

LAPORAN TUGAS BESAR

“IMPLEMENTASI ALGORITMA K-MEANS DALAM PROSES *CLUSTERING* HASIL PENILAIAN KINERJA KARYAWAN DI PT. XYZ”

Disusun Untuk Memenuhi Tugas Mata Kuliah Data Mining

Dosen Pengampu : Dwi Welly Sukma Nirad, M.T, Aina Hubby Aziira, M. Eng, dan
Febby Apri Wenando, M.Eng.



Oleh :

Kelompok 6

1. Nadia Nur Saida (2111521007)
2. Azzahra Athifah Dzaki (2111521009)
3. Fayza Zeevania Putri (2111521013)
4. Rakhilca Yanedika (2111522005)

**DEPARTEMEN SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS ANDALAS
2023**

KATA PENGANTAR

Puji dan syukur penyusun ucapkan atas kehadiran Allah SWT. Karena berkat rahmat, karunia dan hidayah-Nya, penulis dapat menyelesaikan laporan Tugas Besar Data Mining yang berjudul “Implementasi Algoritma K-Means Dalam Proses *Clustering* Hasil Penilaian Kinerja Karyawan di PT. XYZ” dengan baik meskipun masih banyak kekurangan didalamnya.

Laporan ini dibuat untuk memenuhi Tugas Besar mata kuliah Data Mining. Dalam penyusunan laporan ini, penulis mengucapkan terima kasih kepada Ibu Dwi Welly Sukma Nirad. .MT, Ibu Aina Hubby Azira. .M.Eng, dan Bapak Febby Apri Wenando. .M.Eng selaku dosen pengampu mata kuliah Data Mining yang telah memberikan bimbingan dalam pembuatan Tugas Besar ini. Serta juga sangat berterimakasih kepada semua pihak yang telah membantu dalam pembuatan laporan ini, sehingga laporan ini dapat selesai tepat pada waktunya.

Selaku penulis sangat mengetahui bahwa laporan ini jauh dari kesempurnaan. Oleh karena itu, mohon kritik dan saran yang membangun agar dapat memperbaiki untuk penyusunan laporan yang lebih baik. Semoga laporan ini dapat bermanfaat bagi semua pihak, terutama bagi penulis.

Padang, 07 Juni 2023

Penulis

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI	ii
DAFTAR GAMBAR	iv
DAFTAR TABEL	v
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.5 Manfaat.....	3
BAB II.....	4
LANDASAN TEORI.....	4
2.1 Tinjauan Pustaka.....	4
2.1.1 Pengertian Data Mining.....	4
2.1.2 Tujuan Data Mining.....	4
2.1.3 Manfaat Data Mining.....	5
2.1.4 Penerapan Data Mining	5
2.1.5 Tahap-tahap Data Mining	6
2.1.6 Fungsi Data Mining	7
2.1.7 Metode Clustering	7
2.1.8 Algoritma K-Means.....	7
2.2 Tinjauan Organisasi Penelitian.....	8
2.2.1 Latar Belakang Perusahaan	8
2.2.2 Latar Belakang Masalah Perusahaan	9
2.2.3 Batasan Masalah Penelitian Perusahaan.....	11
BAB III	12
METODOLOGI	12
3.1 Metode Penelitian	12
3.2 Tahap Pengumpulan Data	12
3.2.1 Lokasi dan Waktu Pengumpulan Data.....	12
3.2.2 Kebutuhan Laporan	13
3.2.3 Metode Pengumpulan Data	13
3.3 Tahapan Pembuatan Laporan	13

3.4 Tahap Preprocessing Data.....	14
3.5 Tahap Klasifikasi Data.....	15
BAB IV	17
HASIL DAN PEMBAHASAN	17
4.1 Pengumpulan Data.....	17
4.2 Pre-Processing Data	17
4.2.1 Data Reduction	18
4.2.3 Data Cleaning	22
4.3 Perhitungan Manual Algoritma K-Means Clustering	22
4.4 Pehitungan Menggunakan Python Algoritma K-Means Clustering	39
4.4.1 Pengolahan Data	39
4.4.2 Hasil Akhir	56
4.5 Perbandingan Hasil Perhitungan	63
BAB V	64
PENUTUP	64
5.1 Kesimpulan	64
5.2 Saran.....	65
DAFTAR PUSTAKA	65

DAFTAR GAMBAR

Gambar 2 1 Kajian Umum Data Mining.....	4
Gambar 4 1 Data Hasil Penilaian Kinerja Karyawan PT.XYZ.....	17
Gambar 4 2 Perhitungan Jarak Centroid Awal.....	24
Gambar 4 3 Hasil Cluster Data Awal	25
Gambar 4 4 Perhitungan Jarak Centroid Iterasi ke-1	27
Gambar 4 5 Hasil Cluster Data Iterasi ke-1	28
Gambar 4 6 Perhitungan Jarak Centroid Iterasi ke-2	30
Gambar 4 7 Hasil Cluster Data Iterasi ke-2	31
Gambar 4 8 Perhitungan Jarak Centroid Iterasi ke-3	33
Gambar 4 9 Hasil Cluster Data Iterasi ke-3	34
Gambar 4 10 Matrix Kolerasi.....	45
Gambar 4 11 Grafik Elbow Method	48
Gambar 4 12 Bar Chart Perbandingan Jumlah Setiap Cluster	52
Gambar 4 13 Grafik Hasil K-Means Clustering.....	54
Gambar 4 14 Grafik Hasil K-Means Clustering versi 2	55

DAFTAR TABEL

Tabel 4 1 Penyeleksian Data	21
Tabel 4 2 Inisialisasi Centroid awal	23
Tabel 4 3 Hasil Pengelompokkan Data Cluster 0	36
Tabel 4 4 Hasil Pengelompokkan Data Cluster 1	38
Tabel 4 5 Hasil Pengelompokkan Data Cluster 2	39
Tabel 4 6 Hasil Akhir Pengelompokkan kinerja karyawan Cluster 0	60
Tabel 4 7 Hasil Akhir Penilaian Kinerja Karyawan Cluster 1	62
Tabel 4 8 Hasil Akhir Penilaian Kinerja Karyawan Cluster 2	63
Tabel 4 9 Perbandingan Hasil Akhir Pengelompokkan	64

BAB I

PENDAHULUAN

1.1 Latar Belakang

Untuk mencapai tujuan organisasi diperlukan karyawan yang sesuai dengan persyaratan dalam organisasi, dan juga harus mampu menjalankan tugas-tugas yang telah ditentukan oleh organisasi [1]. Karyawan yang memiliki kinerja tinggi cenderung bekerja lebih efisien, menghasilkan output yang lebih banyak, dan meningkatkan efektivitas operasional bagi perusahaan. Hal ini dapat membantu perusahaan mencapai target produksi, mengurangi biaya, dan meningkatkan keuntungan. Oleh karena itu, penting bagi perusahaan untuk memantau dan mengevaluasi kinerja karyawan mereka secara berkala.

Begitu juga pada PT. XYZ yang bergerak di bidang perkebunan kelapa sawit dan pengolahan minyak kelapa sawit. Perusahaan ini memiliki tujuan untuk menjadi industri pengolahan minyak kelapa sawit yang sehat dan inovatif serta mengutamakan mutu dan kelestarian lingkungan. Dalam upaya mencapai tujuan tersebut, PT. XYZ telah melaksanakan penilaian kinerja guna mengidentifikasi karyawan-karyawan yang memiliki kinerja yang baik. Penilaian kinerja adalah proses pengevaluasian kinerja, penyusunan rencana pengembangan, dan pengkomunikasian hasil proses tersebut kepada karyawan itu sendiri [1]. Dari hasil penilaian tersebut akan diperoleh berbagai wawasan dan informasi mengenai kinerja karyawan, di antaranya perusahaan bisa mendapat gambaran tentang sejauh mana karyawan mencapai target dan tujuan kerja yang ditetapkan, serta bagaimana karyawan berkinerja dalam memenuhi harapan perusahaan.

Namun dalam pelaksanaannya, PT. XYZ mengalami beberapa permasalahan dalam melakukan penilaian karyawan. Permasalahannya yaitu penilaian masih dilakukan secara subjektif berdasarkan preferensi individu manajer, sehingga penilaian yang dihasilkan kurang objektif. Ketidakobjektifan ini tentu akan berdampak pada keputusan perusahaan terkait pengelolaan kinerja karyawan. Penilaian juga masih dilakukan secara semi-manual, sehingga perusahaan kesulitan dalam mengidentifikasi pola kinerja yang berulang atau tren yang berkaitan dengan faktor-faktor tertentu karena jumlah data karyawan yang cukup banyak. Selain itu, belum adanya model prediktif yang diterapkan membuat perusahaan kesulitan dalam mengidentifikasi karyawan yang berpotensi untuk mendapatkan pengembangan karir.

Berdasarkan permasalahan tersebut, penerapan data mining clustering dalam proses penilaian kinerja karyawannya dapat menjadi solusi dari permasalahan pelaksanaan penilaian kinerja oleh PT. XYZ. Penelitian sebelumnya yang berkaitan dengan K-Means Clustering

telah banyak dilakukan seperti pada penelitian yang dilakukan oleh Devi Sartika dan Juju Jumadi [2] bahwa pengelompokan kinerja karyawan menghasilkan pengelompokan data 92 orang karyawan berdasarkan 3 kriteria yaitu sangat baik (C1) sebanyak 15 data, cukup baik (C2) sebanyak 44 data, dan kurang baik (C3) sebanyak 33 data.

Penelitian terkait clustering penilaian kinerja juga dilakukan oleh [3] dengan hasil penelitian beberapa kategori kinerja karyawan yaitu sangat produktif, cukup produktif dan kurang produktif dan hasil pengelompokan 0,42% untuk cluster 1 kategori sangat produktif, yang terdiri dari 16 data karyawan, 0,47% untuk cluster 2 kategori cukup produktif, yang terdiri dari 18 data karyawan, 0,11% untuk cluster 3 kategori kurang produktif, yang terdiri dari 4 data karyawan. Penelitian ini bertujuan untuk mengetahui produktivitas seorang karyawan dan keefektifan kinerja seorang karyawan dimasa yang akan datang.

Penelitian lainnya oleh [4] dengan metode clustering K-Means untuk mengelompokkan gaji karyawan berdasarkan faktor-faktor yang mempengaruhi kinerja karyawan dalam bekerja. Penelitian menghasilkan dua kategori gaji karyawan, yaitu bergaji kategori tinggi dan bergaji kategori rendah dimana masing-masing cluster memiliki nilai rata-rata yang berbeda.

Pada penelitian ini penulis tertarik melakukan penelitian mengenai pengklasteran penilaian kinerja karyawan sebagai solusi untuk mengatasi permasalahan dalam proses penilaian kinerja karyawan di PT. XYZ. Pengklasteran akan menggunakan algoritma K-Means, karena K-Means dapat mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok. Dengan algoritma K-Means, data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam kelompok yang lain [5] .

1.2 Rumusan Masalah

Perumusan masalah berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah yang didapat pada penelitian ini adalah bagaimana penerapan algoritma K-Means clustering pada data mining terhadap pengelompokan penilaian kinerja karyawan di PT.XYZ menggunakan bahasa pemogramann python.

1.3 Batasan Masalah

Batasan masalah yang diangkat dari penelitian ini adalah :

1. Penerapan data mining pada data hasil penilaian kinerja karyawan di PT. XYZ.
2. Menggunakan salah satu algoritma data mining, yaitu K-Means Clustering.

3. Pengolahan data menggunakan algoritma K-Means Clustering secara manual pada aplikasi Microsoft Excel dan bahasa pemrograman python pada Jupyter Notebook.
4. Hasil pengolahan data untuk mengelompokkan hasil kinerja karyawan kedalam beberapa cluster.

1.4 Tujuan

Tujuan dari penelitian ini adalah untuk membantu PT. XYZ dalam mengelompokkan hasil penilaian kinerja karyawan guna mengevaluasi kinerja karyawan untuk perbaikan dan pengambilan keputusan terkait hasil kinerja karyawan dalam satu periode penilaian.

1.5 Manfaat

Manfaat yang diperoleh dari penelitian ini adalah dengan penerapan data mining menggunakan algoritma K-Means Clustering pada mengelompokkan hasil penilaian kinerja karyawan PT. XYZ dapat meningkatkan objektivitas penilaian kinerja karyawan dan identifikasi pola kinerja karyawan berdasarkan model prediktif yang telah diterapkan, sehingga pengembangan kinerja karyawan menjadi lebih tepat dan efektif.

BAB II

LANDASAN TEORI

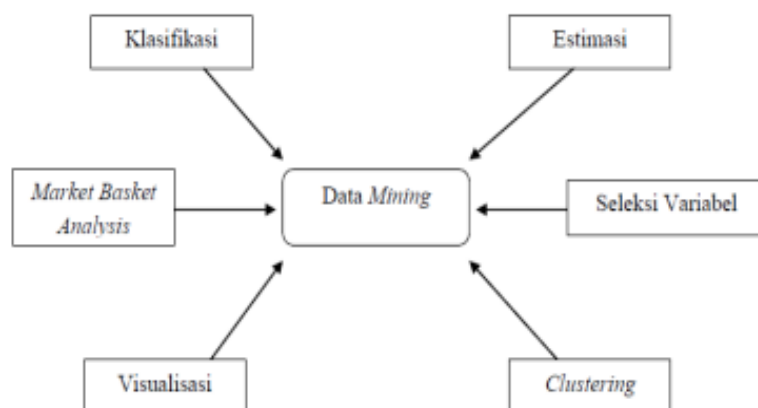
2.1 Tinjauan Pustaka

2.1.1 Pengertian Data Mining

Data mining adalah serangkaian proses untuk mengeksplorasi nilai tambah dari kumpulan data dalam bentuk pengetahuan yang tidak diketahui secara artifisial [6] Harus diingat bahwa kata mining itu sendiri berkonotasi dengan upaya mengekstraksi sejumlah kecil komoditas berharga dari sejumlah besar bahan dasar. Karena data mining sebenarnya berasal dari bidang ilmiah seperti kecerdasan buatan, pembelajaran mesin, statistik, dan basis data[7]. Dengan kata lain, data mining adalah proses penggalian pola dari data.

Data mining menjadi alat yang semakin penting untuk mengubah data menjadi informasi. Ini sering digunakan dalam berbagai praktik analitis seperti pemasaran, pengawasan, deteksi penipuan, dan penemuan ilmiah. Selama bertahun-tahun, bisnis, ilmuwan, dan pemerintah telah menggunakan penambangan data untuk mengekstrak data dalam jumlah besar, seperti catatan perjalanan penumpang maskapai, data sensus, dan data pemindai supermarket, untuk menghasilkan laporan riset pasar.

Data mining menjadi pusat dari beberapa kajian[8]. Diantaranya adalah estimasi, seleksi variabel, clustering, visualisasi, market basket analysis dan klasifikasi, Berikut gambaran pada gambar 2.1



Gambar 2 1 Kajian Umum Data Mining

2.1.2 Tujuan Data Mining

Adapun beberapa masalah yang sesuai dan dapat diselesaikan dengan penerapan teknik data mining yaitu:

1. Memerlukan keputusan yang bersifat *knowledge-based*

2. Mempunyai lingkungan yang berubah
3. Metode yang ada sekarang bersifat suboptimal.
4. Tersedia data yang bisa diakses, cukup dan relevan.
5. Memberikan keuntungan yang tinggi jika keputusan yang diambil tepat [9] .

Salah satu tujuan utama data mining adalah untuk mendapatkan hubungan atau pola yang dilihat antara dua atau lebih atribut dan dua atau lebih objek yang mungkin memberikan indikasi yang bermanfaat.

2.1.3 Manfaat Data Mining

Menurut Tan 2004 pemanfaatan dari data mining sendiri bisa dilihat dari dua sudut pandang, baik sudut pandang komersial dan sudut pandang keilmuan. Dari sudut pandang komersial, data mining bisa digunakan untuk menangani adanya peledakan dari volume data. Dengan melihat bagaimana menyimpannya, mengekstraknya dan memanfaatkannya. Tentunya berbagai ilmu komputasi dapat untuk menghasilkan informasi yang dibutuhkan.

Data mining juga digunakan untuk membantu dalam melakukan analisis koleksi pengamatan. Data mining memiliki manfaat yang signifikan dalam berbagai bidang, mulai dari bidang bisnis, kesehatan, keuangan, pemasaran, dan lain-lain. Pertama, data mining memungkinkan pengambilan keputusan yang lebih baik dan berbasis fakta. Dengan menganalisis data historis dan mengidentifikasi pola-pola yang tersembunyi, data mining dapat membantu mengungkap wawasan dan tren yang berharga. Hal ini memungkinkan organisasi untuk membuat keputusan yang lebih informasional, memperbaiki efisiensi operasional, mengoptimalkan strategi pemasaran, dan meningkatkan kepuasan pelanggan. Selain itu, data mining juga dapat digunakan untuk deteksi penipuan, pengelolaan risiko, prediksi pasar, dan pengembangan produk yang lebih baik. Dengan demikian, data mining berperan penting dalam mendukung pengambilan keputusan yang lebih cerdas dan berdampak positif pada keberhasilan organisasi.

2.1.4 Penerapan Data Mining

Penerapan data mining telah berhasil dalam berbagai konteks dan sektor. Contohnya, dalam industri retail, data mining digunakan untuk menganalisis pola pembelian pelanggan, memprediksi tren penjualan, dan mengatur stok barang secara efisien. Di bidang kesehatan, data mining dapat membantu mengidentifikasi faktor risiko penyakit, menganalisis efektivitas pengobatan, dan mendukung pengambilan keputusan medis. Dalam industri keuangan, data mining dapat digunakan untuk deteksi penipuan, analisis risiko kredit, dan perencanaan investasi. Bahkan, penerapan data mining juga

terlihat dalam analisis sosial media, di mana data mining digunakan untuk mengidentifikasi sentimen pelanggan, memprediksi tren pasar, dan melakukan pemetaan perilaku pengguna. Dengan kemampuannya untuk mengeksplorasi dan menganalisis data yang kompleks, data mining memberikan manfaat nyata dalam berbagai aspek kehidupan dan bisnis.

2.1.5 Tahap-tahap Data Mining

Berikut tahapan proses data mining [10]

1. Pemilihan data (*Data Selection*)

Perlu dilakukan sebelum memilih (selecting) data dari sekumpulan data operasional yaitu tahap penggalian informasi dalam data mining . Data yang dipilih akan digunakan dalam proses data mining, disimpan dalam file dan terpisah dari database operasional.

2. Pembersihan data (*Pre-processing*)

Proses pembersihan meliputi, antara lain, membersihkan duplikasi data, periksa data yang tidak konsisten dan perbaiki kesalahan dalam data, seperti kesalahan cetak. Ada juga proses pengayaan (enrichment), yaitu proses “memperkaya” data dimana data yang sudah ada dengan data atau informasi lain yang relevan dan dibutuhkan untuk data mining, seperti data atau informasi eksternal.

3. Transformasi (*Transformation*)

Coding adalah proses mengubah data yang dipilih sebelum diproses dalam data mining. Proses pengkodean dalam data mining adalah sebuah proses yang membutuhkan kreativitas sehingga sangat bergantung pada jenis atau pola informasi yang akan dicari di basis data.

4. Data Mining

Data mining adalah proses menemukan pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. teknik, metode atau algoritma dalam data mining sangat bervariasi. Sangat penting untuk memilih metode atau algoritma yang tepat tergantung pada tujuan dan keseluruhan proses data mining.

5. Evaluasi (*Evaluation*)

Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan di dalam format yang mudah dipahami oleh pihak yang berkepentingan. Tahap ini adalah bagian dari proses data mining yang disebut interpretasi. Tahap ini meliputi periksa pola atau informasi yang ditemukan yang bertentangan dengan fakta atau asumsi yang sudah ada sebelumnya.

2.1.6 Fungsi Data Mining

Fungsi umum dalam data mining yaitu :

1. Association, adalah proses untuk menemukan aturan asosiasi Antara suatu kombinasi item dalam suatu waktu.
2. Sequence, adalah proses untuk menentukan aturan asosiasi Antara suatu kombinasi item dalam suatu waktu dan diterapkan lebih dari satu periode.
3. Clustering, adalah proses pengelompokan sejumlah data/objek ke dalam kelompok data sehingga setiap kelompok berisi data yang mirip.
4. Classification, adalah proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.
5. Regression, adalah proses pemetaan data dalam suatu nilai prediksi.
6. Forecasting, adalah proses pengestimasian nilai prediksi berdasarkan pola-pola di dalam sekumpulan data.
7. Solution, adalah proses penemuan akar masalah dan problem solving dari persoalan bisnis yang diharapkan atau paling tidak sebagai informasi dalam pengambilan keputusan. [11]

2.1.7 Metode Clustering

Clustering atau pengklasteran merupakan salah satu metode data mining dengan tujuan untuk mengelompokkan data dengan karakteristik yang sama di suatu ‘wilayah’ yang sama dengan karakteristik yang berbeda di ‘wilayah’ lain. Proses clustering dimana pengelompokan objek data ke dalam dua atau lebih cluster yang mempunyai tingkat kemiripan yang tinggi dan dengan data cluster berbeda tidak memiliki kemiripan dengan objek lainnya.

Algoritma clustering digunakan untuk mengorganisasikan dan mengkategorikan data , namun juga bermanfaat untuk kompresi data dan konstruksi model. Dengan adanya pencarian kesamaan dalam data, data yang sama dapat direpresentasikan dengan lebih sedikit simbol. Dan juga, dengan adanya pengelompokkan data, kita dapat membangun sebuah model masalah berdasarkan pengelompokkan-pengelompokkan tersebut[12]

2.1.8 Algoritma K-Means

Dari beberapa teknik clustering yang paling sederhana dan umum dikenal adalah clustering K-Means. Algoritma k-means adalah algoritma sederhana dengan jenis clustering data non hirarki yang membagi data ke dalam beberapa kelompok dimana

banyak digunakan karena mudah untuk diterapkan dan menggunakan waktu yang relatif singkat.

Penggunaan algoritma dalam proses clustering tergantung dari data yang ada konklusi yang ingin dicapai. Berikut algoritma K-Means yang terdapat didalamnya aturan yaitu :

1. Jumlah Cluster perlu diinputkan
2. Hanya memiliki atribut yang bertipe numerik.

Algoritma k-means memilih secara acak k buah data sebagai centroid, lalu jarak antara data dan centroid dihitung dengan menggunakan : Euclidean Distance. Data dikelompokkan berdasarkan jarak minimum dari titik semua cluster.

Untuk algoritma *K-Means* dijelaskan sebagai berikut :

1. Memasukkan data
2. Menentukan jumlah *cluster*
3. Mengambil data secara acak sesuai dengan jumlah *cluster* yang akan digunakan untuk pusat klaster (*centroid*)
4. Menghitung jarak antara data dengan pusat *cluster* dengan menggunakan persamaan di bawah ini.

$$D_{ij} = \sqrt{(x_{1j} - v_{1j})^2 + \dots + (x_{ij} - v_{kj})^2}$$

dimana :

D_{ij} : jarak data ke i ke pusat *cluster* ke k

x_{ij} : data ke i , pada atribut ke j

v_{ij} : titik pusat ke k , pada atribut ke j .

5. Menghitung pusat *cluster* dengan anggota *cluster* yang baru
6. Jika pusat *cluster* tidak berubah maka proses *cluster* sudah selesai tetapi jika pusat *cluster* masih berubah maka ulangi langkah menghitung jarak pada langkah ke empat sampai pusat *cluster* tidak berubah lagi.

2.2 Tinjauan Organisasi Penelitian

2.2.1 Latar Belakang Perusahaan

PT. XYZ merupakan perusahaan swasta lokal di daerah Salareh Aia, Kecamatan Palembayan, Kabupaten Agam, Sumatera Barat yang bergerak di bidang perkebunan kelapa sawit dan pengolahan minyak kelapa sawit. Perusahaan ini termasuk perusahaan dengan level menengah ke bawah dengan fokus usahanya terletak pada pengelolaan usaha

perkebunan kelapa sawit dan pengolahan produk minyak kelapa sawit. PT. XYZ memiliki tujuan perusahaan untuk menjadi industri pengolahan minyak kelapa sawit yang sehat dan inovatif serta mengutamakan mutu dan kelestarian lingkungan.

Sebagai perusahaan manufaktur PT. XYZ mempunyai lini bisnis mulai dari operasional di plantation (perkebunan) sampai proses produksi bahan mentah di pabrik. Produk utama yang berada dalam core bisnis perusahaan ini, yaitu :

1. Tandan buah segar (Fresh fruit bunch)

Tandan buah segar adalah produk bahan mentah yang dihasilkan oleh unit bisnis di plantation. Tandan buah segar sebagai bahan baku yang akan di proses di pabrik.

2. Minyak mentah sawit (Crude palm oil)

Minyak mentah adalah keluaran produk dari pengolahan tandan buah segar yang juga merupakan inti bisnis perusahaan ini.

Dengan level perusahaan menengah ke bawah, pangsa pasar dari produk minyak kelapa sawit PT. XYZ merupakan pasar nasional, dengan pemasaran secara partai besar maupun partai kecil. Untuk itu, PT. XYZ sedang mengupayakan pengembangan yang berkelanjutan dengan fokus pada ekspansi pasar. Keuntungan yang diharapkan dari upaya tersebut ialah agar dapat memperluas distribusi produk dan memperluas jejaring mitra perusahaan.

Melalui investasi dalam teknologi, SDM, dan infrastruktur, PT. XYZ berkomitmen untuk meningkatkan efisiensi operasional dan kemampuan produksinya. Untuk mencapai hal tersebut, PT. XYZ juga mengusahakan peningkatan mutu SDM yang menjadi bagian dari perusahaan tersebut dengan melaksanakan program pelatihan dan pengembangan SDM agar menghasilkan SDM yang kompeten dan memiliki kinerja yang baik.

2.2.2 Latar Belakang Masalah Perusahaan

Salah satu faktor penggerak utama roda bisnis suatu perusahaan adalah karyawan atau SDM, sehingga apabila kualitas SDM dalam suatu perusahaan terjaga dengan baik, hal ini dapat memberikan dampak positif terhadap peningkatan produktivitas operasi perusahaan. Dimana, SDM yang berkualitas merupakan karyawan yang memiliki pengetahuan, dan kompetensi yang sesuai dengan tuntutan pekerjaan. Selain itu kualitas SDM perusahaan juga dapat dinilai berdasarkan kontribusinya dalam berinovasi di perusahaan, seperti mengidentifikasi peluang baru.

Untuk mengetahui dan mengevaluasi kinerja karyawan di PT. XYZ maka dilakukan penilaian kinerja karyawan secara berkala. Dari hasil penilaian tersebut akan diperoleh berbagai wawasan dan informasi mengenai kinerja karyawan, di antaranya

perusahaan bisa mendapat gambaran tentang sejauh mana karyawan mencapai target dan tujuan kerja yang ditetapkan, serta bagaimana karyawan berkinerja dalam memenuhi harapan perusahaan. Selain itu, dari hasil penilaian kinerja perusahaan juga dapat membuat pertimbangan untuk menemukan potensi pengembangan karir karyawan. Dimana nantinya akan memungkinkan perusahaan untuk mengenali karyawan yang memiliki potensi untuk tanggung jawab yang lebih besar. Dengan demikian, perusahaan dapat merencanakan program pengembangan dan pelatihan yang sesuai untuk membantu karyawan mencapai potensi penuh mereka serta menindaklanjuti karyawan yang tidak mencapai target dan tujuan kerja perusahaan.

Agar dapat memperoleh wawasan dan informasi tersebut, maka diperlukan teknik data mining. Dengan menggunakan teknik data mining, perusahaan dapat menggali dan menganalisis data yang terkait dengan kinerja karyawan untuk mendapatkan wawasan yang lebih mendalam dengan lebih akurat dan efisien dari segi waktu. Namun, PT. XYZ belum menerapkan teknik data mining pada penilaian kinerja karyawannya, sehingga ditemukan beberapa permasalahan, di antaranya :

1. Kurangnya informasi atau wawasan yang diperoleh dari data hasil penilaian kinerja yang akan berdampak pada pengambilan keputusan penilaian kinerja yang kurang objektif.
2. Penilaian kinerja karyawan didasarkan pada persepsi subjektif atau preferensi individu manajer, sehingga menyebabkan ketidakadilan dalam penilaian karyawan karena tidak didasarkan pada data yang terukur.
3. Perusahaan kesulitan dalam mengidentifikasi pola kinerja yang berulang atau tren yang berkaitan dengan faktor-faktor tertentu karena jumlah data karyawan yang cukup banyak. Hal ini dapat menghambat kemampuan perusahaan untuk membuat perubahan yang efektif dalam penilaian kinerja dan pengembangan karyawan.
4. Perusahaan kesulitan dalam mengidentifikasi karyawan yang berpotensi untuk mendapatkan pengembangan karir karena belum adanya model prediktif yang diterapkan.

Sehingga, dapat disimpulkan bahwa PT. XYZ membutuhkan penerapan data mining dalam proses penilaian kinerja karyawannya. Data mining memungkinkan analisis cepat dan efisien terhadap data kinerja karyawan. Algoritma dan teknik data mining dapat digunakan untuk mengidentifikasi pola, hubungan, dan tren yang relevan dalam data kinerja. Analisis ini dapat membantu menghasilkan wawasan yang berarti dengan lebih efisien daripada metode manual seperti menganalisis data melalui Microsoft Excel yang

membutuhkan waktu lebih lama. Kemudian, data mining dapat digunakan untuk mengotomatisasi beberapa aspek dalam proses penilaian kinerja. Misalnya, PT. XYZ dapat menggunakan algoritma data mining untuk menghasilkan klasifikasi karyawan berdasarkan data kinerja, sehingga mengurangi ketergantungan pada penilaian subjektif yang memiliki bias dan membutuhkan waktu lebih banyak.

2.2.3 Batasan Masalah Penelitian Perusahaan

Berdasarkan rincian permasalahan yang sudah dijelaskan, maka permasalahan yang diangkat pada laporan ini, yaitu penerapan data mining dalam penilaian kinerja karyawan di PT. XYZ. Dimana algoritma data mining yang akan digunakan merupakan algoritma K-Means Clustering. Data hasil penilaian kinerja karyawan akan diolah dan dianalisis menggunakan algoritma tersebut hingga ditemukan tiga cluster atau pengelompokkan karyawan berdasarkan penilaiannya tersebut, yaitu (cluster 1,2,3). Kemudian, dari hasil klasterisasi tersebut PT. XYZ dapat mengambil upaya tindak lanjut, seperti merancang program pengembangan khusus untuk setiap cluster, memberikan umpan balik yang sesuai, atau mengidentifikasi area peningkatan kinerja yang spesifik untuk setiap kelompok karyawan.

BAB III

METODOLOGI

3.1 Metode Penelitian

Metode yang digunakan dalam laporan ini adalah metode deskriptif kuantitatif. Metode deskriptif kuantitatif adalah suatu metode yang bertujuan untuk membuat gambar atau deskriptif tentang suatu keadaan secara objektif yang menggunakan angka, mulai dari pengumpulan data, penafsiran terhadap data tersebut serta penampilan dan hasilnya [13].

Penelitian deskriptif kuantitatif adalah penelitian yang hanya menggambarkan isi suatu variabel dalam penelitian, tidak dimaksudkan untuk menguji hipotesis tertentu [14]. Dengan demikian dapat diketahui bahwa penelitian deskriptif kuantitatif adalah penelitian yang menggambarkan, mengkaji dan menjelaskan suatu fenomena dengan data (angka) apa adanya tanpa bermaksud menguji suatu hipotesis tertentu.

Metode ini tepat untuk digunakan dalam analisis *clustering* dengan algoritma K-Means karena metode ini menggunakan data numerik atau kuantitatif untuk menggambarkan dan menganalisis fenomena atau variabel yang sedang diteliti. Dalam tahapan pada metode ini dalam proses *clustering*, setelah dilakukan analisis clustering menggunakan algoritma K-means, kemudian hasilnya akan dianalisis dengan metode deskriptif kuantitatif seperti menghitung statistik deskriptif (misalnya, rata-rata, standar deviasi) atau menggambarkan hasil tersebut dalam bentuk grafik atau tabel yang menunjukkan kelompok-kelompok yang terbentuk.

3.2 Tahap Pengumpulan Data

Pada bagian tahap pengumpulan data, akan dijelaskan proses yang dilakukan dalam pengumpulan data yang relevan untuk laporan ini. Tahap ini merupakan langkah penting dalam analisis *clustering* menggunakan algoritma K-means, karena data yang dikumpulkan akan menjadi dasar untuk analisis clustering yang dilakukan.

3.2.1 Lokasi dan Waktu Pengumpulan Data

Pengumpulan data dilakukan di PT. XYZ yang beralamat di Salareh Aia, Kecamatan Palembayan, Kabupaten Agam, Sumatera Barat. Adapun pengumpulan data ini dilakukan pada tanggal 4 Mei 2023 - 8 Mei 2023 dengan judul laporan “Implementasi Algoritma K-Means dalam Proses Clustering Hasil Penilaian Kinerja Karyawan di PT. XYZ”

3.2.2 Kebutuhan Laporan

Dalam laporan ini kebutuhan yang diperlukan untuk menjadi bahan analisis serta pengolahan data adalah :

1. Data hasil penilaian kinerja karyawan di PT. XYZ
2. Perusahaan PT. XYZ sebagai penyedia data
3. Microsoft Excel sebagai alat pengolahan data manual
4. Jupyter Notebook sebagai alat pengolahan data mining

3.2.3 Metode Pengumpulan Data

Adapun metode yang digunakan dalam pengumpulan data untuk laporan ini adalah sebagai berikut :

1. Wawancara

Wawancara adalah metode pengumpulan data dengan cara melakukan tanya jawab secara langsung antara pewawancara dengan narasumber atau pihak yang terkait. Pada laporan ini, dilakukan wawancara dengan seorang narasumber dari Departemen Human Resource di PT. XYZ

2. Dokumentasi

Dalam pengumpulan data ini, juga dilakukan teknik dokumentasi dengan melihat sejumlah arsip data PT. XYZ yang diberikan oleh narasumber guna memenuhi kebutuhan analisis. Dokumentasi yang didapatkan berupa data rekapan hasil penilaian kinerja karyawan dalam satu periode penilaian di PT. XYZ

3. Studi Literatur

Untuk mendukung proses analisis dan pengolahan data yang didapat, maka dibutuhkan data - data yang bersifat teoritis terkait metode yang digunakan. Sehingga, dilakukan pengumpulan data melalui studi literatur dengan cara mereview dan mempelajari beberapa artikel jurnal, makalah, dan referensi lain yang terkait dengan permasalahan yang diangkat serta metode yang digunakan.

3.3 Tahapan Pembuatan Laporan

Tahapan pembuatan laporan menggambarkan proses pembuatan laporan yang akan dilaksanakan serta menggambarkan prosesnya secara keseluruhan. Berikut tahapan pembuatan laporan yang akan dilaksanakan :

1. Persiapan

Di tahap ini dilakukan penyusunan rencana pembuatan laporan dan perancangan *timeline* pembuatan laporan, termasuk menentukan objek, metode serta algoritma yang digunakan.

2. Studi Literatur

Di tahap ini dilakukan kajian dan studi literatur mengenai permasalahan yang diangkat, metode, serta algoritma yang digunakan.

3. Pengumpulan Data

Di tahap ini dilakukan melalui wawancara, dokumentasi, dan studi literatur.

4. Preprocessing Data

Di tahap ini dilakukan pengolahan awal termasuk *data reduction* untuk mengurangi data yang tidak dibutuhkan dan *data cleaning* untuk mengisi missing value atau yang kurang pas, mengoreksi data yang tidak konsisten, dan mendeteksi redundansi data.

5. Pengolahan Data Manual

Tahap ini dilakukan menggunakan Microsoft Excel agar di akhir dapat dibandingkan keakuratannya dengan hasil pengolahan data menggunakan *Python* di Jupyter Notebook.

6. Pengujian Cluster

Di tahap ini dilakukan pengujian cluster menggunakan metode *Silhouette Coefficient*.

7. Visualisasi Hasil Cluster

Di tahap ini hasil dari tahap *clustering* akan divisualisasikan dalam bentuk *scatter plot* untuk melihat penyebaran dan pengelompokan dari masing - masing cluster.

8. Hasil dan Pembahasan

Pada tahapan ini akan dijelaskan mengenai hasil dari proses data mining yang dilakukan menggunakan analisis *clustering* dengan Algoritma K-Means.

9. Kesimpulan dan Saran

Di tahap ini dibuat kesimpulan dari hasil yang didapatkan dan memberikan saran untuk pihak PT. XYZ terkait dengan permasalahan yang diangkat serta pembuatan laporan selanjutnya.

3.4 Tahap Preprocessing Data

Data Preprocessing merupakan salah satu tahapan penting dalam melakukan mining data. Sebelum menuju ke tahap pemrosesan, data mentah akan diolah terlebih dahulu, termasuk mengeliminasi data yang tidak sesuai. Preprocessing data seringkali digunakan untuk

mengurangi kesalahan data dan sistematis bias dalam data mentah sebelum analisis apapun terjadi [15]. Tujuan dilakukan preprocessing dalam data mining adalah untuk mentransformasi data ke suatu format yang prosesnya lebih mudah dan efektif untuk kebutuhan pemakai.

Adapun tahap-tahap dalam melakukan preprocessing data dalam data mining, sebagai berikut:

1. Data Cleaning

Tahap ini melibatkan identifikasi dan penanganan data yang tidak lengkap, data yang hilang, atau data yang tidak valid dari data mentah yang akan digunakan.

2. Data Integration

Apabila data yang digunakan berasal dari multiple data sources, langkah ini melibatkan menggabungkan data dari berbagai sumber menjadi satu dataset yang terintegrasi. Pada tahap ini menangani perbedaan skema data, format data, dan resolusi referensi yang berbeda.

3. Transformasi Data

Pada tahap ini, dilakukan transformasi pada data untuk memperbaiki distribusi data, mengurangi noise, atau mengubah skala data. Contohnya, dapat dilakukan normalisasi data maupun pengurangan dimensi data.

4. Data Reduction

Tahap ini dilakukan untuk mengurangi jumlah fitur dan menghilangkan fitur yang tidak memberikan kontribusi signifikan pada analisis dengan cara mengurangi sampel data yang diambil. Ada tiga teknik yang bisa diterapkan saat melakukan data reduction, yakni dimensionality reduction (pengurangan dimensi), numerosity reduction (pengurangan jumlah), dan data compression (kompresi data).

3.5 Tahap Klasifikasi Data

Bagian klasifikasi merupakan bagian tahap pembentukan klasifikasi dalam data mining. Klasifikasi adalah proses pengelompokan objek atau instance data ke dalam kelas-kelas yang telah ditentukan berdasarkan atribut-atributnya. Klasifikasi berguna untuk mengenali pola atau hubungan dalam data dan membuat prediksi pada data baru.

Pada laporan ini menggunakan algoritma K-Means Clustering sebagai algoritma klasifikasinya. K-Means Clustering adalah metode untuk mengkategorikan atau pengelompokan sekelompok objek sesuai dengan atribut yang sama atau karakteristik ke dalam sejumlah groups (jumlah bilangan bulat positif). Ini mendefinisikan sebuah cluster oleh massa yang yang mewakili mean dari cluster.

Hal ini sesuai dengan permasalahan yang diangkat, yakni identifikasi klasifikasi hasil penilaian kinerja karyawan di PT. XYZ, di antaranya kinerja sangat baik, baik, dan kurang baik. Dimana metode data mining dengan algoritma K-Means clustering cocok untuk mengumpulkan dan mengelompokkan kinerja karyawan tersebut menjadi beberapa kategori atau kelompok, yaitu kinerja sangat baik, kinerja baik, dan kinerja kurang baik.

Oleh karena itu pada laporan ini, digunakan suatu metode pengolahan data dengan mengelompokkan data menggunakan algoritma K-Means Clustering dari hasil penilaian kinerja karyawan di PT. XYZ. Hasil observasi dibandingkan dengan identifikasi menggunakan algoritma K-Means clustering, sehingga diharapkan agar dapat melakukan tindak lanjut terhadap masing-masing kelompok karyawan dengan lebih bijak dan meminimalisir penilaian yang subjektif.

HASIL DAN PEMBAHASAN

Pada penelitian ini data penilaian kinerja karyawan yang ada di PT. XYZ yang telah dikumpulkan kemudian diolah untuk menghasilkan pengelompokan data yang konsisten, relevan, dan dapat memperbaiki kesalahan data. Data yang diperoleh ditransformasikan untuk dapat diproses lebih lanjut dari bentuk asalnya kedalam bentuk yang sesuai untuk dikelompokkan. Tujuan dari pengolahan data ini ialah supaya data yang sudah diperoleh mempunyai makna agar dapat menarik kesimpulan dalam pengelompokan hasil penilaian kinerja karyawan di PT. XYZ.

4.1 Pengumpulan Data

Sumber data utama yang digunakan dalam penelitian ini adalah dataset penilaian kinerja karyawan di PT. XYZ dengan jumlah keseluruhan data 128 *record* dari keseluruhan total karyawan yang ada pada PT tersebut. Data yang diterima dalam bentuk format .xlsx atau *microsoft excel* sehingga mempermudah dalam melakukan pembersihan dan pemfilteran data. Data yang diperoleh sudah melalui perizinan kepada pihak PT. XYZ dibawah wewenang manajer pimpinan PT tersebut.

[illegible]

Gambar 4.1 Data Hasil Penilaian Kinerja Karyawan PT.XYZ

4.2 Pre-Processing Data

Pre-Processing Data adalah tahapan dari Data Mining yaitu suatu proses atau tahapan yang dilakukan untuk mengolah data mentah menjadi data yang berkualitas yang baik dilanjutkan ke proses selanjutnya. Tahapan preprocessing data sebagai berikut:

4.2.1 Data Reduction

Data Reduction adalah proses untuk mengurangi atau mereduksi sejumlah data yang tidak dibutuhkan. Data Reduction sangat berguna untuk mendapatkan item data yang akan digunakan dalam penelitian ini.

Data penilaian karyawan yang telah diperoleh akan disaring terlebih dahulu dan diambil beberapa item data yang diperlukan dalam proses pengolahan data. Terdapat tiga item data yang digunakan dalam proses pengolahan data nantinya, yaitu NIK, index performance, dan index potensial. Dimana, item data NIK berguna sebagai item data yang membedakannya dengan item data lain. Kemudian item data index performance yang merupakan hasil rata-rata dari penilaian hasil kerja yang terdiri dari kuantitas kerja, ketepatan kerja, dan kualitas kerja. Terakhir item data index potensial yang merupakan hasil rata-rata dari penilaian perilaku kerja yang terdiri dari pengetahuan tentang pekerjaan, komunikasi, kerjasama tim, dan disiplin kerja.

No	NIK	INDEX	
		Peformance	Potensial
1	009AMP18068708	33	39
2	327AMP10037607	24	22
3	002AMP13077908	27	39
4	GA/GMP/0604/403	24	28
5	307AMP09087102	27	33
6	002AMP14069207	27	33
7	324AMP11028703	24	30
8	0493AMP03057303	21	25
9	313AMP01028104	18	25
10	180AMP06067908	18	14
11	329AMP08057901	24	25
12	0431AMP97036608	39	47
13	322AMP11027610	18	22
14	305AMP01117902	27	27
15	177AMP06067308	21	30
16	317AMP02098010	21	52
17	321AMP11028204	27	33
18	179AMP06068109	18	22
19	0483AMP01048303	18	22

20	313AMP11028207	18	22
21	319AMP05078006	33	44
22	324AMP10098608	30	33
23	322AMP05078407	24	31
24	323AMP11029004	27	33
25	159AMP07016903	27	30
26	325AMP11026702	18	25
27	328AMP08108012	21	39
28	320AMP11028709	18	25
29	310AMP00118201	33	39
30	0495AMP01097805	27	33
31	0471AMP00057605	27	33
32	AG/PAMP/0119/183	18	22
33	0490AMP03057607	36	44
34	169AMP03058009	33	44
35	0519AMP01048106	33	41
36	320AMP03058205	21	22
37	326AMP05078010	27	33
38	0485AMP00097403	30	39
39	323AMP05078008	30	36
40	0473AMP00087103	27	33
41	0468AMP02077212	27	30
42	0476AMP00087702	33	44
43	157AMP99017712	30	41
44	320AMP02098512	18	25
45	0455AMP96117312	18	25
46	175AMP06128207	27	33
47	0677AMP00086412	36	44
48	319AMP11028710	36	44
49	331AMP05078703	18	22
50	0475AMP00057611	27	33
51	326AMP08107710	27	38
52	0437AMP96106312	18	22
53	0422AMP00086912	30	36
54	0424AMP97036811	27	36

55	0488AMP02077810	27	33
56	327AMP08107905	27	41
57	310AMP11027706	18	22
58	0465AMP99017406	18	14
59	0478AMP00097404	27	33
60	162AMP99117412	27	25
61	325AMP05078303	27	39
62	178AMP06068003	27	39
63	309AMP11028305	18	44
64	324AMP05098208	27	33
65	181AMP06048507	33	36
66	323AMP08058410	33	35
67	160AMP07048611	27	31
68	308AMP11028912	30	38
69	176AMP06088112	27	33
70	319AMP02097901	27	33
71	161AMP07048903	27	33
72	327AMP12028904	24	25
73	307AMP11028402	27	33
74	317AMP11028705	27	36
75	0421AMP00097902	27	33
76	174AMP06118011	27	33
77	312AMP11028607	30	33
78	308AMP03058201	27	33
79	320AMP05078505	30	36
80	0457AMP98077707	27	33
81	319AMP03057904	27	33
82	AG/PAMP/0119/189	24	27
83	325AMP08108208	27	33
84	318AMP11028503	27	42
85	0523AMP96107111	27	33
86	173AMP06067908	27	42
87	327AMP11028504	27	33
88	332AMP05078408	27	27
89	329AMP05088507	33	33

90	0430AMP98126712	30	36
91	311AMP11029004	27	42
92	155AMP98037611	27	25
93	311AMP02037810	21	28
94	333AMP05077710	18	30
95	326AMP11028902	21	27
96	0521AMP01047806	18	27
97	330AMP08087908	27	28
98	322AMP97108001	18	31
99	0480AMP00097206	36	44
100	311AMP00027307	36	44
101	167AMP99097508	27	33
102	318AMP03038106	21	36
103	320AMP08088409	30	33
104	327AMP05078002	18	36
105	0482AMP00097304	27	33
106	309AMP01098005	27	33
107	0497AMP04128308	18	31
108	0479AMP00098113	27	33
109	0526AMP02077411	36	47
110	0411AMP99036612	18	30
111	330AMP05078512	27	33
112	314AMP05077702	36	44
113	302AMP04127906	30	36
114	304AMP05098401	18	30
115	0466AMP00058212	36	44
116	321AMP05048108	21	36
117	0524AMP00087707	30	35
118	0522AMP96056712	27	36
119	0427GMP00088108	33	33
120	003PMJ02017208	18	22
121	0410AMP01067604	18	42

Tabel 4 1 Penyeleksian Data

4.2.3 Data Cleaning

Setelah data dipilih dan diseleksi sesuai dengan item data yang digunakan, maka dilakukan proses *cleaning* data agar tidak ada duplikasi data, missing value, data outliers, dan lainnya. Data cleaning adalah proses untuk mengisi missing value atau yang kurang pas, mengoreksi data yang tidak konsisten, dan mendeteksi redudansi data yaitu penumpukan data atau duplikasi data.

Dari data yang diperoleh ada dua item data yang mengalami missing value, yaitu pendidikan dan keterangan. Namun, dalam proses pengolahan data kedua item data tersebut tidak digunakan. Karena item data yang digunakan dalam pengolahan data yaitu, NIK, index performance dan Index potensial, sehingga proses cleaning data tidak dilakukan terhadap data yang telah di seleksi. Data dapat langsung diolah dan dilakukan proses data mining menggunakan metode Clustering dengan algoritma K-Means.

4.3 Perhitungan Manual Algoritma K-Means Clustering

Penerapan Algoritma K-Means Clustering pada proses manual dalam pengelompokan data hasil penilaian kinerja karyawan menggunakan aplikasi Microsoft Excel 2021. Pada pengujian data secara manual ini, menggunakan sampel data sebagian dari 121 records data yang diperoleh. Adapun tahapan yang diperlukan adalah sebagai berikut :

1. Menentukan Jumlah Cluster

Jumlah cluster yang ditetapkan pada pengelompokkan hasil penilaian kinerja karyawan adalah sebanyak 3 cluster yaitu diantaranya sangat baik, baik, kurang baik berdasarkan data hasil penilaian kinerja karyawan PT. XYZ.

2. Menentukan Nilai Titik Pusat Awal Cluster (Centroid)

Titik pusat cluster atau juga disebut sebagai centroid, digunakan sebagai nilai pengurang untuk penghitungan jarak antara data ke setiap cluster atau disebut distance. Dalam proses ini penentuan titik pusat awal bias ditentukan secara acak sesuai keinginan peneliti dengan syarat nilai centroid masih dalam range nilai data dari setiap item data .

Inisialisasi Centroids 0 merupakan nilai rata-rata dari item data karyawan ke 66 dan 68, Centroids 1 merupakan nilai rata-rata dari item data karyawan ke 27 dan 33, dan dari item data karyawan ke 1 dan 3. Berikut centroid awal pada setiap item data :

Titik Pusat	Index Performance	Index Potensial
Centroids 0	25.5	32
Centroids 1	28.5	41.5
Centroids 2	30	39

Tabel 4.2 Inisialisasi Centroid awal

3. Menghitung Jarak Setiap Data Points ke Centroids Pertama

Setelah titik pusat awal cluster telah ditentukan, proses selanjutnya adalah menghitung jarak setiap cluster atau distance. Dalam menghitung nilai jarak setiap cluster, peneliti menggunakan rumus Euclidean Distance. Adapun rumus tersebut adalah sebagai berikut :

$$\text{Distance} = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + (x_{ni} - x_{nj})^2}$$

Keterangan :

- x_{1i} = Data ke i pada atribut data 1
- x_{2j} = Data ke j pada atribut data 1
- Distance = jarak antar fasilitas i dan j

Perhitungan Jarak Centroids Awal

$$\begin{aligned}
 S1|66\ 68 &= ((33 - 25.5)^2 + (39 - 32)^2) = 105.25 \\
 S1|27\ 33 &= ((33 - 28.5)^2 + (39 - 41.5)^2) = 26.5 \\
 S1|1\ 3 &= ((33 - 30)^2 + (39 - 39)^2) = 9 \\
 S2|66\ 68 &= ((24 - 25.5)^2 + (22 - 32)^2) = 102.25 \\
 S2|27\ 33 &= ((24 - 28.5)^2 + (22 - 41.5)^2) = 400.5 \\
 S2|1\ 3 &= ((24 - 30)^2 + (22 - 39)^2) = 325 \\
 S3|66\ 68 &= ((27 - 25.5)^2 + (39 - 32)^2) = 51.25 \\
 S3|27\ 33 &= ((27 - 28.5)^2 + (39 - 41.5)^2) = 8.5 \\
 S3|1\ 3 &= ((27 - 30)^2 + (39 - 39)^2) = 9 \\
 S4|66\ 68 &= ((24 - 25.5)^2 + (28 - 32)^2) = 18.25 \\
 S4|27\ 33 &= ((24 - 28.5)^2 + (28 - 41.5)^2) = 202.5 \\
 S4|1\ 3 &= ((24 - 30)^2 + (28 - 39)^2) = 157 \\
 S5|66\ 68 &= ((27 - 25.5)^2 + (33 - 32)^2) = 13.25 \\
 S5|27\ 33 &= ((27 - 28.5)^2 + (33 - 41.5)^2) = 74.5 \\
 S5|1\ 3 &= ((27 - 30)^2 + (33 - 39)^2) = 45
 \end{aligned}$$

Dst

No	NIK	INDEX		Karyawan 66 68	Karyawan 27 33	Karyawan 1 3
		66 68	27 33			
1	004WFL0068708	33	39	105,25	26,5	9
2	327WFL0017607	24	22	102,25	400,5	325
3	002WFL1077908	27	39	51,25	8,5	9
4	01/00P/0604/403	24	28	18,25	202,5	157
5	307WFL0087102	27	33	3,25	74,5	45
6	002WFL1069207	27	33	3,25	74,5	45
7	324WFL1028703	24	30	6,25	152,5	117
8	0493WFL01057103	21	25	69,25	328,5	277
9	313WFL1028104	18	25	105,25	382,5	340
10	180WFL06067908	18	14	380,25	866,5	769
11	329WFL00807901	24	25	51,25	292,5	232
12	0431WFL0066608	39	47	407,25	140,5	145
13	322WFL1027610	18	22	156,25	490,5	433
14	105WFL117902	27	27	27,25	212,5	153
15	177WFL06067308	21	30	24,25	188,5	162
16	317WFL0088010	21	52	420,25	166,5	250
17	321WFL1028204	27	33	3,25	74,5	45
18	179WFL06068109	18	22	156,25	490,5	433
19	0493WFL01048103	18	22	156,25	490,5	433
20	313WFL1028207	18	22	156,25	490,5	433
21	319WFL05078006	33	44	200,25	26,5	34
22	324WFL10098608	30	33	21,25	74,5	36
23	322WFL05078407	24	31	3,25	130,5	100
24	323WFL1029004	27	33	3,25	74,5	45
25	159WFL0106903	27	30	6,25	134,5	90
26	325WFL1026702	18	25	105,25	382,5	340
27	328WFL108012	21	39	69,25	62,5	81
28	320WFL1028709	18	25	105,25	382,5	340
29	310WFL0118201	33	39	105,25	26,5	9
30	0495WFL1027805	27	33	3,25	74,5	45
31	0471WFL00807605	27	33	3,25	74,5	45
32	16/08WFL0119/183	18	22	156,25	490,5	433
33	0490WFL0057607	36	44	254,25	62,5	61
34	169WFL0058009	33	44	200,25	26,5	34
35	0519WFL1048106	33	41	137,25	20,5	13
36	320WFL0058205	21	22	120,25	436,5	370
37	326WFL05078010	27	33	3,25	74,5	45
38	0495WFL0087403	30	39	69,25	8,5	0
39	323WFL05078008	30	36	36,25	32,5	9
40	0473WFL00807103	27	33	3,25	74,5	45
41	0468WFL0207712	27	30	6,25	134,5	90
42	0476WFL00807702	33	44	200,25	26,5	34
43	167WFL009017712	30	41	101,25	2,5	4
44	320WFL02098512	18	25	105,25	382,5	340
45	0455WFL06117312	18	25	105,25	382,5	340
46	175WFL0138207	27	33	3,25	74,5	45
47	0677WFL00808412	36	44	254,25	62,5	61
48	319WFL1028710	36	44	254,25	62,5	61
49	331WFL05078703	18	22	156,25	490,5	433
50	0475WFL00807611	27	33	3,25	74,5	45
51	326WFL08107710	27	38	38,25	14,5	10
52	0437WFL06106112	18	22	156,25	490,5	433
53	0422WFL00808612	30	36	36,25	32,5	9
54	0424WFL07036911	27	36	18,25	32,5	18
55	0498WFL008077810	27	33	3,25	74,5	45
56	327WFL08107905	27	41	83,25	2,5	13
57	310WFL11027706	18	22	156,25	490,5	433
58	0465WFL00117406	18	14	380,25	866,5	769
59	0478WFL00807404	27	33	3,25	74,5	45
60	162WFL09117412	27	25	51,25	274,5	205
61	325WFL05078103	27	39	51,25	8,5	9
62	178WFL06068003	27	39	51,25	8,5	9
63	309WFL11028105	18	44	200,25	116,5	169
64	324WFL05098208	27	33	3,25	74,5	45
65	181WFL06048507	33	36	72,25	50,5	18
66	323WFL08058410	33	35	65,25	62,5	25
67	160WFL07048611	27	31	3,25	112,5	73
68	308WFL11028912	30	38	56,25	14,5	1
69	176WFL06088112	27	33	3,25	74,5	45
70	319WFL02097901	27	33	3,25	74,5	45

Gambar 4 2 Perhitungan Jarak Centroid Awal

4. Mengelompokkan Data Kedalam Cluster

Setelah melakukan proses perhitungan jarak data ke setiap cluster, selanjutnya dilakukan proses mengelompokkan data ke setiap cluster yang terbentuk. Pengelompokkan data ini bertujuan untuk menentukan nilai pusat cluster baru pada proses selanjutnya. Berdasarkan perhitungan jarak sebelumnya, didapati cluster data pertama sebagai berikut :

No	NIK	INDEX		Kluster 0	Kluster 1	Kluster 2
		Performance	Potensial			
1	008AMF18008708	33	39			OK
2	307AMF10037007	24	22	OK		
3	002AMF13077808	27	39		OK	
4	041GMP0004/403	24	28	OK		
5	307AMF00087102	27	33	OK		
6	002AMF14008307	27	33			
7	304AMF11028703	24	30	OK		
8	0483AMF03027303	21	25	OK		
9	313AMF01028104	18	25	OK		
10	180AMF00007808	18	14	OK		
11	303AMF08027801	24	25	OK		
12	0431AMF87030008	39	47		OK	
13	303AMF11027010	18	22	OK		
14	303AMF01117302	27	27	OK		
15	177AMF00007308	21	30	OK		
16	317AMF00088010	21	52		OK	
17	301AMF11028304	27	33	OK		
18	178AMF00008108	18	22	OK		
19	0483AMF01048303	18	22	OK		
20	313AMF11028307	18	22	OK		
21	318AMF03078000	33	44		OK	
22	304AMF10038008	30	33	OK		
23	303AMF03078407	24	31	OK		
24	303AMF11028004	27	33	OK		
25	178AMF07014803	27	30	OK		
26	303AMF11028702	18	25	OK		
27	308AMF08108012	21	39		OK	
28	300AMF11028708	18	25	OK		
29	310AMF00118301	33	39			OK
30	0483AMF01087803	27	33	OK		
31	0471AMF00007603	27	33	OK		
32	0471AMF00007603	18	22	OK		
33	0480AMF03027607	36	44			OK
34	108AMF03038008	33	44		OK	
35	0218AMF01048100	33	41			OK
36	300AMF03038004	21	22	OK		
37	300AMF03078010	27	33	OK		
38	0483AMF00087403	30	39			OK
39	303AMF03078008	30	36			OK
40	0473AMF00087103	27	33	OK		
41	0488AMF03027312	27	30	OK		
42	0470AMF00087702	33	44		OK	
43	177AMF89017712	30	41		OK	
44	300AMF03038312	18	25	OK		
45	0433AMF89117312	18	25	OK		
46	173AMF00138307	27	33	OK		
47	0077AMF00088412	36	44			OK
48	318AMF11028710	36	44			OK
49	331AMF03078703	18	22	OK		
50	0473AMF00007011	27	33	OK		
51	300AMF08107710	27	38			OK
52	0437AMF89108312	18	22	OK		
53	0403AMF00088312	30	36			OK
54	0404AMF87030811	27	36			OK
55	0488AMF00007810	27	33	OK		
56	307AMF08107803	27	41		OK	
57	310AMF11027700	18	22	OK		
58	0403AMF89017400	18	14	OK		
59	0478AMF00087404	27	33	OK		
60	102AMF89117412	27	25	OK		
61	303AMF03078303	27	39	OK		
62	178AMF00008003	27	39	OK		
63	308AMF11028304	18	44		OK	
64	304AMF03038008	27	33	OK		
65	181AMF00048307	33	36			OK

Gambar 4.3 Hasil Cluster Data Awal

5. Menentukan Nilai Titik Pusat Cluster Baru (Centroid)

Setelah melakukan proses perhitungan pertama dengan titik pusat cluster (centroids) awal, maka diperoleh titik pusat (centroids) baru dan dilakukan pengelompokkan data iterasi ke-1.

Inisialisasi titik pusat cluster baru dengan, Centroids 0 merupakan nilai rata-rata dari item data karyawan ke 6 dan 17, Centroids 1 merupakan nilai rata-rata dari item data karyawan ke 2 dan 11, dan dari item data karyawan ke 29 dan 109. Berikut centroid awal pada setiap item data :

Titik Pusat	Index Performance	Index Potensial
Centroids 0	27	33
Centroids 1	25.5	29
Centroids 2	34.5	43

6. Proses Iterasi ke-1 Menghitung Jarak Setiap Data Points ke Centroids Baru

Setelah titik pusat cluster yang baru didapatkan, proses selanjutnya adalah menghitung jarak setiap cluster atau distance. Berikut jarak setiap data ke titik pusat cluster terbaru :

Perhitungan Jarak Centroids Iterasi ke-1

$$\begin{aligned}
 S1|1 \ 16 &= ((33 - 27)^2 + (39 - 33)^2) = 72 \\
 S1|2 \ 8 &= ((33 - 25.5)^2 + (39 - 29)^2) = 156.25 \\
 S1|17 \ 22 &= ((33 - 34.5)^2 + (39 - 45)^2) = 18.25 \\
 S2|1 \ 16 &= ((24 - 27)^2 + (22 - 33)^2) = 130 \\
 S2|2 \ 8 &= ((24 - 25.5)^2 + (22 - 29)^2) = 51.25 \\
 S2|17 \ 22 &= ((24 - 34.5)^2 + (22 - 45)^2) = 551.25 \\
 S3|1 \ 16 &= ((27 - 27)^2 + (39 - 33)^2) = 36 \\
 S3|2 \ 8 &= ((27 - 25.5)^2 + (39 - 29)^2) = 102.25 \\
 S3|17 \ 22 &= ((27 - 34.5)^2 + (39 - 45)^2) = 72.25 \\
 S4|1 \ 16 &= ((24 - 27)^2 + (28 - 33)^2) = 34 \\
 S4|2 \ 8 &= ((24 - 25.5)^2 + (28 - 29)^2) = 3.25 \\
 S4|17 \ 22 &= ((24 - 34.5)^2 + (28 - 45)^2) = 335.25 \\
 S5|1 \ 16 &= ((27 - 27)^2 + (33 - 33)^2) = 0 \\
 S5|2 \ 8 &= ((27 - 25.5)^2 + (33 - 29)^2) = 18.25 \\
 S5|17 \ 22 &= ((27 - 34.5)^2 + (33 - 45)^2) = 156.25
 \end{aligned}$$

Dst

No	NIK	INDEX		Karyawan 1 12	Karyawan 2 11	Karyawan 30 31
		Performance	Potensial			
1	009ANF10008708	33	39	72	150,25	18,25
2	307ANF10037007	24	22	130	51,25	351,25
3	002ANF10077908	27	39	30	102,25	72,25
4	04/CMF/0004/403	24	28	34	3,25	335,25
5	307ANF00087102	27	33	0	18,25	150,25
6	002ANF14009207	27	33	0	18,25	150,25
7	304ANF11028703	24	30	18	3,25	279,25
8	0493ANF00007303	21	25	100	30,25	306,25
9	313ANF01028104	18	25	143	72,25	390,25
10	180ANF000007908	18	14	442	281,25	1113,25
11	309ANF00007901	24	25	73	18,25	434,25
12	0431ANF970300008	39	47	340	306,25	30,25
13	302ANF110287010	18	22	202	103,25	713,25
14	302ANF01117902	27	27	30	6,25	312,25
15	177ANF00007308	21	30	43	21,25	351,25
16	317ANF00008010	21	52	397	349,25	203,25
17	301ANF11028004	27	33	0	18,25	150,25
18	179ANF00008109	18	22	202	103,25	713,25
19	0483ANF01048303	18	22	202	103,25	713,25
20	313ANF11028007	18	22	202	103,25	713,25
21	319ANF00078006	33	44	157	281,25	3,25
22	304ANF10008008	30	33	9	30,25	120,25
23	302ANF00078407	24	31	13	0,25	254,25
24	303ANF11029004	27	33	0	18,25	150,25
25	179ANF07010903	27	30	9	3,25	223,25
26	303ANF11028702	18	25	143	72,25	390,25
27	308ANF06108012	21	39	72	120,25	198,25
28	300ANF11028709	18	25	143	72,25	390,25
29	310ANF00118001	33	39	72	150,25	18,25
30	0492ANF01097803	27	33	0	18,25	150,25
31	0471ANF00007003	27	33	0	18,25	150,25
32	06/FAHF/0119/183	18	22	202	103,25	713,25
33	0490ANF00007007	36	44	202	335,25	3,25
34	109ANF00008009	33	44	157	281,25	3,25
35	0019ANF01048100	33	41	100	200,25	0,25
36	300ANF00008003	21	22	157	09,25	023,25
37	300ANF00078010	27	33	0	18,25	150,25
38	0482ANF00007403	30	39	43	120,25	30,25
39	303ANF00078006	30	36	18	09,25	09,25
40	0472ANF00087103	27	33	0	18,25	150,25
41	0409ANF00077212	27	30	9	3,25	223,25
42	0470ANF00087102	33	44	157	281,25	3,25
43	107ANF99017712	30	41	73	104,25	24,25
44	300ANF00008012	18	25	143	72,25	390,25
45	0437ANF990117312	18	25	143	72,25	390,25
46	173ANF00128007	27	33	0	18,25	150,25
47	0077ANF000080412	36	44	202	335,25	3,25
48	319ANF11028710	36	44	202	335,25	3,25
49	331ANF00078703	18	22	202	103,25	713,25
50	0470ANF00007011	27	33	0	18,25	150,25
51	300ANF06107710	27	38	23	83,25	81,25
52	0437ANF990100312	18	22	202	103,25	713,25
53	0422ANF00008012	30	36	18	09,25	09,25
54	0404ANF970300811	27	36	9	31,25	103,25
55	0489ANF00077810	27	33	0	18,25	150,25
56	307ANF00107903	27	41	04	140,25	00,25
57	310ANF11027706	18	22	202	103,25	713,25
58	0402ANF99017400	18	14	442	281,25	1113,25
59	0472ANF00007404	27	33	0	18,25	150,25
60	102ANF990117412	27	25	04	18,25	380,25
61	303ANF00078303	27	39	30	102,25	72,25
62	178ANF00008003	27	39	30	102,25	72,25
63	309ANF11028305	18	44	202	281,25	273,25
64	304ANF00008008	27	33	0	18,25	150,25
65	181ANF00048307	33	36	43	103,25	51,25
66	303ANF00078410	33	35	40	92,25	00,25
67	100ANF07048011	27	31	4	0,25	200,25
68	308ANF11028912	30	38	34	101,25	43,25
69		27	33			

Gambar 4 4 Perhitungan Jarak Centroid Iterasi ke-1

7. Proses Iterasi ke-1 Mengelompokkan Data Kedalam Cluster

Berdasarkan perhitungan jarak sebelumnya, didapati cluster data kedua sebagai berikut :

No	NIK	INDEX		Kluster0	Kluster1	Kluster2
		Performance	Intensitas			
1	009ANF18008708	33	39		OK	
2	027ANF10037807	24	22	OK		
3	000ANF13077908	27	39		OK	
4	047ANF/0004/403	24	28	OK		
5	007ANF00087102	27	33	OK		
6	000ANF14009007	27	33		OK	
7	004ANF11008703	24	30		OK	
8	0483ANF00007303	21	25		OK	
9	013ANF01008104	18	25		OK	
10	180ANF00007308	18	14		OK	
11	009ANF00007901	24	25			OK
12	0431ANF07000008	39	47		OK	
13	000ANF11007710	18	22		OK	
14	000ANF01117902	27	27		OK	
15	077ANF00007308	21	30			OK
16	017ANF00008010	21	52	OK		
17	001ANF11008004	27	33		OK	
18	178ANF00008108	18	22		OK	
19	0483ANF01048303	18	22		OK	
20	013ANF11008007	18	22			OK
21	019ANF00078000	33	44	OK		
22	004ANF10008008	30	33		OK	
23	000ANF00078407	24	31	OK		
24	000ANF11008004	27	33		OK	
25	109ANF07016903	27	30		OK	
26	000ANF11008702	18	25	OK		
27	000ANF08108012	21	39		OK	
28	000ANF11008708	18	25			OK
29	010ANF00118001	33	39	OK		
30	0483ANF01087803	27	33	OK		
31	0471ANF00007603	27	33		OK	
32	047ANF/0119/183	18	22			OK
33	0490ANF03007807	36	44			OK
34	109ANF03008008	33	44			OK
35	0019ANF01048100	33	41		OK	
36	000ANF03008003	21	22	OK		
37	000ANF00078010	27	33			OK
38	0483ANF00007403	30	39	OK		
39	000ANF00078008	30	36	OK		
40	0473ANF00087103	27	33		OK	
41	0483ANF00077212	27	30			OK
42	0470ANF00087702	33	44			OK
43	107ANF00017712	30	41		OK	
44	000ANF00008012	18	25		OK	
45	0431ANF00117312	18	25	OK		
46	173ANF00108007	27	33			OK
47	0077ANF000080412	36	44			OK
48	019ANF11008710	36	44		OK	
49	001ANF00078703	18	22	OK		
50	0473ANF00007811	27	33	OK		
51	000ANF08107710	27	38		OK	
52	0437ANF00108012	18	22	OK		
53	0483ANF00008012	30	36	OK		
54	0484ANF07008011	27	36	OK		
55	0483ANF00077810	27	33			OK
56	007ANF08107900	27	41		OK	
57	010ANF11007708	18	22		OK	
58	0483ANF00017400	18	14	OK		
59	0478ANF00007404	27	33		OK	
60	109ANF00117412	27	25	OK		
61	000ANF00078303	27	39	OK		
62	178ANF00008003	27	39	OK		
63	000ANF11008300	18	44	OK		
64	004ANF00008008	27	33	OK		
65	001ANF00048007	33	36	OK		
66	000ANF000080410	33	35	OK		
67	100ANF07048011	27	31	OK		
68	000ANF11008012	30	38	OK		
69	173ANF00008110	27	33	OK		

Gambar 4 5 Hasil Cluster Data Iterasi ke-1

8. Menentukan Nilai Titik Pusat Cluster Baru (Centroid)

Setelah melakukan proses perhitungan pertama dengan titik pusat cluster (centroids) ke-2, maka diperoleh titik pusat (centroids) baru dan dilakukan pengelompokkan data iterasi ke-2.

Inisialisasi titik pusat cluster baru dengan, Centroids 0 merupakan nilai rata-rata dari item data karyawan ke 1 dan 96, Centroids 1 merupakan nilai rata-rata dari item data karyawan ke 4 dan 13, dan dari item data karyawan ke 1 dan 33. Berikut centroid awal pada setiap item data :

Titik Pusat	Index Performance	Index Potensial
Centroids 0	25.5	33
Centroids 1	21	25
Centroids 2	34.5	41.5

9. Proses Iterasi ke-2 Menghitung Jarak Setiap Data Points ke Centroids Baru

Setelah titik pusat cluster yang baru didapatkan, proses selanjutnya adalah menghitung jarak setiap cluster atau distance. Berikut jarak setiap data ke titik pusat cluster terbaru :

Perhitungan Jarak Centroids Iterasi ke-2

$$\begin{aligned} S1|1 \ 96 &= ((33 - 25.5)^2 + (39 - 33)^2) = 92.25 \\ S1|4 \ 13 &= ((33 - 21)^2 + (39 - 25)^2) = 340 \\ S1|1 \ 33 &= ((33 - 34.5)^2 + (39 - 41.5)^2) = 74.5 \\ S2|1 \ 96 &= ((24 - 25.5)^2 + (22 - 33)^2) = 123.25 \\ S2|4 \ 13 &= ((24 - 21)^2 + (22 - 25)^2) = 18 \\ S2|1 \ 33 &= ((24 - 34.5)^2 + (22 - 41.5)^2) = 416.5 \\ S3|1 \ 96 &= ((27 - 25.5)^2 + (39 - 33)^2) = 38.25 \\ S3|4 \ 13 &= ((27 - 21)^2 + (39 - 25)^2) = 232 \\ S3|1 \ 33 &= ((27 - 34.5)^2 + (39 - 41.5)^2) = 266.5 \\ S4|1 \ 96 &= ((24 - 25.5)^2 + (28 - 33)^2) = 27.25 \\ S4|4 \ 13 &= ((24 - 21)^2 + (28 - 25)^2) = 18 \\ S4|1 \ 33 &= ((24 - 34.5)^2 + (28 - 41.5)^2) = 416.5 \\ S5|1 \ 96 &= ((27 - 25.5)^2 + (33 - 33)^2) = 2.25 \\ S5|4 \ 13 &= ((27 - 21)^2 + (33 - 25)^2) = 100 \\ S5|1 \ 33 &= ((27 - 34.5)^2 + (33 - 41.5)^2) = 266.5 \end{aligned}$$

Dst

No	NIK	INDEX		Karyawan 1 96	Karyawan 4 13	Karyawan 1 33
		Indeks	Estimasi			
1	0001VP1000708	33	39	92,25	340	74,5
2	0271VP10007607	24	22	123,25	18	416,5
3	0021VP10007908	27	39	38,25	232	266,5
4	0111VP10004100	24	28	27,25	18	416,5
5	0071VP00087102	27	33	2,25	100	266,5
6	0021VP10009207	27	33	2,25	100	266,5
7	0241VP10008700	24	30	11,25	34	416,5
8	0493VP00057303	21	25	84,25	0	602,5
9	0111VP00028104	18	25	120,25	9	824,5
10	0001VP00067908	18	14	417,25	130	824,5
11	0291VP00057901	24	25	66,25	9	416,5
12	0431VP00066608	39	47	378,25	808	26,5
13	0221VP10027610	18	22	177,25	18	824,5
14	0051VP01117902	27	27	38,25	40	266,5
15	0771VP00067308	21	30	29,25	25	602,5
16	0171VP00099010	21	52	381,25	729	602,5
17	0211VP10008704	27	33	2,25	100	266,5
18	0791VP00068109	18	22	177,25	18	824,5
19	0481VP01048309	18	22	177,25	18	824,5
20	0131VP10028207	18	22	177,25	18	824,5
21	0191VP00079006	33	44	177,25	505	74,5
22	0241VP10008608	30	33	20,25	145	152,5
23	0221VP00078407	24	31	6,25	45	416,5
24	0231VP10029004	27	33	2,25	100	266,5
25	0591VP07016903	27	30	11,25	61	266,5
26	0251VP10026702	18	25	120,25	9	824,5
27	0281VP00108012	21	39	56,25	196	602,5
28	0201VP10028709	18	25	120,25	9	824,5
29	0101VP00118201	33	39	92,25	340	74,5
30	0495VP01097905	27	33	2,25	100	266,5
31	0471VP00057605	27	33	2,25	100	266,5
32	0111VP00101018	18	22	177,25	18	824,5
33	0490VP00057607	36	44	231,25	586	32,5
34	0691VP00058009	33	44	177,25	505	74,5
35	0519VP01048106	33	41	120,25	400	74,5
36	0201VP00058205	21	22	141,25	9	602,5
37	0281VP00078010	27	33	2,25	100	266,5
38	0485VP00097403	30	39	56,25	277	152,5
39	0231VP00078008	30	36	29,25	202	152,5
40	0473VP00087103	27	33	2,25	100	266,5
41	0468VP00077212	27	30	11,25	61	266,5
42	0478VP00087702	33	44	177,25	505	74,5
43	0571VP00017712	30	41	84,25	337	152,5
44	0201VP00088512	18	25	120,25	9	824,5
45	0455VP06117312	18	25	120,25	9	824,5
46	0751VP06128207	27	33	2,25	100	266,5
47	0677VP00088612	36	44	231,25	586	32,5
48	0191VP10028710	36	44	231,25	586	32,5
49	0311VP00078703	18	22	177,25	18	824,5
50	0475VP00057611	27	33	2,25	100	266,5
51	0261VP00107710	27	38	27,25	205	266,5
52	0407VP06106012	18	22	177,25	18	824,5
53	0422VP00086912	30	36	29,25	202	152,5
54	0424VP07006811	27	36	11,25	157	266,5
55	0488VP00077910	27	33	2,25	100	266,5
56	0271VP00107905	27	41	66,25	292	266,5
57	0101VP10027706	18	22	177,25	18	824,5
58	0465VP00017406	18	14	417,25	130	824,5
59	0479VP00097404	27	33	2,25	100	266,5
60	0621VP00117412	27	25	66,25	36	266,5
61	0251VP00078303	27	39	38,25	232	266,5
62	0791VP00068003	27	39	38,25	232	266,5
63	0091VP10028305	18	44	177,25	370	824,5
64	0241VP00088208	27	33	2,25	100	266,5
65	0811VP00048507	33	36	65,25	265	74,5
66	0231VP00058410	33	35	60,25	244	74,5
67	0691VP07048611	27	31	6,25	72	266,5

Gambar 4 6 Perhitungan Jarak Centroid Iterasi ke-2

10. Proses Iterasi ke-2 Mengelompokkan Data Kedalam Cluster

Berdasarkan perhitungan jarak sebelumnya, didapati cluster data ketiga sebagai berikut :

No	NIK	INDEX		Kluster0	Kluster1	Kluster2
		Euclidean	Euclidean1			
1	000WF1000700	33	39			OK
2	027WF1000700	24	22		OK	
3	002WF1007900	27	39	OK		
4	017WF10004100	24	28		OK	
5	007WF00007100	27	33	OK		
6	002WF1000007	27	33	OK		
7	024WF1000700	24	30	OK		
8	000WF00007000	21	25		OK	
9	013WF01000100	18	25		OK	
10	000WF00007000	18	14		OK	
11	020WF00007000	24	25		OK	
12	000WF00000000	39	47			OK
13	022WF1007010	18	22		OK	
14	000WF0111700	27	27	OK		
15	077WF00007000	21	30		OK	
16	007WF0000000	21	52	OK		
17	021WF1000000	27	33	OK		
18	070WF0000000	18	22		OK	
19	000WF00000000	18	22		OK	
20	013WF1000007	18	22		OK	
21	000WF00007000	33	44			OK
22	024WF10000000	30	33	OK		
23	022WF00007000	24	31	OK		
24	023WF1000000	27	33	OK		
25	000WF00000000	27	30	OK		
26	025WF1000000	18	25		OK	
27	000WF00000000	21	39	OK		
28	020WF1000000	18	25		OK	
29	000WF0010000	33	39			OK
30	000WF00000000	27	33	OK		
31	007WF00007000	27	33	OK		
32	017WF01000100	18	22		OK	
33	000WF00007000	36	44			OK
34	000WF00000000	33	44			OK
35	000WF00000000	33	41			OK
36	000WF00000000	21	22		OK	
37	026WF00007000	27	33	OK		
38	000WF00007000	30	39	OK		
39	023WF00007000	30	36	OK		
40	007WF00007000	27	33	OK		
41	000WF00007000	27	30	OK		
42	007WF00007000	33	44			OK
43	007WF00007000	30	41	OK		
44	000WF00000000	18	25		OK	
45	000WF00007000	18	25		OK	
46	070WF00000000	27	33	OK		
47	007WF00000000	36	44			OK
48	010WF1000000	36	44			OK
49	011WF00007000	18	22		OK	
50	007WF00007000	27	33	OK		
51	000WF00007000	27	38	OK		
52	007WF00007000	18	22		OK	
53	002WF00000000	30	36	OK		
54	002WF00000000	27	36	OK		
55	000WF00007000	27	33	OK		
56	007WF00007000	27	41	OK		
57	000WF1000700	18	22		OK	
58	000WF00007000	18	14		OK	
59	007WF00007000	27	33	OK		
60	000WF00007000	27	25		OK	
61	025WF00007000	27	39	OK		
62	070WF00000000	27	39	OK		
63	000WF1000000	18	44	OK		
64	000WF00000000	27	33	OK		
65	010WF00000000	33	36	OK		
66	000WF00000000	33	35	OK		
67	000WF00000000	27	31	OK		

Gambar 4 7 Hasil Cluster Data Iterasi ke-2

11. Menentukan Nilai Titik Pusat Cluster Baru (Centroid)

Setelah melakukan proses perhitungan pertama dengan titik pusat cluster (centroids) ke-3, maka diperoleh titik pusat (centroids) baru dan dilakukan pengelompokan data iterasi ke-3.

Inisialisasi titik pusat cluster baru dengan, Centroids 0 merupakan nilai rata-rata dari item data karyawan ke 29 dan 96, Centroids 1 merupakan nilai rata-rata dari item data karyawan ke 7 dan 13, dan dari item data karyawan ke 1 dan 34. Berikut centroid awal pada setiap item data :

Titik Pusat	Index Performance	Index Potensial
Centroids 0	25.5	33
Centroids 1	21	26
Centroids 2	33	41.5

12. Proses Iterasi ke-3 Menghitung Jarak Setiap Data Points ke Centroids Baru

Setelah titik pusat cluster yang baru didapatkan, proses selanjutnya adalah menghitung jarak setiap cluster atau distance. Berikut jarak setiap data ke titik pusat cluster terbaru :

Perhitungan Jarak Centroids Iterasi ke-3

$$\begin{aligned} S1|29 \ 96 &= ((33 - 25.5)^2 + (39 - 33)^2) = 92.25 \\ S1|7 \ 13 &= ((33 - 21)^2 + (39 - 26)^2) = 313 \\ S1|1 \ 34 &= ((33 - 33)^2 + (39 - 41.5)^2) = 72.25 \\ S2|29 \ 96 &= ((24 - 25.5)^2 + (22 - 33)^2) = 123.25 \\ S2|7 \ 13 &= ((24 - 21)^2 + (22 - 26)^2) = 25 \\ S2|1 \ 34 &= ((24 - 33)^2 + (22 - 41.5)^2) = 387.25 \\ S3|29 \ 96 &= ((27 - 25.5)^2 + (39 - 33)^2) = 38.25 \\ S3|7 \ 13 &= ((27 - 21)^2 + (39 - 26)^2) = 205 \\ S3|1 \ 34 &= ((27 - 33)^2 + (39 - 41.5)^2) = 246.25 \\ S4|29 \ 96 &= ((24 - 25.5)^2 + (28 - 33)^2) = 27.25 \\ S4|7 \ 13 &= ((24 - 21)^2 + (28 - 26)^2) = 13 \\ S4|1 \ 34 &= ((24 - 33)^2 + (28 - 41.5)^2) = 387.25 \\ S5|29 \ 96 &= ((27 - 25.5)^2 + (33 - 33)^2) = 2.25 \\ S5|7 \ 13 &= ((27 - 21)^2 + (33 - 26)^2) = 85 \\ S5|1 \ 34 &= ((27 - 33)^2 + (33 - 41.5)^2) = 246.25 \end{aligned}$$

No	NIK	INDEX		Karyawan 1 12	Karyawan 2 11	Karyawan 30 31
		Performance	Potensial			
1	009ANF18008708	33	39	92,25	313	72,25
2	327ANF10037907	24	22	123,25	25	387,25
3	009ANF13077908	27	39	38,25	203	240,25
4	04/GNF/0004/403	24	28	27,25	13	387,25
5	307ANF09087102	27	33	2,25	85	240,25
6	009ANF14009207	27	33	2,25	85	240,25
7	324ANF11028703	24	30	11,25	25	387,25
8	0493ANF03007303	21	25	84,25	1	504,25
9	313ANF01028104	18	25	120,25	10	777,25
10	180ANF00007908	18	14	417,25	153	777,25
11	329ANF08007901	24	25	66,25	10	387,25
12	0431ANF97030009	39	47	378,25	763	48,25
13	329ANF11028710	18	22	177,25	25	777,25
14	309ANF01117902	27	27	38,25	37	240,25
15	177ANF00007308	21	30	29,25	16	504,25
16	317ANF02099010	21	52	381,25	676	504,25
17	323ANF11028204	27	33	2,25	85	240,25
18	179ANF00008109	18	22	177,25	25	777,25
19	0493ANF01048303	18	22	177,25	25	777,25
20	313ANF11028207	18	22	177,25	25	777,25
21	319ANF03078000	33	44	177,25	468	72,25
22	324ANF11028208	30	33	20,25	130	141,25
23	329ANF03078407	24	31	6,25	34	387,25
24	323ANF11028904	27	33	2,25	85	240,25
25	129ANF07010903	27	30	11,25	52	240,25
26	329ANF11028702	18	25	120,25	10	777,25
27	329ANF08108012	21	39	56,25	169	504,25
28	320ANF11028709	18	25	120,25	10	777,25
29	310ANF00118201	33	39	92,25	313	72,25
30	0493ANF01097803	27	33	2,25	85	240,25
31	0471ANF00007003	27	33	2,25	85	240,25
32	AG/TANF/0119/183	18	22	177,25	25	777,25
33	0490ANF03007007	36	44	231,25	549	39,25
34	109ANF03008009	33	44	177,25	468	72,25
35	0519ANF01048100	33	41	120,25	369	72,25
36	320ANF03008203	21	22	141,25	16	504,25
37	320ANF03078010	27	33	2,25	85	240,25
38	0483ANF00097403	30	39	56,25	250	141,25
39	323ANF03078008	30	36	29,25	181	141,25
40	0473ANF00087103	27	33	2,25	85	240,25
41	0408ANF02077212	27	30	11,25	52	240,25
42	0478ANF00087302	33	44	177,25	468	72,25
43	137ANF99017712	30	41	84,25	306	141,25
44	320ANF02098512	18	25	120,25	10	777,25
45	0405ANF99117312	18	25	120,25	10	777,25
46	179ANF00128207	27	33	2,25	85	240,25
47	0077ANF00008412	36	44	231,25	549	39,25
48	319ANF11028710	36	44	231,25	549	39,25
49	331ANF03078703	18	22	177,25	25	777,25
50	0473ANF00007011	27	33	2,25	85	240,25
51	320ANF08107710	27	38	27,25	180	240,25
52	0437ANF99108312	18	22	177,25	25	777,25
53	0493ANF00088912	30	36	29,25	181	141,25
54	0494ANF97030011	27	36	11,25	136	240,25
55	0488ANF02077810	27	33	2,25	85	240,25
56	327ANF08107903	27	41	66,25	261	240,25
57	310ANF11028709	18	22	177,25	25	777,25
58	0403ANF99017400	18	14	417,25	153	777,25
59	0478ANF00097404	27	33	2,25	85	240,25
60	109ANF99117412	27	25	66,25	37	240,25
61	329ANF03078303	27	39	38,25	203	240,25
62	179ANF00008003	27	39	38,25	203	240,25
63	309ANF11028303	18	44	177,25	333	777,25
64	324ANF03098208	27	33	2,25	85	240,25
65	181ANF00048207	33	36	65,25	244	72,25
66	323ANF08038410	33	35	60,25	223	72,25

Gambar 4 8 Perhitungan Jarak Centroid Iterasi ke-3

13. Proses Iterasi ke-3 Mengelompokkan Data Kedalam Cluster

Berdasarkan perhitungan jarak sebelumnya, didapati cluster data keempat sebagai berikut :

No	NIK	INDEX		Kluster0	Kluster1	Kluster2
		Euclidean	Centroid			
1	000WF1006700	33	39			OK
2	027WF1007607	24	22		OK	
3	000WF1007700	27	39	OK		
4	000WF1007700	24	28		OK	
5	007WF00067100	27	33	OK		
6	002WF10069207	27	33	OK		
7	024WF11028700	24	30	OK		
8	049WF00057300	21	25		OK	
9	011WF01008100	18	25		OK	
10	080WF00067900	18	14		OK	
11	020WF00057900	24	25		OK	
12	041WF07006600	39	47			OK
13	022WF11027610	18	22		OK	
14	000WF01117902	27	27	OK		
15	077WF00067300	21	30		OK	
16	017WF02008010	21	52	OK		
17	021WF11028204	27	33	OK		
18	070WF00081000	18	22		OK	
19	048WF01008300	18	22		OK	
20	010WF11028207	18	22		OK	
21	010WF00078006	33	44			OK
22	020WF10008600	30	33	OK		
23	022WF00078407	24	31	OK		
24	020WF11028004	27	33	OK		
25	050WF07016900	27	30	OK		
26	020WF11028700	18	25		OK	
27	020WF08108012	21	39	OK		
28	020WF11028700	18	25		OK	
29	010WF00118201	33	39			OK
30	040WF01007800	27	33	OK		
31	047WF00057605	27	33	OK		
32	000WF01101100	18	22		OK	
33	040WF00057607	36	44			OK
34	080WF00088000	33	44			OK
35	051WF01008106	33	41			OK
36	020WF00088000	21	22		OK	
37	020WF00078010	27	33	OK		
38	040WF00007400	30	39	OK		
39	020WF00078008	30	36	OK		
40	017WF00087100	27	33	OK		
41	040WF020077012	27	30	OK		
42	017WF00087702	33	44			OK
43	057WF00017712	30	41	OK		
44	020WF00088112	18	25		OK	
45	040WF000177012	18	25		OK	
46	070WF00012807	27	33	OK		
47	067WF00088412	36	44			OK
48	010WF11028710	36	44			OK
49	011WF00078700	18	22		OK	
50	047WF00057611	27	33	OK		
51	020WF000107710	27	38	OK		
52	047WF000106012	18	22		OK	
53	042WF00088012	30	36	OK		
54	042WF07006811	27	36	OK		
55	040WF000107710	27	33	OK		
56	027WF000107905	27	41	OK		
57	010WF11027700	18	22		OK	
58	040WF00017406	18	14		OK	
59	047WF00088404	27	33	OK		
60	062WF00117412	27	25		OK	
61	020WF00078001	27	39	OK		
62	070WF00088000	27	39	OK		
63	000WF11028005	18	44	OK		
64	020WF00088208	27	33	OK		
65	011WF000108007	33	36	OK		
66	020WF00088410	33	35	OK		
67	060WF07008611	27	31	OK		

Gambar 4 9Hasil Cluster Data Iterasi ke-3

14. Hasil Akhir Perhitungan

Hasil dari hasil akhir perhitungan iterasi pengelompokkan data penilaian kinerja karyawan di PT.XYZ menggunakan sampel data sebanyak 61 records (setengah dari keseluruhan records data yang dimiliki), maka diperoleh hasil bahwa :

a) Cluster 0

Pada cluster 0 memiliki 29 data yang yang diinisialisasikan sebagai kelompok pertama. Kelompok pertama ini diartikan sebagai kelompok dengan hasil penilaian kinerja karyawan 'baik'.

No	NIK	Nama Tenaga Kerja	Jabatan	Index Performance	Index Potensial
1	002AMP13077908	Hirwansyah	Sortase Helper	27	39
2	307AMP09087102	Saudi	Security	27	33
3	002AMP14069207	Syafrino	Sortase Helper	27	33
4	324AMP11028703	Ad Edianto	Process Operator	24	30
5	305AMP01117902	Amra Luis	Security	27	27
6	317AMP02098010	Andi Sumarno	Process Operator	21	52
7	321AMP11028204	Ardiones	Capstan Operator	27	33
8	324AMP10098608	Bal Efendi	Sample Boy	30	33
9	322AMP05078407	Bulkaini	Sortase Foreman	24	31
10	323AMP11029004	Candra Kurniawan	Hoisting Crane Operator	27	33
11	159AMP07016903	Darmawi	Driver	27	30
12	328AMP08108012	Daswir	Boiler Operator	21	39
13	0495AMP01097805	Deris Suanto	KCP Operator	27	33
14	0471AMP00057605	Deswito	Pressing Operator	27	33

15	326AMP05078010	Ermansyah	Process Helper	27	33
16	0485AMP00097403	Erwin	Land Application Operator	30	39
17	323AMP05078008	Gusmaldi	Process Helper	30	36
18	0473AMP00087103	Gusti Metri	Transfer Carriage Operator	27	33
19	0468AMP02077212	Gustiar	Clarification Operator	27	30
20	157AMP99017712	Hendra	Pressing Operator	30	41
21	175AMP06128207	Hendri Yonanda	Sample Boy	27	33
22	0475AMP00057611	Irmeh Suhita	KCP Operator	27	33
23	326AMP08107710	Jumanto	Maintenance Helper	27	38
24	0422AMP00086912	Khairul Amri	Effluent Operator	30	36
25	0424AMP97036811	M. Nur Alamsyah	Sortase Operator	27	36
26	0488AMP02077810	Madrizel Miyanto	Treshing Operator	27	33
27	327AMP08107905	Maihendri	Engine Room Operator	27	41
28	0478AMP00097404	Masriadi	Boiler Operator	27	33
29	325AMP05078303	Masrizel	Process Helper	27	39

Tabel 4.3 Hasil Pengelompokan Data Cluster 0

Berdasarkan hasil cluster 0 dapat diketahui terdapat 29 data karyawan. Tabel diatas berisikan informasi karyawan mulai dari NIK, Nama, Jabatan, serta index penilaian kinerja karyawan. Jumlah data pada cluster 0 merupakan jumlah data yang paling banyak dibanding cluster lain. Berdasarkan tabel diatas, dapat diketahui rentang nilai index performance yang dimiliki mulai dari 21-30 dan index potensial mulai dari 27- 52 dengan kombinasi nilai dari index performance dan index potensial yang menjadi faktor penentu dalam pengelompokan.

b) Cluster 1

Pada cluster 1 memiliki 22 data yang diinisialisasikan sebagai kelompok kedua. Kelompok kedua ini diartikan sebagai kelompok dengan hasil penilaian kinerja karyawan ‘kurang baik’.

No	NIK	Nama Tenaga Kerja	Jabatan	Index Perfor mance	Index Potensi al
1	327AMP10037607	Dedi Iman Widiyanto	Security	24	22
2	GA/GMP/0604/403	KIWAL	Security	24	28
3	0493AMP03057303	Adrison	L. Ramp Operator	21	25
4	313AMP01028104	Afrizal Arlin	Process Operator	18	25
5	180AMP06067908	Agus Wandu	Security	18	14
6	329AMP08057901	Agusman	Sortase Helper	24	25
7	322AMP11027610	Amir Johan	Sterilizer Operator	18	22
8	177AMP06067308	Amrizal	Boiler Operator	21	30
9	179AMP06068109	Aris	Transfer Carriage Operator	18	22
10	0483AMP01048303	Asril	L. Ramp Operator	18	22
11	313AMP11028207	Asril	L. Ramp Operator	18	22
12	325AMP11026702	Dasril	Mechanical Helper	18	25
13	320AMP11028709	Dedrinaldi	Capstan Operator	18	25
14	AG/PAMP/0119/18 3	Dodi Agusman	Security	18	22
15	320AMP03058205	Emrizal	Process Helper	21	22

16	320AMP02098512	Hendra. A	Process Helper	18	25
17	0455AMP96117312	Hendra.y	Engine Driver	18	25
18	331AMP05078703	Ipen Riawan	Capstan Operator	18	22
19	0437AMP96106312	Kenek	PGA Staff	18	22
20	310AMP11027706	Mardianto	Sterilizer Operator	18	22
21	0465AMP99017406	Mardinas	Clarification Operator	18	14
22	162AMP99117412	Masrizal	Security	27	25

Tabel 4.4 Hasil Pengelompokan Data Cluster 1

Berdasarkan hasil cluster 1 dapat diketahui terdapat 22 data karyawan. Tabel diatas berisikan informasi karyawan mulai dari NIK, Nama, Jabatan, serta index penilaian kinerja karyawan. Jumlah data pada cluster 1 lebih banyak dari pada cluster ke 2 namun lebih sedikit dari cluster 0. Berdasarkan tabel diatas, dapat diketahui rentang nilai index performance yang dimiliki mulai dari 18-27 dan index potensial mulai dari 14- 30 dengan kombinasi nilai dari index performance dan index potensial yang menjadi faktor penentu dalam pengelompokan.

c) Cluster 2

Pada cluster 2 memiliki 10 data yang diinisialisasikan sebagai kelompok ketiga. Kelompok ketiga ini diartikan sebagai kelompok dengan hasil penilaian kinerja karyawan ‘sangat baik’.

No	NIK	Nama Tenaga Kerja	Jabatan	Index Performance	Index Potensial
1	009AMP18068708	Afrizal	Mechanical Helper	33	39
2	0431AMP97036608	Ajang Nugraha	Sample Boy	39	47
3	319AMP05078006	Asril	Maintenance POM	33	44

4	310AMP00118201	Delita Effendi	Process Operator	33	39
5	0490AMP03057607	Dodi Eka Edison	KCP Operator	36	44
6	169AMP03058009	Doni Hendra	KCP Operator	33	44
7	0519AMP01048106	Eka Nofianda	KCP Operator	33	41
8	0476AMP00087702	Haryodis	Sterilizer Operator	33	44
9	0677AMP00086412	Hendriful	Clarification Operator	36	44
10	319AMP11028710	Indra Bangsawan	L. Ramp Operator	36	44

Tabel 4 5 Hasil Pengelompokkan Data Cluster 2

Berdasarkan hasil cluster 2 dapat diketahui terdapat 10 data karyawan. Tabel diatas berisikan informasi karyawan mulai dari NIK, Nama, Jabatan, serta index penilaian kinerja karyawan. Jumlah data pada cluster 2 merupakan jumlah data paling sedikit dalam pengelompokkan ini. Berdasarkan tabel diatas, dapat diketahui rentang nilai index performance yang dimiliki mulai dari 33-39 dan index potensial mulai dari 39- 47 dengan kombinasi nilai dari index performance dan index potensial yang menjadi faktor penentu dalam pengelompokkan.

4.4 Pehitungan Menggunakan Python Algoritma K-Means Clustering

Jupyter Notebook merupakan singkatan dari tiga bahasa pemrograman, yakni Julia (Ju), Python (Py), dan R. Jupyter adalah sebuah aplikasi web gratis yang paling banyak dipakai oleh data scientist. Aplikasi ini dipakai untuk membuat dan membagikan dokumen yang memiliki kode, hasil hitungan, visualisasi, dan teks. Jupyter Notebook berfungsi membantu data scientist dalam membuat narasi komputasi. Narasi komputasi menjelaskan makna dari data di dalamnya dan memberikan insight (wawasan) mengenai data tersebut.

4.4.1 Pengolahan Data

Berikut langkah-langkah implementasi k-means clustering dalam pengolahan data penilaian kinerja karyawan pada PT.XYZ menggunakan bahasa pemograman python :

1. Import library yang dibutuhkan dalam pengolahan data

```
# import library yang dibutuhkan
import pandas as pd
```

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.cluster import KMeans
from sklearn.preprocessing import MinMaxScaler
```

Ada enam library yang dibutuhkan dalam pengolahan data penilaian kinerja karyawan di PT.XYZ. Pertama library Pandas yang digunakan untuk memanipulasi data, seperti proses membaca data dari berbagai format, function dataframe, memproses penggabungan data, memfilter data, menghapus data yang tidak dibutuhkan, ataupun agregasi. Library Numpy digunakan untuk perhitungan aljabar seperti operasi vektor dan matriks. Library Matplotlib dan Seaborn digunakan untuk visualisasi data seperti pembuatan grafik, chart, histogram, dan lain-lain. Dan terakhir, tentu saja membutuhkan library KMeans untuk memproses pengelompokan data menggunakan algoritma k-means clustering.

2. Gathering Data

Gathering data ialah proses menggabungkan sejumlah data pada data collection. Berikut proses dalam mengimport data penilaian kinerja karyawan dengan nama file 'Book1ARA_DataAsli_Olah' dengan format csv ke jupyter notebook. Data yang diimport akan disimpan kedalam variabel inData.

```
# read data
inData = pd.read_csv('Book1ARA_DataAsli_Olah.csv', delimiter= ';')
inData.head()
```

Output :

Hasil dari import data, menampilkan dataset seperti berikut.

NIK	NAMA TENAGA KERJA	TANGGAL LAHIR	UMUR(TH/BLN)	TANGGAL MASUK	MASA KERJA(TH/BLN)	JABATAN	PENDIDIKAN	HK- Kuantitas Kerja	Ketej V
009AMP18068708	Afrizal	18/08/1987	32-2	02/06/2018	01-Apr	Mechanical Helper	NaN	30	
327AMP10037807	Dedi Iman Widiyanto	07/07/1976	43-3	01/03/2010	09-Jul	Security	NaN	30	
002AMP13077908	Hirwansyah	05/08/1979	40-2	01/07/2013	06-Mar	Sortase Helper	NaN	30	
GA/GMP/0604/403	KIWAL	02/05/1973	46-5	01/06/2004	15-Apr	Security	NaN	30	
307AMP09087102	Saudi	04/02/1971	48-8	01/08/2009	10-Feb	Security	NaN	30	

Mengecek dataset, apakah ada item data yang missing value atau tidak

```
# melakukan cleaning data
# mengecek missing value pada data
inData.isnull().any()
```

Output :

Hasil yang didapat ialah seperti gambar di bawah. Dapat dilihat bahwa terdapat 2 item data yang memiliki missing value yaitu 'Pendidikan' dan 'Keterangan'. Karena dari data yang diperoleh kedua item data ini memang dikosongkan dari pihak PT.XYZ, karena merupakan data privasi karyawannya.

```
NIK                False
NAMA TENAGA KERJA  False
TANGGAL LAHIR      False
UMUR(TH/BLN)       False
TANGGAL MASUK      False
MASA KERJA(TH/BLN) False
JABATAN            False
PENDIDIKAN         True
HK-Kuantitas Kerja False
HK-Ketepatan Waktu False
HK-Kualitas Kerja  False
PK-Pengetahuan Tentang Pekerjaan False
PK-Komunikasi       False
PK-Kerjasama Tim   False
PK-Disiplin Kerja  False
INDEX Performannse False
INDEX Potensial    False
KETERANGAN         True
dtype: bool
```

Menghapus kolom yang mengalami missing value, yaitu 'Pendidikan' dan 'Keterangan'.

```
# menghilangkan item data pendidikan dan keterangan, karena pada data yang
didapatkan kedua item ini tidak memiliki value
df = inData.drop(['KETERANGAN', 'PENDIDIKAN'], axis = 1)
df.head()
```

Output :

Hasil dari proses ini ialah hilangnya kolom 'Pendidikan' dan 'Keterangan' pada dataset. Tampilan dataset seperti berikut.

NIK	NAMA TENAGA KERJA	TANGGAL LAHIR	UMUR(TH/BLN)	TANGGAL MASUK	MASA KERJA(TH/BLN)	JABATAN	HK- Kuantitas Kerja	HK- Ketepatan Waktu	HK- Kualitas Kerja	PK- Pengetahuan Tentang Pekerjaan	PK- Komunikasi	Ke
009AMP18068708	Afrizal	18/08/1987	32-2	02/06/2018	01-Apr	Mechanical Helper	30	40	40	30	30	
327AMP10037607	Dedi Iman Widiyanto	07/07/1976	43-3	01/03/2010	09-Jul	Security	30	30	20	20	20	
002AMP13077908	Hirwansyah	05/08/1979	40-2	01/07/2013	06-Mar	Sortase Helper	30	30	30	30	30	
GA/GMP/0604/403	KIWAL	02/05/1973	46-5	01/06/2004	15-Apr	Security	30	30	20	20	20	
307AMP09087102	Saudi	04/02/1971	48-8	01/08/2009	10-Feb	Security	30	30	30	30	30	

3. Data Transformation dan Data Preprocessing Untuk Matrik Korelasi

Data transformation adalah upaya yang bertujuan untuk mengubah skala pengukuran data asli (format data) menjadi bentuk lain sehingga data dapat memenuhi asumsi-asumsi yang mendasari analisis ragam atau dapat dipahami dan dianalisis oleh machine learning. Data transformation pada tahap ini dilakukan untuk memproses sebelum menghasilkan matrik korelasi antar item data penilaian kinerja karyawan di PT.XYZ.

Pertama lakukan pengecekan terhadap tipe data yang dimiliki dataset.

```
# cek tipe data awal
tipe = df.dtypes
print(tipe)
```

Output :

Hasil dari proses ini akan menampilkan tipe data pada setiap item data pada dataset penilaian kinerja karyawan di PT.XYZ seperti di bawah.

```
NIK                                object
NAMA TENAGA KERJA                 object
TANGGAL LAHIR                    object
UMUR(TH/BLN)                     object
TANGGAL MASUK                    object
MASA KERJA(TH/BLN)               object
JABATAN                           object
HK-Kuantitas Kerja                int64
HK-Ketepatan Waktu                int64
HK-Kualitas Kerja                 int64
PK-Pengetahuan Tentang Pekerjaan int64
PK-Komunikasi                     int64
PK-Kerjasama Tim                 int64
PK-Disiplin Kerja                 int64
INDEX Performannse               int64
INDEX Potensial                  int64
dtype: object
```

Dapat dilihat pada gambar diatas masih terdapat item data yang bertipe data object. Dalam menghasilkan matrix korelasi, semua tipe data setiap item set harus

dalam bentuk int atau float (dalam bentuk angka). Sehingga item data yang masih bertipe data obj diubah ke dalam tipe data integer.

```
# mengubah tipe data obj menjadi integer
from sklearn import preprocessing as pp
le = pp.LabelEncoder()
df['TANGGAL LAHIR'] = le.fit_transform(df['TANGGAL LAHIR'])
df['UMUR(TH/BLN)'] = le.fit_transform(df['UMUR(TH/BLN)'])
df['TANGGAL MASUK'] = le.fit_transform(df['TANGGAL MASUK'])
df['MASA KERJA(TH/BLN)'] = le.fit_transform(df['MASA KERJA(TH/BLN)'])
df['JABATAN'] = le.fit_transform(df['JABATAN'])
df
```

Output :

Berikut hasil dari dataset yang telah di transformasi.

NIK	NAMA TENAGA KERJA	TANGGAL LAHIR	UMUR(TH/BLN)	TANGGAL MASUK	MASA KERJA(TH/BLN)	JABATAN	HK- Kuantitas Kerja	HK- Ketepatan Waktu	HK- Kualitas Kerja	PK- Pengetahuan Tentang Pekerjaan	PK- Komunikasi	PK- Kerj
009AMP18068708	Afrizal	86	9	38	0	19	30	40	40	30	30	
327AMP10037807	Dedi Iman Widiyanto	44	75	9	7	28	30	30	20	20	20	
002AMP13077908	Hirwansyah	28	58	21	3	32	30	30	30	30	30	
GA/GMP/0604/403	KIWAL	10	87	17	23	28	30	30	20	20	20	
307AMP09087102	Saudi	18	93	25	8	28	30	30	30	30	30	
...	
0524AMP00087707	Yuliswardi	99	69	22	38	20	30	30	40	30	40	
0522AMP96056712	Yusrizal	39	96	37	52	8	30	30	30	40	30	
0427GMP00088108	Edra Mulyadi	78	44	22	38	21	40	40	30	30	30	
003PMJ02017208	Muller Sitirjak	94	91	39	42	30	20	20	20	20	20	
0410AMP01067604	Zaharudin	72	77	16	33	30	20	20	20	40	30	

2WS x 16 columns

Mengecek info dataset.

```
# menampilkan info semua item data
df.info()
```

Output :

Berikut info dataset penilaian kinerja karyawan di PT.XYZ.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 121 entries, 0 to 120
Data columns (total 16 columns):
 #   Column                                          Non-Null Count  Dtype
---  -
0    NIK                                           121 non-null    object
1    NAMA TENAGA KERJA                           121 non-null    object
2    TANGGAL LAHIR                               121 non-null    int32
3    UMUR(TH/BLN)                                121 non-null    int32
4    TANGGAL MASUK                              121 non-null    int32
5    MASA KERJA(TH/BLN)                         121 non-null    int32
6    JABATAN                                       121 non-null    int32
7    HK-Kuantitas Kerja                          121 non-null    int64
8    HK-Ketepatan Waktu                         121 non-null    int64
9    HK-Kualitas Kerja                          121 non-null    int64
10   PK-Pengetahuan Tentang Pekerjaan           121 non-null    int64
11   PK-Komunikasi                              121 non-null    int64
12   PK-Kerjasama Tim                          121 non-null    int64
13   PK-Disiplin Kerja                         121 non-null    int64
14   INDEX Performannse                        121 non-null    int64
15   INDEX Potensial                          121 non-null    int64
dtypes: int32(5), int64(9), object(2)
memory usage: 12.9+ KB

```

Untuk menghasilkan matrik korelasi, item data 'NIK' dan 'NAMA TENAGA KERJA' tidak diperlukan. Sehingga pada item data 'NIK' dan 'NAMA' dihapus.

```

# menghilangkan item data yang tidak mempengaruhi visualisasi matrik korelasi
dfVisual = df.drop(['NIK','NAMA TENAGA KERJA'], axis = 1)
dfVisual.head()

```

Output :

Berikut hasil dataset yang telah diubah.

TANGGAL LAHIR	UMUR(TH/BLN)	TANGGAL MASUK	MASA KERJA(TH/BLN)	JABATAN	HK- Kuantitas Kerja	HK- Ketepatan Waktu	HK- Kualitas Kerja	PK- Pengetahuan Tentang Pekerjaan	PK- Komunikasi	PK- Kerjasama Tim	PK- Disiplin Kerja	INDEX Performannse	INDEX Potensial
88	9	38	0	19	30	40	40	30	30				
44	75	9	7	28	30	30	20	20	20				
28	58	21	3	32	30	30	30	30	30				
10	87	17	23	28	30	30	20	20	20				
18	93	25	8	28	30	30	30	30	30				

Setelah dilakukan data transformation dan data preprocessing untuk menghasilkan matrik korelasi yang menampilkan hubungan antar item pada dataset penilaian kinerja di PT.XYZ, dilanjutkan dengan visualisasi matrix korelasinya.

```

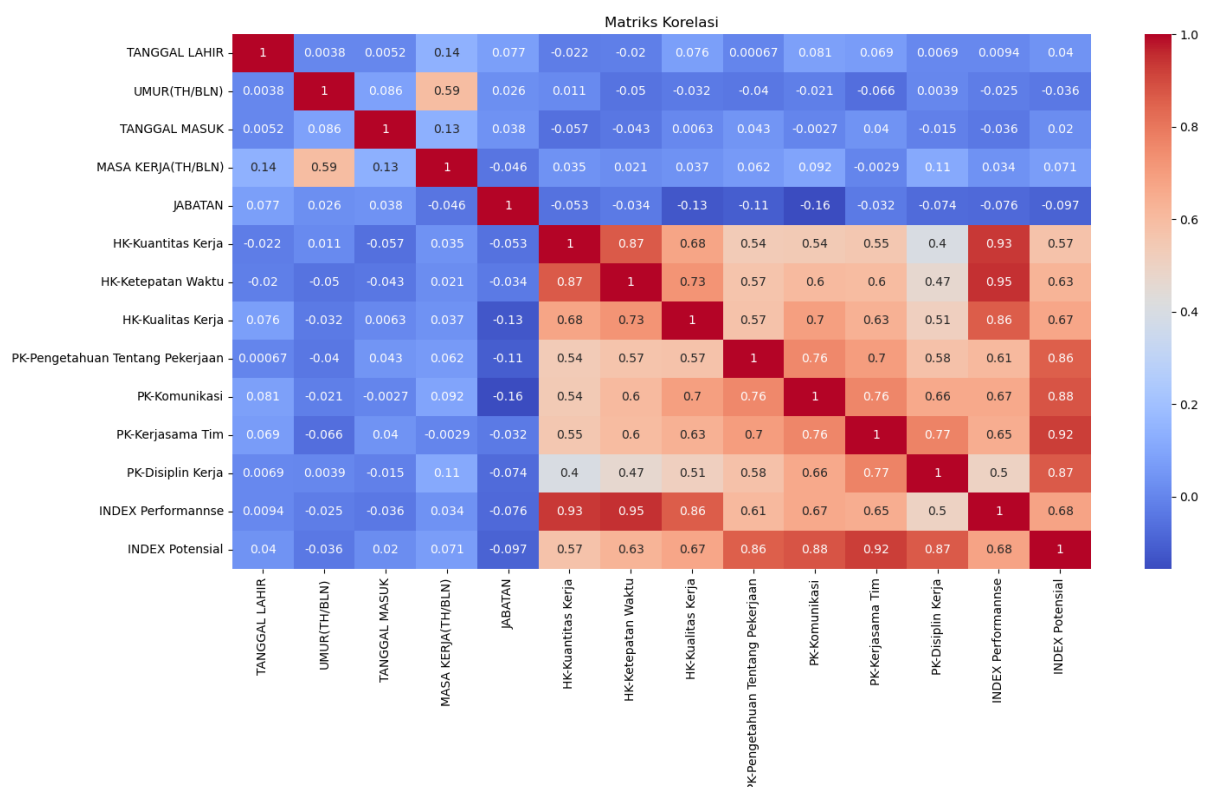
# visualisasi hubungan setiap item data dengan matrik korelasi

```

```
plt.figure(figsize=(16, 8))
sns.heatmap(dfVisual.corr(), annot=True, cmap='coolwarm')
plt.title('Matriks Korelasi')
plt.show()
```

Output :

Berikut hasil visualisasi matrik korelasi. Pada matrix ini range warna dari biru - merah menandakan tingkat hubungan antar item. Semakin mengarah ke warna merah menandakan semakin kuat hubungan antar kedua item tersebut. Sebaliknya, semakin mengarah ke warna biru menandakan semakin tidak ada hubungan antar kedua item.



Gambar 4 10 Matrix Kolerasi

4. Data Preparation

Preparation data merupakan proses menyiapkan data mentah sehingga layak untuk diproses dan dianalisis lebih lanjut. Dari banyak item data yang terdapat pada dataset penilaian kinerja karyawan di PT.XYZ, hanya membutuhkan tiga item untuk diproses lebih lanjut menggunakan metode Clustering algoritma K-Means. Item yang diperlukan yaitu 'NIK' sebagai pembeda antar item, 'Index Performance', dan 'Index Potensial' yang merupakan hasil akhir penilaian kinerja karyawan.

Mengambil item data yang diperlukan kemudian disimpan kedalam variabel dataOlah.

```
dataOlah = df[['NIK','INDEX Performance', 'INDEX Potensial' ]]  
dataOlah
```

Output :

Berikut hasil tabel data terbaru yang akan digunakan untuk pengolahan data lebih lanjut.

	NIK	INDEX Performannse	INDEX Potensial
0	009AMP18088708	33	39
1	327AMP10037607	24	22
2	002AMP13077908	27	39
3	GA/GMP/0604/403	24	28
4	307AMP09087102	27	33
...
116	0524AMP00087707	30	35
117	0522AMP96056712	27	36
118	0427GMP00088108	33	33
119	003PMJ02017208	18	22
120	0410AMP01067604	18	42

121 rows × 3 columns

5. Analisis Data

Setelah melalui data preparation, data siap untuk dianalisis. Analisis data adalah proses pengolahan data bertujuan untuk menemukan informasi yang berguna untuk dijadikan sebagai dasar pengambilan keputusan untuk memecahkan suatu masalah.

Membuat variabel baru yaitu dataClus yang menyimpan nilai dari 'Index Performance' dan 'Index Potensial' yang digunakan untuk pembuatan grafik elbow untuk menentukan jumlah cluster data yang akan digunakan.

```
dataClus = dataOlah.drop(['NIK'], axis = 1)  
dataClus
```

Output :

Berikut tabel dari variabel dataClus.

	INDEX Performannse	INDEX Potensial
0	33	39
1	24	22
2	27	39
3	24	28
4	27	33
...
116	30	35
117	27	36
118	33	33
119	18	22
120	18	42

121 rows × 2 columns

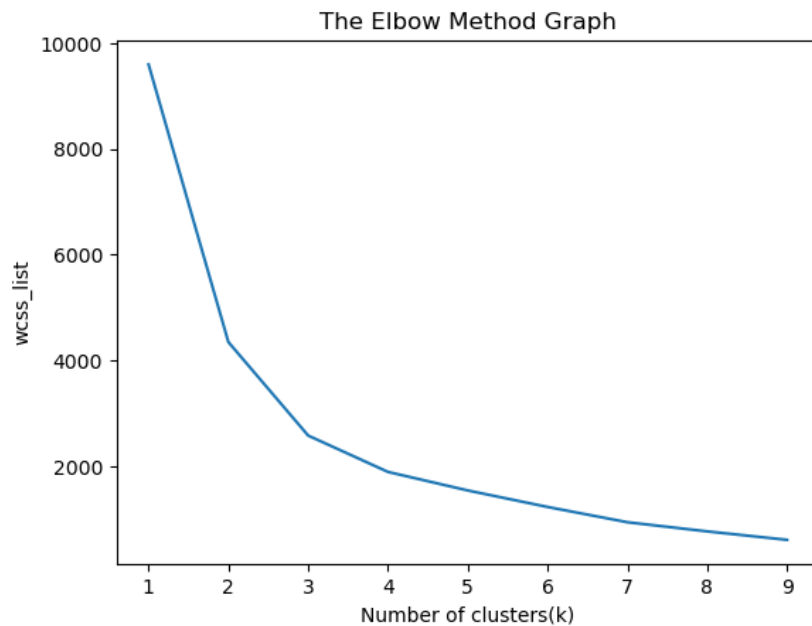
Membuat grafik elbow method untuk mencari nilai k yang akan digunakan dalam pengklasteran data, dengan bantuan library KMeans dari sklearn.cluster.

```
# menggunakan grafik elbow untuk mencari nilai k yang tepat
from sklearn.cluster import KMeans
wcss_list = []

for i in range(1,10):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
    kmeans.fit(dataClus)
    wcss_list.append(kmeans.inertia_)
plt.plot(range(1,10),wcss_list)
plt.title('The Elbow Method Graph')
plt.xlabel('Number of clusters(k)')
plt.ylabel('wcss_list')
plt.show()
```

Output :

Berikut hasil grafik menggunakan metode elbow. Dapat dilihat pada grafik dibawah, pada saat nilai $k = 3$ line grafik sudah mengalami penurunan yang stabil. Maka jumlah cluster yang akan dibuat ialah $k = 3$.



Gambar 4 11 Grafik Elbow Method

Menyimpan variabel feature yang akan digunakan ke dalam variabel X.

```
# memilih variabel feature yang akan digunakan
X = df[['INDEX Performannse', 'INDEX Potensial']]
X
```

Output :

Berikut tabel dari variabel feature X yang digunakan.

	INDEX Performannse	INDEX Potensial
0	33	39
1	24	22
2	27	39
3	24	28
4	27	33
...
116	30	35
117	27	36
118	33	33
119	18	22
120	18	42

Membuat model algoritma K-Means dengan membuat objek 'Kmeans' dengan inisialisasi parameter jumlah cluster $k = 3$, `init = 'k-means++'` sebagai metode inisialisasi pusat cluster dengan metode `k-means++` yang memilih pusat cluster awal dengan cerdas agar tersebar secara merata di seluruh ruang fitur, dan `random_state = 42` sebagai argumen opsional untuk mengontrol inisialisasi acak dalam algoritma dengan hasil yang konsisten.

Setelah objek 'Kmeans' diinisialisasikan, latih model 'Kmeans' menggunakan metode 'fit' dengan data yang sebelumnya telah disimpan dalam variabel X.

```
# membuat model algoritma K-Means dengan nilai K yang telah ditentukan
# dengan nilai K = 3
kmeans = KMeans(n_clusters=3, init='k-means++', random_state = 42)
predict = kmeans.fit_predict(X)
```

Menyimpan hasil clustering kedalam variabel labels

```
# menyimpan labels clustering kedalam variabel labels
labels = kmeans.labels_
```

Menampilkan titik pusat clusters awal (centroids) di setiap item data pada tiap clusternya.

```
# menampilkan pusat cluster
kmeans.cluster_centers_
```

Output :

Hasil dari titik pusat clusters awal disimpan kedalam array. Centroid cluster pertama dengan nilai 27.046 dan 33.984, kemudian centroid cluster kedua dengan nilai 19.852 dan 24.676, dan centroid cluster ketiga dengan nilai 32.454, 43.272. Berikut hasilnya.

```
array([[27.04615385, 33.98461538],
       [19.85294118, 24.67647059],
       [32.45454545, 43.27272727]])
```

Menyimpan hasil cluster dari setiap atribut kedalam item baru yang diberi nama 'Clusters' yang disimpan kedalam variabel dataset dataOlah.

```
# menampilkan tabel data beserta label clusternya
dataOlah['Clusters'] = kmeans.labels_
dataOlah.head(121)
```

Output :

Berikut tabel dari dataset variabel dataOlah yang telah ditambahkan dengan item 'Clusters'.

	NIK	INDEX Performannse	INDEX Potensial	Clusters
0	009AMP18068708	33	39	2
1	327AMP10037607	24	22	1
2	002AMP13077908	27	39	0
3	GA/GMP/0604/403	24	28	1
4	307AMP09087102	27	33	0
...
116	0524AMP00087707	30	35	0
117	0522AMP96056712	27	36	0
118	0427GMP00088108	33	33	0
119	003PMJ02017208	18	22	1
120	0410AMP01067604	18	42	0

121 rows × 4 columns

Menampilkan hasil cluster dalam bentuk array. Memanggil predict yang merupakan hasil prediksi label cluster pada dataset yang telah diolah berdasarkan model K-means yang telah dilatih sebelumnya

```
# menampilkan cluster item data dalam bentuk array
predict
```


Output :

Berikut hasil cluster tiap atribut dalam bentuk array.

```
array([2, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 2, 1, 0, 1, 2, 0, 1, 1, 1, 2, 0,
       0, 0, 0, 1, 0, 1, 2, 0, 0, 1, 2, 2, 2, 1, 0, 2, 0, 0, 0, 2, 2, 1,
       1, 0, 2, 2, 1, 0, 0, 1, 0, 0, 0, 2, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0,
       0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 0, 2, 0, 0,
       0, 0, 2, 1, 1, 1, 1, 1, 0, 1, 2, 2, 0, 0, 0, 0, 0, 0, 1, 0, 2, 1,
       0, 2, 0, 1, 2, 0, 0, 0, 0, 1, 0])
```

Menampilkan total karyawan di setiap clusternya. Menggunakan metode value_counts yang ada dalam library pandas untuk menghitung jumlah kemunculan nilai unik dalam setiap item data.

```
# menampilkan cluster beserta jumlah item tiap cluster
dataOlah['Clusters'].value_counts()
```

Output :

Hasil menampilkan bahwa cluster 0 sejumlah 65 record data, cluster 1 sebanyak 34 record data, dan cluster 2 sebanyak 22 record data. Berikut hasil outputnya.

```
0    65
1    34
2    22
Name: Clusters, dtype: int64
```

Memvisualisasikan jumlah data tiap cluster menggunakan bar chart.

```
# visualisasi jumlah tiap clusters
ax = dataOlah['Clusters'].value_counts().plot(kind='bar',
                                              figsize=(14,8),
                                              title="Total Karyawan tiap Clusters",color='palevioletred')
ax.set_xlabel("Cluster")
ax.set_ylabel("Frequency")
```

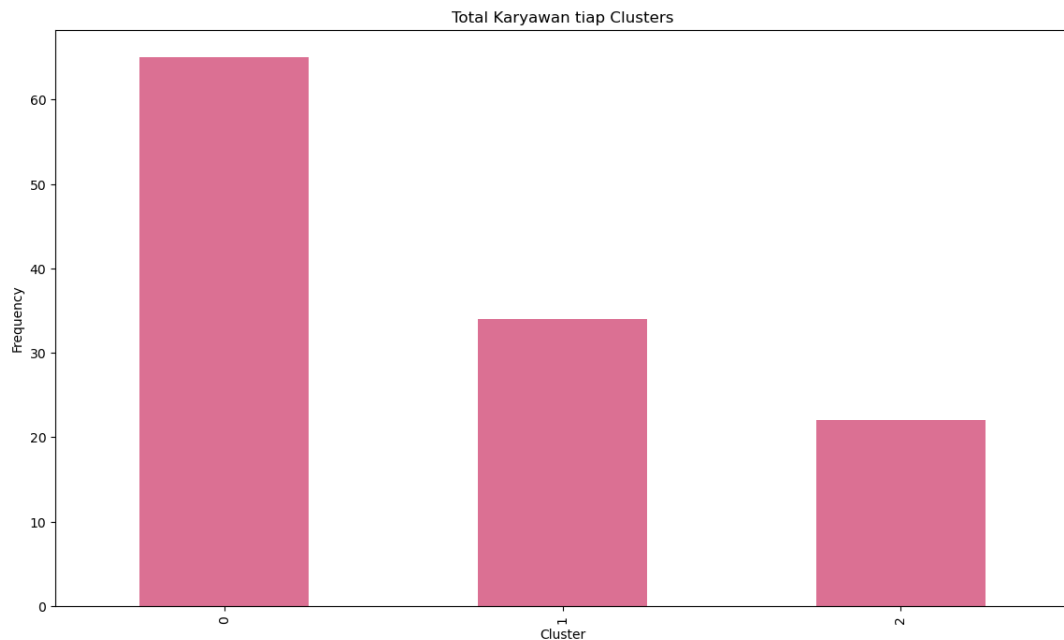
Output :

Cluster 0 = 65 record

Cluster 1 = 34 record

Cluster 2 = 22 record

Berikut bar chart visualisasinya.



Gambar 4 12 Bar Chart Perbandingan Jumlah Setiap Cluster

Menampilkan NIK yang terdapat di setiap clusters. NIK tersebut akan disimpan ke dalam arraylist yang diberi nama `clusters_list`. Kemudian inisialisasikan perulangan `for` untuk menampilkan NIK yang ada tiap clusters, dari clusters 0 - 2. Data NIK dan Cluster diambil berdasarkan dataset yang terdapat pada variabel `dataOlah`.

```
# membuat arraylist kosong untuk menyimpan data tiap cluster
clusters_list = []

# menampilkan list dari anggota tiap cluster data
for i in range(0, max(dataOlah['Clusters'])+1):
    print('Cluster', i)
    print(dataOlah[dataOlah['Clusters']==i]['NIK'].to_list())
    print("")
```

Output :

Berikut hasil array yang menampilkan NIK karyawan di setiap clusternya

```

Cluster 0
['002AMP13077908', '307AMP09087102', '002AMP14069207', '324AMP11028703', '305AMP01117902', '321AMP110
28204', '324AMP10098608', '322AMP05078407', '323AMP11029004', '159AMP07016903', '328AMP08108012', '04
95AMP01097805', '0471AMP00057605', '326AMP05078010', '323AMP05078008', '0473AMP00087103', '0468AMP020
77212', '175AMP06128207', '0475AMP00057611', '326AMP08107710', '0422AMP00086912', '0424AMP97036811',
'0488AMP02077810', '0478AMP00097404', '325AMP05078303', '178AMP06068003', '309AMP11028305', '324AMP05
098208', '181AMP06048507', '323AMP08058410', '160AMP07048611', '308AMP11028912', '176AMP06088112', '3
19AMP02097901', '161AMP07048903', '307AMP11028402', '317AMP11028705', '0421AMP00097902', '174AMP06118
011', '312AMP11028607', '308AMP03058201', '320AMP05078505', '0457AMP98077707', '319AMP03057904', '325
AMP08108208', '0523AMP96107111', '327AMP11028504', '332AMP05078408', '329AMP05088507', '0430AMP981267
12', '330AMP08087908', '167AMP99097508', '318AMP03038106', '320AMP08088409', '327AMP05078002', '0482A
MP00097304', '309AMP01098005', '0479AMP00098113', '330AMP05078512', '302AMP04127906', '321AMP0504810
8', '0524AMP00087707', '0522AMP96056712', '0427GMP00088108', '0410AMP01067604']

Cluster 1
['327AMP10037607', 'GA/GMP/0604/403', '0493AMP03057303', '313AMP01028104', '180AMP06067908', '329AMP0
8057901', '322AMP11027610', '177AMP06067308', '179AMP06068109', '0483AMP01048303', '313AMP11028207',
'325AMP11026702', '320AMP11028709', 'AG/PAMP/0119/183', '320AMP03058205', '320AMP02098512', '0455AMP9
6117312', '331AMP05078703', '0437AMP96106312', '310AMP11027706', '0465AMP99017406', '162AMP99117412',
'327AMP12028904', 'AG/PAMP/0119/189', '155AMP98037611', '311AMP02037810', '333AMP05077710', '326AMP11
028902', '0521AMP01047806', '322AMP97108001', '0497AMP04128308', '0411AMP99036612', '304AMP05098401',
'003PMJ02017208']

Cluster 2
['009AMP18068708', '0431AMP97036608', '317AMP02098010', '319AMP05078006', '310AMP00118201', '0490AMP0
3057607', '169AMP03058009', '0519AMP01048106', '0485AMP00097403', '0476AMP00087702', '157AMP9901771
2', '0677AMP00086412', '319AMP11028710', '327AMP08107905', '318AMP11028503', '173AMP06067908', '311AM
P11029004', '0480AMP00097206', '311AMP00027307', '0526AMP02077411', '314AMP05077702', '0466AMP0005821
2']

```

Membuat visualisasi data setiap cluster menggunakan library matplotlib. Pada grafik ini menampilkan titik letak data beserta centroidsnya.

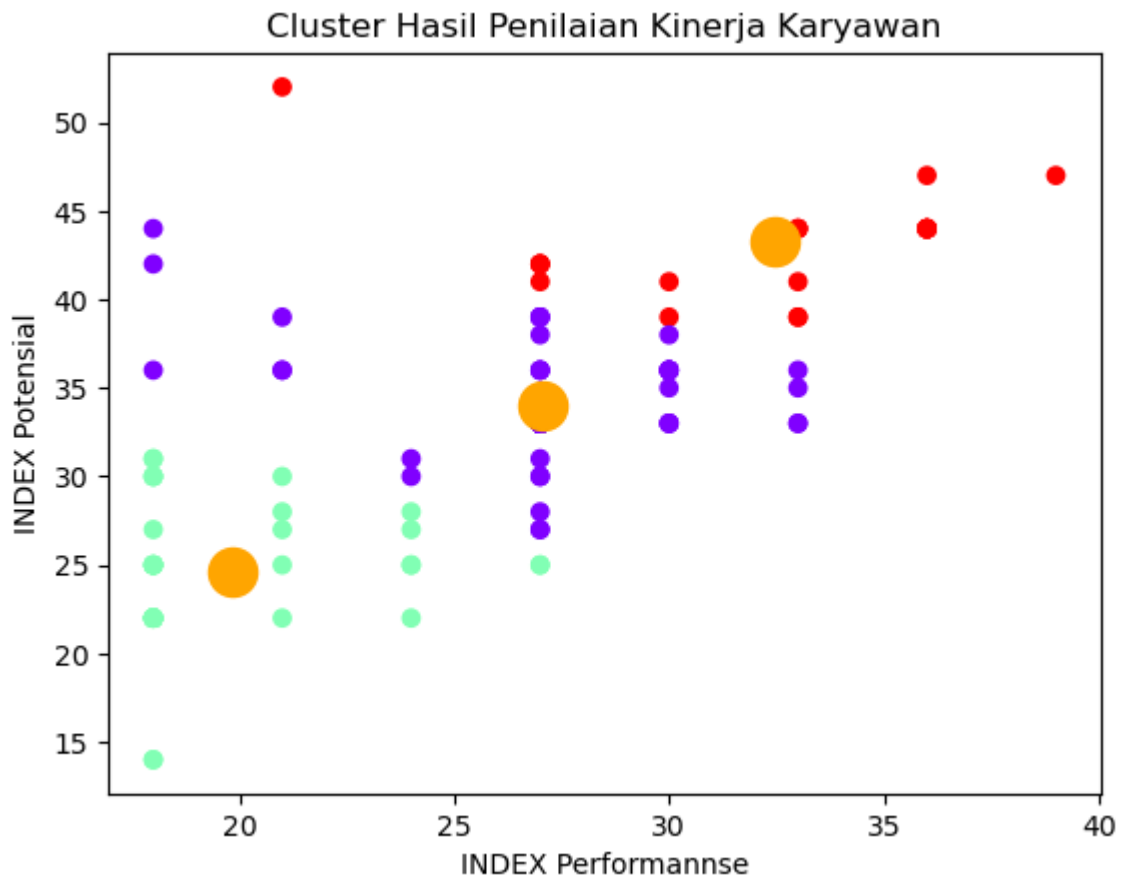
```

# visualisasi cluster data
plt.scatter(X.iloc[:, 0], X.iloc[:, 1], c=kmeans.labels_, cmap='rainbow')
plt.scatter(kmeans.cluster_centers_[0], kmeans.cluster_centers_[1], s=300,
c='orange')
plt.title("Cluster Hasil Penilaian Kinerja Karyawan")
plt.xlabel("INDEX Performannse")
plt.ylabel("INDEX Potensial")
plt.show()

```

Output :

Gambar grafik visualisasi data penilaian kinerja karyawan menggunakan metode clustering algoritma k-means



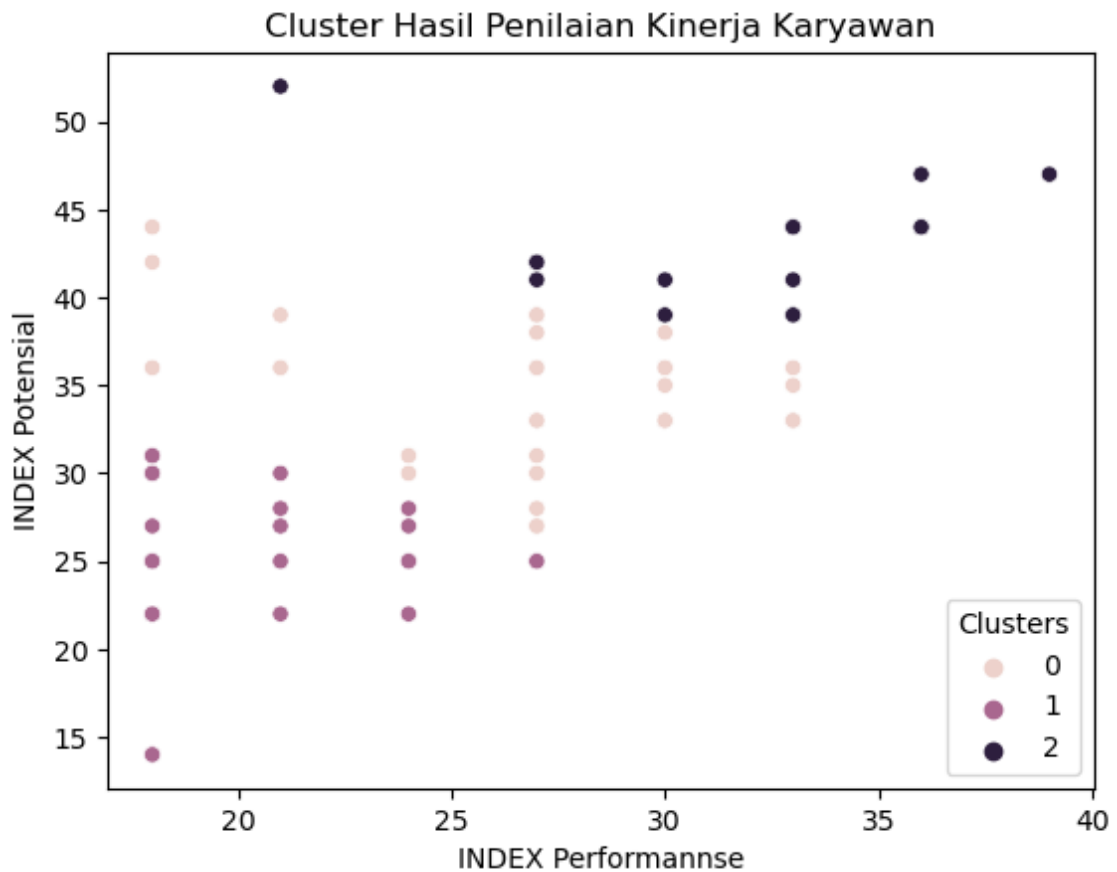
Gambar 4 13 Grafik Hasil K-Means Clustering

Membuat visualisasi data dalam bentuk ke-2 (bentuk yang berbeda) setiap cluster menggunakan library matplotlib. Pada grafik ini menampilkan titik letak data setiap clusters dengan warna berbeda antar clusternya

```
# visualisasi cluster data dlm bentuk ke 2 (tanpa titik centroids)
sns.scatterplot(x='INDEX Performannse', y='INDEX Potensial', hue = 'Clusters',
data = dataOlah)
plt.title("Cluster Hasil Penilaian Kinerja Karyawan")
plt.xlabel("INDEX Performannse")
plt.ylabel("INDEX Potensial")
```

Output :

Gambar grafik visualisasi data penilaian kinerja karyawan menggunakan metode clustering algoritma k-means



Gambar 4 14 Grafik Hasil K-Means Clustering versi 2

Menyimpan tabel hasil pengolahan data yang disimpan dalam variabel dataOlah. Variabel dataOlah ini akan di ekstrak dalam file csv yang diberi nama 'HasilClusterData.csv'.

```
# menyimpan hasil pengclusteran data
dataOlah.to_csv('HasilClusterData.csv', index = False)
```

6. Perbandingan hasil k-means clustering secara manual di excel dengan python

Mengimpor modul yang diperlukan yaitu 'adjusted_rand_score' dari modul 'sklearn.metrics' yang terdapat pada 'KMeans' untuk melakukan perhitungan nilai Adjusted Rand Score. Adjusted Rand Score berfungsi untuk mengukur kesamaan antara label cluster yang diperoleh dari perhitungan manual menggunakan excel dengan label cluster yang dihasilkan K-Means. ground_truth_labels berisi cluster data pada pengolahan secara manual di excel. Nilai ars berkisar dari -1 hingga 1, dengan 1 menunjukkan kesesuaian sempurna antara label cluster dan label ground truth.

Menampilkan hasil nilai akurasi tingkat kesamaan data yang diperoleh secara manual dengan k-means label yang disimpan dalam variabel ars.

```
from sklearn.metrics import adjusted_rand_score

# Menghitung nilai ars dengan label kluster dan label ground truth
ground_truth_labels = [2, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 2, 1, 0, 1, 0, 0, 1, 1, 1, 2, 0, 0, 0,
0, 1, 0, 1, 2, 0, 0, 1, 2, 2, 2, 1, 0, 0, 0, 0, 0, 2, 2, 1, 1, 0, 2, 2, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1,
0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 2, 0, 2, 0, 0, 0, 0, 2, 1,
1, 1, 1, 1, 0, 1, 2, 2, 0, 0, 0, 0, 0, 0, 1, 0, 2, 1, 0, 2, 0, 1, 2, 0, 0, 0, 0, 1, 0]
ars = adjusted_rand_score(ground_truth_labels, labels)

print("Adjusted Rand Score:", ars)
```

Output :

Hasil akurasi kesamaan datanya ialah 0.928 yang dikatakan hampir mendekati kesesuaian yang sempurna.

Adjusted Rand Score: 0.9280109924183163

4.4.2 Hasil Akhir

Hasil dari hasil akhir perhitungan pengelompokkan data penilaian kinerja karyawan di PT. XYZ menggunakan dataset 'Book1ARA_DataAsli_Olah' yang terdisri dari 121 records, maka diperoleh hasil bahwa :

a) Cluster 0

Pada cluster 0 memiliki 65 data yang diinisialisasikan sebagai kelompok pertama. Kelompok pertama ini diartikan sebagai kelompok dengan hasil penilaian kinerja karyawan 'baik'.

No	NIK	Nama Tenaga Kerja	Jabatan	Index Performance	Index Potensial
1	002AMP13077908	Hirwansyah	Sortase Helper	27	39

2	307AMP0908710 2	Saudi	Security	27	33
3	002AMP1406920 7	Syafrino	Sortase Helper	27	33
4	324AMP1102870 3	Ad Edianto	Process Operator	24	30
5	305AMP0111790 2	Amra Luis	Security	27	27
6	181AMP0604850 7	Narti	Office Girl	33	36
7	321AMP1102820 4	Ardiones	Capstan Operator	27	33
8	324AMP1009860 8	Bal Efendi	Sample Boy	30	33
9	322AMP0507840 7	Bulkaini	Sortase Foreman	24	31
10	323AMP1102900 4	Candra Kurniawan	Hoisting Crane Operator	27	33
11	159AMP0701690 3	Darmawi	Driver	27	30
12	328AMP0810801 2	Daswir	Boiler Operator	21	39
13	0495AMP010978 05	Deris Suanto	KCP Operator	27	33
14	0471AMP000576 05	Deswito	Pressing Operator	27	33
15	326AMP0507801 0	Ermansyah	Process Helper	27	33
16	324AMP0509820 8	Musliadi	Process Helper	27	33
17	323AMP0507800 8	Gusmaldi	Process Helper	30	36
18	0473AMP000871 03	Gusti Metri	Transfer Carriage Operator	27	33
19	0468AMP020772 12	Gustiar	Clarification Operator	27	30

20	309AMP1102830 5	Muhammad Reki	Boiler Operator	18	44
21	175AMP0612820 7	Hendri Yonanda	Sample Boy	27	33
22	0475AMP000576 11	Irmien Suhita	KCP Operator	27	33
23	326AMP0810771 0	Jumanto	Maintenance Helper	27	38
24	0422AMP000869 12	Khairul Amri	Effluent Operator	30	36
25	0424AMP970368 11	M. Nur Alamsyah	Sortase Operator	27	36
26	0488AMP020778 10	Madrizel Miyanto	Treshing Operator	27	33
27	178AMP0606800 3	Maysar	Capstan Operator	27	39
28	0478AMP000974 04	Masriadi	Boiler Operator	27	33
29	325AMP0507830 3	Masrizel	Process Helper	27	39
30	323AMP0805841 0	Nofi Hendri	Logistic Clerk	33	35
31	160AMP0704861 1	Oki Ilham	Mechanical Helper	27	31
32	308AMP1102891 2	Rahmat Effendi	Mechanic	30	38
33	176AMP0608811 2	Ratijan	Boiler Operator	27	33
34	319AMP0209790 1	Rifnaldi	Process Helper	27	33
35	161AMP0704890 3	RIJAL ISLAMI	Capstan Operator	27	33
36	307AMP1102840 2	Rinaldi Rahmanal Pendi	Electrician	27	33
37	317AMP1102870 5	Riska Joni Putra	Sortase Operator	27	36

38	0421AMP000979 02	Rismanto	Sortase Operator	27	33
39	174AMP0611801 1	Rivai	Sortase Foreman	27	33
40	312AMP1102860 7	Rodi Hartono	Store Keeper	30	33
41	308AMP0305820 1	Romi	Sortase Operator	27	33
42	320AMP0507850 5	Rudi Hartono	Hoisting Crane Operator	30	36
43	0457AMP980777 07	Sabir	Heavy Equipment Operator	27	33
44	319AMP0305790 4	Sabri	Sortase Foreman	27	33
45	325AMP0810820 8	Saruji Ismael	Well Loader Operator	27	33
46	0523AMP961071 11	Sixteen Nr	KCP Operator	27	33
47	327AMP1102850 4	SUKIRMAN	Capstan Operator	27	33
48	332AMP0507840 8	Supirman	Capstan Operator	27	27
49	329AMP0508850 7	Supri Endi	Sortase Helper	33	33
50	0430AMP981267 12	Supriadi	Land Application Operator	30	36
51	330AMP0808790 8	Warniati	Office Girl	27	28
52	167AMP9909750 8	Yuni Sholekah	Office Girl	27	33
53	318AMP0303810 6	Yunus Effendi	Process Operator	21	36
54	320AMP0808840 9	Zul Hendra	Maintenance POM	30	33
55	327AMP0507800 2	Zulfikar	Boiler Operator	18	36

56	0482AMP000973 04	Zulmi Wardi	WTP Operator	27	33
57	309AMP0109800 5	Adel Kurnia Putra	Administration Clerk	27	33
58	0479AMP000981 13	Basrial B	Hoisting Crane Operator	27	33
59	330AMP0507851 2	Desri	Capstan Operator	27	33
60	302AMP0412790 6	Ismel Antone	Sterilizer Operator	30	36
61	321AMP0504810 8	Saruji Ismail	Security	21	36
62	0524AMP000877 07	Yuliswardi	Mechanical Operator	30	35
63	0522AMP960567 12	Yusrizal	Engine Room Operator	27	36
64	0427GMP000881 08	Edra Mulyadi	Mechanical Shift Filter	33	33
65	0410AMP010676 04	Zaharudin	Security (Ka)	18	42

Tabel 4 6 Hasil Akhir Pengelompokan kinerja karyawan Cluster 0

Berdasarkan hasil cluster 0 dapat diketahui terdapat 65 data karyawan. Tabel diatas berisikan informasi karyawan mulai dari NIK, Nama, Jabatan, serta index penilaian kinerja karyawan. Jumlah data pada cluster 0 merupakan jumlah data yang paling banyak dibanding cluster lain. Berdasarkan tabel diatas, dapat diketahui rentang nilai index performance yang dimiliki mulai dari 18-33 dan index potensial mulai dari 27- 44 dengan kombinasi nilai dari index performance dan index potensial yang menjadi faktor penentu dalam pengelompokan.

b) Cluster 1

Pada cluster 1 memiliki 34 data yang yang diinisialisasikan sebagai kelompok kedua. Kelompok kedua ini diartikan sebagai kelompok dengan hasil penilaian kinerja karyawan ‘kurang baik’.

No	NIK	Nama Tenaga Kerja	Jabatan	Index Perfor mance	Index Potensi al
----	-----	-------------------	---------	--------------------	------------------

1	327AMP10037607	Dedi Iman Widiyanto	Security	24	22
2	GA/GMP/0604/403	KIWAL	Security	24	28
3	0493AMP03057303	Adrison	L. Ramp Operator	21	25
4	313AMP01028104	Afrizal Arlin	Process Operator	18	25
5	180AMP06067908	Agus Wandu	Security	18	14
6	329AMP08057901	Agusman	Sortase Helper	24	25
7	322AMP11027610	Amir Johan	Sterilizer Operator	18	22
8	177AMP06067308	Amrizal	Boiler Operator	21	30
9	179AMP06068109	Aris	Transfer Carriage Operator	18	22
10	0483AMP01048303	Asril	L. Ramp Operator	18	22
11	313AMP11028207	Asril	L. Ramp Operator	18	22
12	325AMP11026702	Dasril	Mechanical Helper	18	25
13	320AMP11028709	Dedrinaldi	Capstan Operator	18	25
14	AG/PAMP/0119/18 3	Dodi Agusman	Security	18	22
15	320AMP03058205	Emrizal	Process Helper	21	22
16	320AMP02098512	Hendra. A	Process Helper	18	25
17	0455AMP96117312	Hendra.y	Engine Driver	18	25
18	331AMP05078703	Ipen Riawan	Capstan Operator	18	22
19	0437AMP96106312	Kenek	PGA Staff	18	22
20	310AMP11027706	Mardianto	Sterilizer Operator	18	22
21	0465AMP99017406	Mardinas	Clarification Operator	18	14
22	162AMP99117412	Masrizal	Security	27	25
23	327AMP12028904	Riko Martono	Security	24	25
24	AG/PAMP/0119/18 9	SARIYUL EFENDI	Security	24	27
25	155AMP98037611	Suwardi	Heavy Equipment Operator	27	25

26	311AMP02037810	Syafriyal	Process Operator	21	28
27	333AMP05077710	Syahrul	Mechanical Helper	18	30
28	326AMP11028902	Syahrul Ujud	Mechanical Helper	21	27
29	0521AMP01047806	Syarhadi Syam	KCP Operator	18	27
30	0497AMP04128308	Anton Masrial	Store Clerk	18	31
31	322AMP97108001	Yati Murni	Weigh Bridge Operator	18	31
32	0411AMP99036612	Desman	Security (Danru)	18	30
33	304AMP05098401	Joni Hendrika. P	Security (Danru)	18	30
34	003PMJ02017208	Muller Sitinjak	Security (Ka)	18	22

Tabel 4 7 Hasil Akhir Penilaian Kinerja Kryawan Cluster 1

Berdasarkan hasil cluster 1 dapat diketahui terdapat 34 data karyawan. Tabel diatas berisikan informasi karyawan mulai dari NIK, Nama, Jabatan, serta index penilaian kinerja karyawan. Jumlah data pada cluster 1 lebih banyak dari pada cluster ke 2 namun lebih sedikit dari cluster 0. Berdasarkan tabel diatas, dapat diketahui rentang nilai index performance yang dimiliki mulai dari 18-27 dan index potensial mulai dari 14- 30 dengan kombinasi nilai dari index performance dan index potensial yang menjadi faktor penentu dalam pengelompokkan.

c) Cluster 2

Pada cluster 2 memiliki 22 data yang yang diinisialisasikan sebagai kelompok ketiga. Kelompok ketiga ini diartikan sebagai kelompok dengan hasil penilaian kinerja karyawan ‘sangat baik’.

No	NIK	Nama Tenaga Kerja	Jabatan	Index Perfor mance	Index Potensi al
1	009AMP18068708	Afrizal	Mechanical Helper	33	39
2	0431AMP97036608	Ajang Nugraha	Sample Boy	39	47
3	317AMP02098010	Andi Sumarno	Process Operator	21	52
4	319AMP05078006	Asril	Maintenance POM	33	44
5	310AMP00118201	Delita Effendi	Process Operator	33	39

6	0490AMP03057607	Dodi Eka Edison	KCP Operator	36	44
7	169AMP03058009	Doni Hendra	KCP Operator	33	44
8	0519AMP01048106	Eka Nofianda	KCP Operator	33	41
9	0476AMP00087702	Haryodis	Sterilizer Operator	33	44
10	0677AMP00086412	Hendriful	Clarification Operator	36	44
11	319AMP11028710	Indra Bangsawan	L. Ramp Operator	36	44
12	0485AMP00097403	Erwin	Land Application Operator	30	39
13	157AMP99017712	Hendra	Pressing Operator	30	41
14	327AMP08107905	Maihendri	Engine Room Operator	27	41
15	318AMP11028503	Simet	Effluent Operator	27	42
16	173AMP06067908	Sufwendi	Sortase Foreman	27	42
17	311AMP11029004	Suprinaldi	Sample Boy	27	42
18	0466AMP00058212	Rahmat Efendi	Clarification Operator	36	44
19	0480AMP00097206	Yondrizal	KCP Operator	36	44
20	311AMP00027307	Yotti Novrianto	Process Helper	36	44
21	0526AMP02077411	Dedi Arnanda	Jr. Analyst	36	47
22	314AMP05077702	Febrinaldi	Sterilizer Operator	36	44

Tabel 4.8 Hasil Akhir Penilaian Kinerja Karyawan Cluster 2

Berdasarkan hasil cluster 2 dapat diketahui terdapat 22 data karyawan. Tabel diatas berisikan informasi karyawan mulai dari NIK, Nama, Jabatan, serta index penilaian kinerja karyawan. Jumlah data pada cluster 2 merupakan jumlah data paling sedikit dalam pengelompokkan ini. Berdasarkan tabel diatas, dapat diketahui rentang nilai index performance yang dimiliki mulai dari 21-39 dan index potensial mulai dari 39- 52 dengan kombinasi nilai dari index performance dan index potensial yang menjadi faktor penentu dalam pengelompokkan.

4.5 Perbandingan Hasil Perhitungan

Berdasarkan dari hasil penelitian yang telah dilakukan pada data penilaian kinerja karyawan di PT. XYZ yang menggunakan dataset penilaian pada satu periode penilaian.

Selain menggunakan perhitungan manual dengan aplikasi Microsoft Excel, penelitian ini juga menggunakan aplikasi data mining Jupyter Notebook. Adapun perbandingan hasil perhitungannya adalah sebagai berikut :

Algoritma K-Means				
Perbandingan	C0	C1	C2	Total
Microsoft Excel	72	34	15	121
Jupyter Notebook	65	34	22	121
Selisih	7	0	7	14

Tabel 4 9 Perbandingan Hasil Akhir Pengelompokkan

Pada tabel diatas menunjukkan bahwa terjadinya selisih hasil perhitungan dalam menggunakan Microsoft Excel dan Jupyter Notebook. Hal ini dikarenakan penentuan centroids awal yang tidak sama. Menentukan centroids awal pada Microsoft Excel, peneliti menentukan sampel titik centroids secara acak. Sedangkan pada Jupyter Notebook dengan bahasa pemograman python dalam menentukan centroids awal menggunakan metode k-means++ yang memilih pusat cluster awal dengan cerdas agar tersebar secara merata di seluruh ruang fitur.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian pada PT. XYZ, maka dapat diambil beberapa kesimpulan sebagai berikut :

1. Metode Clustering K-Means dapat membantu mengelompokkan kinerja karyawan berdasarkan faktor-faktor yang mempengaruhi pada PT. XYZ.

2. Pengelompokkan kinerja karyawan pada penelitian ini dibagi atas: karyawan kinerja sangat baik, baik, dan kurang baik
3. Pengelompokkan kinerja karyawan baik, kurang baik, sangat baik diperoleh setelah melalui proses perhitungan dengan menerapkan algoritma K-Means Clustering selesai.
4. Hasil dari pengujian yang telah dilakukan dengan 61 data sampel, maka terbentuk data kelompok karyawan dengan kinerja baik terdiri dari 29 (dua puluh sembilan) karyawan, data kelompok karyawan dengan kinerja kurang baik terdiri dari 22 (dua puluh dua) karyawan, dan data kelompok sangat baik 10 (sepuluh) karyawan.
5. Hasil dari pengelompokkan kinerja ini akan membantu PT. XYZ dalam pengambilan keputusan untuk pengembangan karir karyawan sehingga target dan tujuan perusahaan dapat tercapai.

5.2 Saran

Penelitian ini masih memiliki kekurangan, sehingga untuk pengembangan penelitian berikutnya dituliskan saran-saran sebagai berikut:

1. Kriteria-kriteria yang mempengaruhi pengelompokkan kinerja karyawan pada PT. XYZ ini dapat ditambahkan sehingga hasil pengelompokkan yang didapatkan lebih objektif.
2. Pada pengembangan penelitian selanjutnya dapat dilakukan dengan jumlah data yang lebih besar lagi.
3. Dapat dilakukannya pengimplementasian berdasarkan hasil penelitian pada aplikasi-aplikasi penerapan data mining, seperti *RapidMiner*.

DAFTAR PUSTAKA

- [1] T. Penulis *et al.*, *KINERJA KARYAWAN*. 2021. [Online]. Available: www.penerbitwidina.com
- [2] A. Ismail *et al.*, “INDONESIA KLASIFIKASI JAJANAN TRADISIONAL JAWA TENGAH DENGAN METODE TRANSFER LEARNING DAN MOBILNETV2 IMPLEMENTASI METODE K-MEANS CLUSTERING PADA PENILAIAN

- KINERJA KARYAWAN PT KOPETRI CITRA ABADI INTEGRASI SISTEM PRESENSI BIOMETRIK DENGAN SISTEM INFORMASI SUMBER DAYA MANUSIA UNIVERSITAS MUHAMMADIYAH SEMARANG ANALISIS PENGARUH PENERAPAN CRM PADA APLIKASI POSAJA TERHADAP LOYALITAS PENGGUNA.” [Online]. Available: <http://e-journal.janabadra.ac.id/>
- [3] S. Regina, E. Sutinah, and N. Agustina, “Clustering Kualitas Kinerja Karyawan Pada Perusahaan Bahan Kimia Menggunakan Algoritma K-Means,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 5, no. 2, p. 573, Apr. 2021, doi: 10.30865/mib.v5i2.2909.
- [4] A. Dan, N. Yona, S. Munti, G. W. Nurcahyo, and J. Santony, “Analisis Dan Penerapan Data Mining Untuk Menentukan Gaji Karyawan Tetap Dan Karyawan Kontrak Menggunakan Algoritma K-Means Clustering (Studi Kasus Di Pt Indomex Dwijaya Lestari).”
- [5] A. Nur Khomarudin, “Teknik Data Mining : Algoritma K-Means Clustering,” 2003. [Online]. Available: <https://agusnkhom.wordpress.com>
- [6] V. S. Moertini, “DATA MINING SEBAGAI SOLUSI BISNIS,” 2002.
- [7] Inc. Jhon Willey & Sons, *Discovering Knowledge In Data: An Introduction To Data Mining*. . Larose, D. T. , 2005.
- [8] B. S. Budi Santosa, *Data mining : Teknik pemanfaatan data untuk keperluan bisnis / Budi Santosa*. Yogyakarta: Garah Ilmu, 2007.
- [9] “Machine Learning, Data Mining, and Knowledge Discovery: An Introduction.” [Online]. Available: http://www.kdnuggets.com/data_mining_course/
- [10] Retno Tri Wulandari, *Data Mining Teori dan Aplikasi Rapidminer*. 2017.
- [11] J. P. H. J. L. I. kosasih Kotter, *Dampak budaya perusahaan terhadap kinerja /John P. Kotter, James I. Heskett; alih bahasa, Benyamin Molan; penyunting bahasa, Kosasih Iskandarsyah*. Jakarta : Prenhallindo, 1997.
- [12] M. Simanjuntak, E. Manik, T. Supratman, and S. Kaputama, “PENERAPAN DATA MINING PENGELOMPOKAN KEJAHATAN ELEKTRONIK SESUAI UU ITE DENGAN MENGGUNAKAN METODE CLUSTERING,” *Jurnal Mahajana Informasi*, vol. 3, no. 2, 2018.

- [13] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.