdasarkan suhu, dan tekanan dapat dilihat pada Tabel 2. Penelitian ini akan berfokus pada kedua atribut tersebut sebagai data latih yang akan dilakukan *clustering* menjadi berbagai *cluster*. Data latih ini selanjutnya akan melalui proses pra-pemrosesan, seperti normalisasi atau *scaling*, serta eksplorasi visual untuk memahami distribusi dan hubungan antar atribut. Proses ini penting untuk memastikan efektifitas data yang ingin digunakan sebagai pembangunan model prediktif mengenai kondisi mesin *screw press*.

3.2 Pra-proses data

Tahap ini merupakan langkah penting dalam memastikan kualitas dan kelayakan data sebelum digunakan dalam analisis lanjutan atau pemodelan. Pada penelitian ini, tahapan pra-pemrosesan yang dilakukan difokuskan pada pembersihan data (*data cleaning*), dengan maksud untuk menghilangkan atau memperbaiki data yang tidak konsisten, data kosong (missing valuse), atau tidak valid. *Cleaning* data dilakukan dalam tahap pra-proses data yang bertujuan untuk meningkatkan kualitas dan keefektifan data sebelum dilakukan tahap analisis berikutnya. Penelitian ini menggunakan data dari PT.XYZ secara mentah, setelah data diseleksi berdasarkan devicenya maka dilakukan pengecekan terhadap kemungkinan duplikasi data, missing value, dan memastikan data yang ingin dianalisis memiliki tipe data yang sama dan relavan untuk diolah. Setelah dilakukan pengecekan dan pembersihan data maka dataset menjadi lebih efektif digunakan, sehingga hasil *clustering* menjadi lebih akurat. Berikut hasil pengecekan dataset berdasarkan data yang sudah diseleksi.

Tabel 3. Dataset setelah dilakukan pengecekan data

NO	Atribut	Jumlah	Null Count	Tipe Data
1	TEMPERATURE	23002	Not-Null	Float64
2	PRESSURE	23002	Not-Null	Float64

Berdasarkan Tabel 3, dapat diketahui bahwa kedua atribut yang digunakan, yaitu TEMPERATURE dan PRESSURE, memiliki jumlah data yang sama sebanyak 23.002 entri, tidak terdapat missing value (Not-Null), serta memiliki tipe data yang seragam yaitu float64. Hal ini menunjukkan bahwa dataset telah melewati tahap data *cleaning* dengan baik dan layak digunakan untuk analisis lebih lanjut. Dengan tidak adanya nilai kosong maupun ketidaksesuaian tipe data, maka dataset yang telah bersih ini siap untuk memasuki tahap transformasi data guna meningkatkan efektivitas dalam proses pemodelan kondisi mesin *screw press*.

3.3 Transformasi data

Tahap selanjutnya, dilakukan pemodelan data yang berfungsi untuk memudahkan proses analisis serta meningkatkan kualitas data dan performa algoritma yang akan digunakan. Salah satu langkah penting dalam proses transformasi data adalah normalisasi, yang berfungsi untuk menyetarakan skala antar variabel agar analisis lebih akurat dan efisien. Metode normalisasi yang digunakan dalam penelitian ini adalah MinMaxScaler, karena metode ini efektif dalam memperkecil rentang nilai antar atribut tanpa mengubah karakteristik distribusi data. Proses normalisasi dilakukan menggunakan rumus (1) pada Subbab 2.3.1 terhadap dataset yang telah diambil sebagai data latih pada Tabel 2. Nilai minimum dan maksimum yang digunakan dalam proses ini adalah 0 dan 4,43 untuk atribut PRESSURE, serta 0 dan 157 untuk atribut TEMPERATURE. Berdasarkan nilai-nilai tersebut, diperoleh hasil normalisasi yang siap digunakan dalam tahap analisis dan pemodelan kondisi mesin *screw press* sebagai berikut.

Tabel 4. Dataset setelah dilakukan normalisasi menggunakan MinMaxScaler

NO	TEMPERATURE	PRESSURE
1	0,783	0,720
2	0,783	0,677
3	0,783	0,666
…	…	…
782	0	0,503
…	…	…
5531	1	0,792
…	…	…
6834	0,783	1
…	…	…
22650	0,783	0
…	…	…
23002	0,636	0,458

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Berdasarkan Tabel 4, dapat dilihat bahwa dataset awal telah berhasil dinormalisasi menggunakan metode MinMaxScaler sesuai dengan rumus (1) yang telah dijelaskan sebelumnya. Proses normalisasi ini bertujuan untuk memperkecil rentang nilai antar variabel sehingga seluruh data berada dalam skala yang seragam, yaitu antara 0 hingga 1. Dengan normalisasi ini, perbedaan skala antara atribut TEMPERATURE dan PRESSURE tidak lagi menjadi kendala dalam proses analisis, sehingga algoritma yang digunakan di tahap selanjutnya dapat bekerja secara lebih optimal dan adil terhadap masing-masing variabel. Oleh karena itu, data yang telah dinormalisasi ini dinyatakan siap untuk digunakan pada tahap berikutnya, seperti eksplorasi data, klasifikasi, atau pemodelan prediktif terhadap kondisi mesin *screw press*.

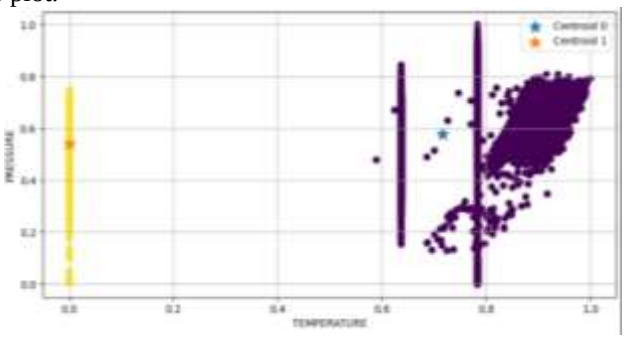
3.4 Penerapan Algoritma *K-Means Clustering*

Proses *clustering* dalam penelitian ini dilakukan menggunakan algoritma *K-Means* terhadap sebanyak 23.002 data hasil normalisasi yang terdiri dari atribut TEMPERATURE dan PRESSURE. Tujuannya adalah untuk mengelompokkan data ke dalam beberapa *cluster* berdasarkan kemiripan pola kondisi mesin *screw press*. Pengujian dilakukan dengan variasi jumlah *cluster* mulai dari 2 hingga 15, di mana masing-masing pengujian menghasilkan label *cluster* dari C0 hingga C14, tergantung pada jumlah *cluster* yang digunakan. Pada setiap iterasi, algoritma *K-Means* diterapkan dengan menggunakan rumus (2) yang telah dijelaskan pada Subbab 2.4, dan diterapkan secara konsisten pada data yang telah dinormalisasi (Tabel 3). Dengan menggunakan pendekatan ini, setiap data akan diberi label berdasarkan *cluster* tempatnya tergolong, sehingga hasil pengelompokan ini dapat memberikan gambaran tentang kondisi operasional mesin dalam kelompok yang lebih terstruktur dan menjadi dasar analisis lebih lanjut, seperti deteksi kondisi tidak normal atau potensi kerusakan.

Tabel 5. Pengujian terhadap 2 jumlah *cluster*

Cluster	Jumlah data
C0	18529
C1	4473

Setelah dilakukan pengujian dengan jumlah *cluster* sebanyak dua (2), diperoleh hasil pengelompokan data seperti yang ditunjukkan pada Tabel 5. Hasil ini menunjukkan bahwa sebagian besar data tergolong ke dalam *cluster* C0 dengan total 18.529 data, sedangkan sisanya masuk ke dalam *cluster* C1 sebanyak 4.473 data. Untuk memberikan gambaran yang lebih jelas terhadap distribusi dan struktur dari hasil *clustering* tersebut, dilakukan proses visualisasi menggunakan scatter plot. Visualisasi ini bertujuan untuk merepresentasikan sebaran data berdasarkan dua atribut utama yaitu TEMPERATURE dan PRESSURE, serta menunjukkan posisi pusat *cluster* (centroid) yang terbentuk dari proses *K-Means*. Berdasarkan hasil perhitungan, pusat *cluster* untuk atribut TEMPERATURE berada pada koordinat [0,716] dan [0,344e-14], sementara untuk atribut PRESSURE berada pada [0,578] dan [0,541]. Melalui scatter plot ini, pola pengelompokan antar data menjadi lebih mudah diamati dan dapat memberikan wawasan awal terhadap kemungkinan karakteristik dari masing-masing *cluster*, misalnya kondisi mesin yang stabil dan tidak stabil. Berikut adalah hasil visualisai menggunakan scatter plot.



Gambar 2. Visualisasi hasil *clustering* menggunakan Scatter plot

Pada Gambar 2, menunjukkan pemisahan yang jelas antara *cluster* C0 dan C1. C0, ditandai dengan warna kuning, merepresentasikan kondisi mesin dalam kategori aman dengan suhu yang stabil. Sementara itu, C1, ditandai dengan warna ungu, menggambarkan kondisi mesin dalam status siaga akibat suhu yang tidak stabil. Visualisasi ini membantu mengidentifikasi pola operasional mesin berdasarkan suhu dan tekanan secara lebih representatif. Pengujian terhadap 3 – 15 *cluster* dilakukan dengan langkah yang sama seperti sebelumnya.

3.5 Evaluasi

Setelah menerapkan algoritma *K-Means* dengan berbagai pengujian *cluster*, langkah berikutnya adalah melakukan evaluasi dan interpretasi terhadap kualitas hasil pengelompokan tersebut. Tahapan ini bertujuan untuk menilai seberapa baik data dikelompokkan berdasarkan kesamaan karakteristiknya, serta memastikan bahwa jumlah *cluster* yang dipilih memberikan hasil yang optimal. Dalam penelitian ini, digunakan dua metode evaluasi, yaitu *Davies-Bouldin Index* (DBI) dan *Elbow Method*. Kedua metode ini digunakan secara komplementer untuk menentukan jumlah *cluster* yang paling

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

representatif terhadap struktur data, sehingga proses *clustering* tidak hanya akurat secara teknis tetapi juga bermakna dalam konteks analisis kondisi mesin *screw press*.

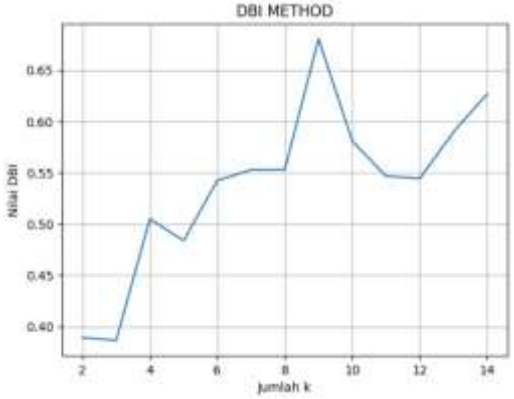
a. *Davies-Bouldin Index*

Evaluasi hasil *clustering* sangat berperan untuk memastikan bahwa pembagian data kedalam *cluster* sudah optimal. Menerapkan metode *Daives-Bouldin Index* (DBI) dapat mengukur seberapa bagus nilai dari tiap tiap *cluster* dengan memilih jumlah *cluster* yang menghasilkan nilai DBI paling kecil, semakin kecil nilai DBI yang dihasilkan maka semakin bagus *cluster* yang telah dilakukan. Berikut perhitungan nilai DBI berdasarkan jumlah *cluster* yang diambil. Berikut adalah hasil evaluasi dengan DBI.

Tabel 6. Evaluasi hasil *cluster* menggunakan *Davies-Bouldin Index*

Cluster	Nilai DBI
2	0,389
3	0,386
4	0,505
5	0,483
6	0,500
7	0,620
8	0,551
9	0,557
10	0,578
11	0,548
12	0,610
13	0,557
14	0,590

Evaluasi hasil *clustering* dengan menggunakan metode *Davies-Bouldin Index* (DBI) dilakukan untuk menilai seberapa baik pembagian data ke dalam masing-masing *cluster*, di mana semakin kecil nilai DBI menunjukkan kualitas *clustering* yang lebih baik. Perhitungan nilai DBI dilakukan berdasarkan rumus (3) hingga rumus (6) terhadap data pada Tabel 3, dan menghasilkan nilai DBI sebagaimana disajikan pada Tabel 7. Dari hasil tersebut, dapat diamati bahwa nilai DBI terendah diperoleh saat jumlah *cluster* adalah tiga (k = 3) dengan nilai 0,386, yang menunjukkan bahwa pemisahan dan kepadatan *cluster* paling optimal terjadi pada konfigurasi tersebut. Untuk memperjelas representasi dan tren dari hasil evaluasi ini, dilakukan visualisasi menggunakan scatter plot, sehingga pola perubahan nilai DBI terhadap variasi jumlah *cluster* dapat dianalisis secara lebih intuitif dan mendukung dalam pengambilan keputusan terhadap jumlah *cluster* yang paling ideal. Berikut hasil visualisasi dengan metode DBI:



Gambar 3. Grafik Evaluasi dengan menggunakan DBI

Berdasarkan Gambar 4., diperoleh hasil DBI dengan nilai terkecil pada *cluster* 3 dengan nilai DBI 0,386. Oleh karena itu, evaluasi hasil *clustering* menggunakan metode DBI memperoleh *cluster* terbaik berada di *cluster* 3.

b. *Elbow*

Metode *Elbow* digunakan dalam evaluasi *clustering* untuk menentukan jumlah *cluster* yang optimal dengan menganalisis hubungan antara jumlah *cluster* dan total variasi dalam *cluster*, yang diukur menggunakan *Within-Cluster Sum of Squares* (WCSS). Prinsip dasar dari metode ini adalah mengidentifikasi titik "siku" pada grafik WCSS, yaitu titik di mana penurunan nilai WCSS mulai melambat secara signifikan seiring dengan penambahan jumlah *cluster*. Titik ini dianggap sebagai jumlah *cluster* yang paling efisien karena menyeimbangkan antara kompleksitas model dan tingkat kesalahan dalam *cluster*. Dalam penelitian ini, pendekatan *Elbow* diterapkan untuk mengevaluasi hasil dari algoritma *K-Means*, dengan tujuan memperoleh jumlah *cluster* yang menghasilkan nilai WCSS yang relatif rendah namun tetap mempertahankan efektivitas penyebaran data dalam masing-masing *cluster* (*within-cluster Sum of Squares* yang rendah). Perhitungan WCSS dilakukan menggunakan rumus (7) berdasarkan data yang terdapat pada Tabel 3, yang kemudian divisualisasikan dalam bentuk grafik untuk membantu mengidentifikasi titik siku secara

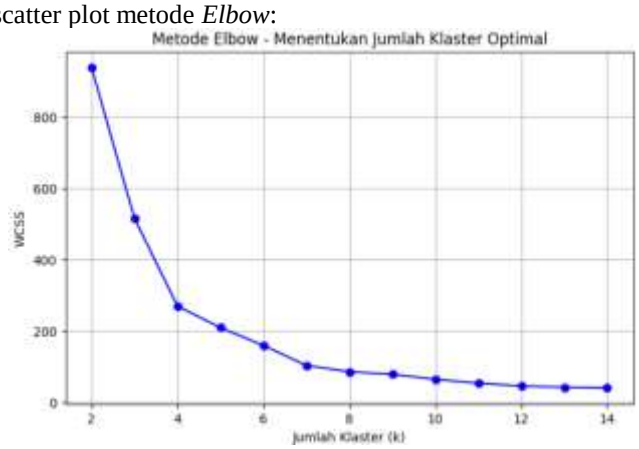
BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

visual. Visualisasi ini sangat penting karena memungkinkan analisis intuitif terhadap tren penurunan WCSS dan penentuan jumlah *cluster* yang optimal. Setelah dilakukan perhitungan, diperoleh nilai WCSS untuk masing-masing jumlah *cluster*, yang selanjutnya dianalisis guna mendukung pemilihan konfigurasi *cluster* terbaik. Berikut adalah perolehan nilai WCSS dari masing-masing *cluster* uji:

Tabel 7. Evaluasi hasil *cluster* berdasarkan nilai WCSS

Cluster	WCSS
2	938
3	515
4	270
5	220
6	153
7	103
8	85
9	76
10	64
11	53
12	46
13	39
14	40

Berdasarkan Tabel 8, perhitungan nilai *Within-Cluster Sum of Squares* (WCSS) terhadap data pada Tabel 3 telah dilakukan untuk setiap variasi jumlah *cluster* mulai dari k = 2 hingga k = 14. Nilai WCSS yang diperoleh menunjukkan penurunan signifikan seiring bertambahnya jumlah *cluster*, yaitu dari 938 (k = 2) hingga 40 (k = 14). Penurunan paling tajam terjadi pada awal peningkatan jumlah *cluster*, terutama antara k = 2 hingga k = 4, setelah itu laju penurunan mulai melambat. Pola ini mencerminkan prinsip dasar dari metode *Elbow*, di mana titik siku (*Elbow* point) dalam grafik WCSS mencerminkan jumlah *cluster* yang optimal—yakni pada titik di mana penambahan jumlah *cluster* tidak lagi memberikan penurunan WCSS yang signifikan. Dalam konteks ini, nilai *within-cluster Sum of Squares* (WCSS) dapat dikatakan mulai stabil dari *cluster* 4 hingga *cluster* 14. Oleh karena itu, untuk mendukung identifikasi visual terhadap titik siku tersebut, dilakukan visualisasi hasil evaluasi menggunakan scatter plot. Visualisasi ini bertujuan untuk menampilkan tren penurunan WCSS secara grafis sehingga mempermudah dalam menentukan jumlah *cluster* terbaik yang mencerminkan keseimbangan antara kompleksitas model dan kualitas pemodelan data. Berikut adalah visualisasi menggunakan scatter plot metode *Elbow*:



Gambar 4. Grafik Evaluasi hasil *clustering* dengan menggunakan *Elbow*

Berdasarkan visualisasi pada Gambar 5, terlihat bahwa titik siku paling jelas dalam grafik metode *Elbow* muncul pada jumlah *cluster* ke-4, yang ditandai dengan nilai WCSS sebesar 270. Titik ini menunjukkan perubahan signifikan dalam tren penurunan nilai WCSS dibandingkan dengan jumlah *cluster* sebelumnya, sementara penurunan WCSS setelah *cluster* ke-4 tidak lagi menunjukkan perbedaan yang besar. Hal ini mengindikasikan bahwa penambahan jumlah *cluster* setelah titik tersebut hanya memberikan peningkatan yang minimal terhadap kualitas *clustering*. Oleh karena itu, dapat disimpulkan bahwa jumlah *cluster* yang paling optimal berdasarkan pendekatan *Elbow* adalah empat *cluster*, karena pada titik ini tercapai keseimbangan terbaik antara kompleksitas model dan efektivitas pemisahan data dalam *cluster*.

3.6 Analisis perbandingan hasil *cluster* terbaik

Perbandingan hasil *clustering* menggunakan metode DBI dan Elbow bertujuan untuk menentukan jumlah *cluster* yang paling optimal berdasarkan karakteristik masing-masing *cluster*. Metode DBI menunjukkan hasil terbaik pada 3 *cluster*, sedangkan Elbow mengindikasikan 4 *cluster* sebagai pilihan optimal. Oleh karena itu, analisis lebih lanjut diperlukan

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

untuk memperdalam pemahaman dan mendukung pengambilan keputusan dalam pengelompokan kondisi mesin screw press. Berikut adalah tabel analisis berdasarkan penggunaan *cluster* terbaik berdasarkan metode tersebut.

Tabel 8. Analisis berdasarkan penggunaan 3 *cluster*

Cluster	Jumlah data	Rata-rata Temperature	Rata-rata Pressure	Karakteristik
C0	4473	0	1,7	Mesin dalam kondisi bagus
C1	1465	100-123	1,1	Mesin memerlukan pengawasan
C2	17064	123	3,2	Mesin memerlukan pemeliharaan segera

Tabel 9. Analisis penggunaaan 4 *cluster*

Cluster	Jumlah data	Rata-rata Temperature	Rata-rata Pressure	Karakteristik
C0	4473	0	1,7	Mesin dalam kondisi bagus
C1	11143	100-123	~1,8	Mesin memerlukan pengawasan tahap 2
C2	6015	123	~3,2	Mesin memerlukan pemeliharaan segera
C3	1371	123	~1,0	Mesin memerlukan pengawasan tahap 1

Analisis hasil *clustering* menunjukkan bahwa penggunaan 3 *cluster* unggul dalam menyederhanakan pengelompokan kondisi mesin secara efisien, meskipun tiap *cluster* bisa mencakup kondisi yang bervariasi. Sebaliknya, 4 *cluster* mampu memberikan informasi yang lebih rinci, namun masih terdapat tumpang tindih kondisi dalam satu *cluster*.

4. KESIMPULAN

Tujuan dari penerapan data mining dengan algoritma *K-Means Clustering* pada dataset yang diperoleh dari PT. XYZ adalah untuk mengidentifikasi pola-pola tersembunyi terkait kondisi mesin yang tidak dapat langsung diamati dari data mentah. Proses *clustering* dalam penelitian ini dilakukan melalui pengujian terhadap 15 konfigurasi *cluster*, yaitu dari C0 hingga C15, yang kemudian dievaluasi menggunakan dua metode utama: *Davies-Bouldin Index* (DBI) dan *Elbow* Method. Berdasarkan hasil evaluasi dengan DBI, nilai indeks terkecil diperoleh saat menggunakan tiga *cluster* dengan skor 0,386, yang menunjukkan bahwa pengelompokan data dalam tiga *cluster* memberikan kualitas segmentasi yang paling optimal menurut metrik tersebut. Di sisi lain, hasil evaluasi menggunakan metode *Elbow* menunjukkan bahwa titik siku pada grafik WCSS berada pada jumlah *cluster* keempat, dengan nilai WCSS sebesar 270, yang mengindikasikan bahwa empat *cluster* merupakan pilihan terbaik berdasarkan keseimbangan antara kompleksitas dan akurasi model. Penggunaan kedua metode evaluasi ini memberikan perspektif komplementer dalam menentukan jumlah *cluster* yang paling representatif terhadap pola data. Meskipun hasil *clustering* telah mampu menggambarkan pola kondisi mesin secara umum, penelitian ini masih memiliki keterbatasan, terutama pada jumlah atribut variabel yang dianalisis. Oleh karena itu, penelitian lanjutan disarankan untuk memperluas cakupan dengan menambahkan lebih banyak atribut, melibatkan berbagai jenis mesin, serta memperpanjang periode pengumpulan data guna memperoleh hasil yang lebih komprehensif dan akurat.

REFERENCES

[1] I. Anwar, A. Afdal, D. Denur, and D. Dermawan, “Analisis Penyebab Kerusakan Hub Screw Press dan Optimasi Teknik Pemeliharaan Terhadap Mesin Pabrik Pengolahan Kelapa Sawit,” *Jurnal Surya Teknika*, vol. 10, no. 1, pp. 733–737, 2023, doi: 10.37859/jst.v10i1.5003.

[2] M. I. Pasaribu, D. A. A. Ritonga, and A. Irwan, “Analisis Perawatan (Maintenance) Mesin Screw Press Di Pabrik Kelapa Sawit Dengan Metode Failure Mode and Effect Analysis (Fmea) Di Pt. Xyz,” *Jitekh*, vol. 9, no. 2, pp. 104–110, 2021, doi: 10.35447/jitekh.v9i2.432.

[3] N. Hairiyah, I. Musthofa, and A. Aminah, “Analisis Kerusakan Bearing Main Shaft Pada Mesin Screw Press Msb 15 Dengan Metode Total Productive Maintenance (Tpm) Di Pabrik Kelapa Sawit Pt. Xyz,” *Jurnal Teknologi Pertanian Andalas*, vol. 28, no. 1, p. 1, 2024, doi: 10.25077/jtpa.28.1.1-7.2024.

[4] A. Ishak, R. M. Sari, and G. H. Sabri, “Perencanaan Sistem Pemeliharaan Mesin Screw Press Menggunakan Metode Reliability Centered Maintenance (Studi Kasus pada PMKS),” pp. 324–334, 2023.

[5] A. Fabrica, “4 1,2,3,” vol. 5, no. 2, 2023.

BULLETIN OF INFORMATION TECHNOLOGY (BIT)
Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

[6] M. Wisnu, G. Supriyanto, and Hermantoro, "Analisis Kebutuhan Uap pada Perebusan Tiga Puncak," *Agroforetech*, vol. 1, pp. 685–692, 2023.
[7] P. Rahayu *et al.*, *Buku Ajar Data Mining*, vol. 1, no. January 2024. 2018.
[8] N. D. Rahayu, A. H. Anshor, and I. Afriantoro, "Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means," *JUKI : Jurnal Komputer dan Informatika*, vol. 6, no. 1, pp. 71–83, 2024, doi: 10.53842/juki.v6i1.474.
[9] D. Suryanto and I. R. Adevi, "Analisa Penjualan Toko Hijab Kiki Hn Dengan Penerapan Data Mining Metode K-Means Clustering Dan Market Basket Analysis," *Jurnal Mirai Management*, vol. 8, no. 3, pp. 167–176, 2023.
[10] Novia Wulandari, Nisa Dienwati Nuris, and Saeful Anwar, "Penerapan Algoritma K-Means Clustering Pada Tingkat Inflasi Kota Di Indonesia," *Akuntansi*, vol. 2, no. 2, pp. 15–34, 2023, doi: 10.55606/akuntansi.v2i2.235.
[11] A. Al Masykur, S. K. Gusti, S. Sanjaya, F. Yanto, and F. Syafria, "Penerapan Metode K-Means Clustering untuk Pemetaan Pengelompokan Lahan Produksi Tandan Buah Segar," *Jurnal Informatika*, vol. 10, no. 1, 2023, doi: 10.31294/inf.v10i1.15621.
[12] Z. Sitorus, "Penerapan Data Mining Untuk Clustering Penduduk Miskin Di Kota Tanjungbalai Menggunakan Metode Algoritma K-Means," *Journal of Science and Social Research*, vol. 4307, no. 1, pp. 212–218, 2024, [Online]. Available: http://jurnal.goretanpena.com/index.php/JSSR
[13] A. Nugraha, O. Nurdiawan, and G. Dwilestari, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Yana Sport," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 849–855, 2022, doi: 10.36040/jati.v6i2.5755.
[14] E. Nurliana, B. Irawan, and A. Bahtiar, "Implementasi Data Mining Algoritma K-Means Untuk Klasifikasi Penduduk Miskin Berdasarkan Tingkat Kemiskinan Di Jawa Barat," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 1116–1122, 2024, doi: 10.36040/jati.v8i1.8883.
[15] E. Muningsih, I. Maryani, and V. R. Handayani, "Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa," *Jurnal Sains dan Manajemen*, vol. 9, no. 1, p. 96, 2021, [Online]. Available: www.bps.go.id
[16] A. Winarta and W. J. Kurniawan, "Optimasi Cluster K-Means Menggunakan Metode Elbow Pada Data Pengguna Narkoba Dengan Pemrograman Python," *JTIK (Jurnal Teknik Informatika Kaputama)*, vol. 5, no. 1, pp. 113–119, 2021, doi: 10.59697/jtik.v5i1.593.
[17] A. Supriyadi, A. Triayudi, and I. D. Sholihati, "Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 6, no. 2, pp. 229–240, 2021, doi: 10.29100/jipi.v6i2.2008.
[18] F. M. Almufqi and A. Voutama, "Perbandingan Metode Data Mining Untuk Memprediksi Prestasi Akademik Siswa," *Jurnal Teknika*, vol. 15, no. 1, pp. 61–66, 2023, doi: 10.30736/jt.v15i1.929.
[19] R. Rahmi, D. Antoni, H. Syaputra, F. Fatoni, and T. B. Kurniawan, "Metode Klasifikasi Gejala Penyakit Coronavirus Disease 19 (COVID-19) Menggunakan Algoritma Neural Network," *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, vol. 12, no. 1, pp. 16–23, 2023, doi: 10.32736/sisfokom.v12i1.1406.
[20] I. Pii, N. Suarna, and N. Rahaningsih, "Penerapan Data Mining Pada Penjualan Produk Pakaian Dameyra Fashion Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 423–430, 2023, doi: 10.36040/jati.v7i1.6336.
[21] R. A. Nawasta, N. H. Cahyana, and H. Heriyanto, "Implementation of Mel-Frequency Cepstral Coefficient as Feature Extraction using K-Nearest Neighbor for Emotion Detection Based on Voice Intonation," *Telematika*, vol. 20, no. 1, p. 51, 2023, doi: 10.31315/telematika.v20i1.9518.
[22] Ary Prandika Siregar, Dwi Priyadi Purba, Jojor Putri Pasaribu, and Khairul Reza Bakara, "Implementasi Algoritma Random Forest Dalam Klasifikasi Diagnosis Penyakit Stroke," *Jurnal Penelitian Rumpun Ilmu Teknik*, vol. 2, no. 4, pp. 155–164, 2023, doi: 10.55606/juprit.v2i4.3039.
[23] C. Nas, "Data Mining Pengelompokan Bidang Keahlian Mahasiswa Menggunakan Algoritma K-Means (Studi Kasus : Universitas Cic Cirebon)," *Syntax : Jurnal Informatika*, vol. 9, no. 1, pp. 1–14, 2020, doi: 10.35706/syji.v9i1.3472.
[24] C. S. D. B. Sembiring, L. Hanum, and S. P. Tamba, "Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Menentukan Judul Skripsi Dan Jurnal Penelitian (Studi Kasus Ftik Unpri)," *Jurnal*

Vol 6, No 2, Juni 2025, Hal. 59 - 70
ISSN 2722-0524 (media online)
DOI 10.47065/bit.v5i2.1783
https://journal.fkpt.org/index.php/BIT

Sistem Informasi dan Ilmu Komputer Prima(JUSIKOM PRIMA), vol. 5, no. 2, pp. 80–85, 2022, doi: 10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2393.

[25] A. Bahauddin, A. Fatmawati, and F. Permata Sari, "Analisis Clustering Provinsi Di Indonesia Berdasarkan Tingkat Kemiskinan Menggunakan Algoritma K-Means," *Jurnal Manajemen Informatika dan Sistem Informasi*, vol. 4, no. 1, pp. 1–8, 2021, doi: 10.36595/misi.v4i1.216.

[26] 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, p. 56, 2023, doi: 10.30998/string.v8i1.16388.

[27] N. A. Maori and E. Evanita, "Metode Elbow dalam Optimasi Jumlah Cluster pada K-Means Clustering," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 14, no. 2, pp. 277–288, 2023, doi: 10.24176/simet.v14i2.9630.

[28] Z. Budiarso, H. Listiyono, and A. Karim, "Optimizing LSTM with Grid Search and Regularization Techniques to Enhance Accuracy in Human Activity Recognition," *Journal of Applied Data Sciences*, vol. 5, no. 4, pp. 2002–2014, Nov. 2024, doi: 10.47738/jads.v5i4.433.

[29] B. Bangun and A. K. Karim, "Pengembalian Data Yang Hilang Pada Dataset Dengan Menggunakan Algoritma K-Nearest Neighbor Imputation Data Mining," *Jurnal Media Informatika Budidarma*, vol. 8, no. 3, p. 1706, 2024, doi: 10.30865/mib.v8i3.8014.

[30] A. Karim, "Penerapan Algoritma Entropy dan Aras Menentukan Desa Terbaik Di Pemerintah Kabupaten Labuhanbatu," vol. 3, no. 1, pp. 33–43, 2022.

[31] A. Karim, "Implementation of the Multi-Objective Optimization Method on the Basic of Ratio Analysis (MOORA) and Entropy Weighting in New Employee Recruitment," vol. 5, no. 2, pp. 704–712, 2024, doi: 10.47065/josh.v5i2.4859.

[32] A. Karim, "Sistem Pendukung Keputusan Penerimaan Analis Di Pusat Penelitian Kelapa Sawit Menggunakan Metode Complex Proportional Assessment (Copras)," *Buletin Ilmiah Informatika Teknologi*, vol. 2, no. 1, pp. 32–42, [Online]. Available: https://ejurnal.amikstiekomsu.ac.id/index.php/BIIT

[33] A. Karim, "Clusterisasi Tingkat Pengangguran Terbuka Menurut Provinsi di Indonesia Menggunakan Algoritma K-Medoids," 2024, doi: 10.47065/bits.v6i3.6198.

[34] Abdul Karim, "Implementasi Metode Multi-Objective Optimization On The Basis Of Ratio Analysis dalam Seleksi Mahasiswa Program Indonesia Pintar," *Bulletin of Computer Science Research*, vol. 3, no. 5, pp. 351–356, 2023, doi: 10.47065/bulletincsr.v3i5.283.