JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186

DASHBOARD UNTUK CLUSTERING PENILAIAN PEGAWAI DINAS PU SUMBER DAYA AIR PROVINSI JAWA TIMUR

Rifdah Alifia*[1)] Rokhmatul Insani[2)] Sri Hidayati[3)]
1. Sistem Informasi, Fakultas Rekayasa Industri, Telkom University Surabaya, Indonesia
2. Sistem Informasi, Fakultas Rekayasa Industri, Telkom University Surabaya, Indonesia
3. Sistem Informasi, Fakultas Rekayasa Industri, Telkom University Surabaya, Indonesia

Article Info
Kata Kunci: dashboard; data mining; evaluasi kinerja; k-means clustering; penilaian pegawai

Keywords: *dashboard; data mining; employee assessment; k-means clustering; performance evaluation*

Article history:
Received 22 July 2024
Revised 10 September 2024
Accepted 3 October 2024
Available online 1 September 2025

DOI :
https://doi.org/10.29100/jipi.v10i3.6201

* Corresponding author.
Corresponding Author
E-mail address:
rifdahalifia@student.telkomuniversity.ac.id

ABSTRAK
Dinas PU Sumber Daya Air Provinsi Jawa Timur merupakan instansi milik pemerintahan yang memiliki wewenang dalam pengelolaan sumber daya air yang menyeluruh. Instansi tersebut memiliki evaluasi dan penilaian kinerja untuk para pegawai yang bekerja di dalamnya, dimana evaluasi tersebut digunakan untuk meningkatkan kinerja sumber daya manusia pada instansi terkait. Akan tetapi, evaluasi tersebut kurang komprehensif karena mereka belum memiliki sistem monitoring dan visualisasi untuk hasil pengelompokan penilaian pegawai. Oleh karena itu, penelitian ini dibuat yaitu dengan tujuan untuk menganalisis evaluasi dan penilaian prestasi kerja berdasarkan kelompok kinerja serta menganalisis variabel-variabel terkait yang digunakan untuk menilai para pegawai. Sasaran dari pengelompokan evaluasi kinerja pegawai yang dilakukan pada penelitian ini yaitu 128 pegawai ASN yang bekerja di Dinas PUSDA pada tahun 2023. Adapun metode yang digunakan dalam penelitian ini yaitu menggunakan algoritma *K-Means Clustering* berdasarkan data penilaian kinerja pegawai dengan 8 variabel, yaitu berorientasi pelayanan, akuntabel, kompeten, harmonis, loyal, adaptif, kolaboratif, dan hasil kerja. Penelitian ini menggunakan bahasa pemrograman Python dalam pengolahan datanya. Penentuan k optimal-nya menggunakan metode elbow, dengan hasil cluster terbaik adalah 3 cluster. Dimana *cluster* 0 memiliki 48 data, *cluster* 1 memiliki 74 data, dan *cluster* 2 memiliki 6 data. Dengan hasil evaluasi meggunakan metode *Silhouette Coefficient* sebanyak 0.593. Hasil dari *clustering* penilaian pegawai menggunakan K-Means digunakan untuk membuatan visualisasi *dashboard* pengelompokan kinerja pegawai menggunakan *tools* Tableau. Dengan adanya pengelompokan kinerja pegawai, diharapkan penelitian ini dapat memberi rekomendasi dalam pengambilan keputusan terkait pengembangan strategi pengelolaan SDM di Dinas PU Sumber Daya Air Provinsi Jawa Timur.

ABSTRACT
The East Java Provincial Water Resources Department is a government agency responsible for comprehensive water resource management. The agency conducts performance evaluations and assessments for its employees, aiming to enhance human resources performance. However, these evaluations are considered less comprehensive due to the absence of a monitoring and visualization system for employee evaluation groupings. Therefore, this study is conducted with the objective of analyzing performance evaluations based on performance groups and examining the related variables used to assess employees. The target of this performance evaluation grouping study includes 128 civil servants working at Dinas PUSDA in 2023. The method used in this research is the K-Means Clustering algorithm, based on employee performance evaluation data with eight variables: service-oriented, accountable, competent, harmonious, loyal, adaptive, collaborative, and work results. This study utilizes the Python programming language for data processing. The optimal number of clusters was determined using the elbow method, resulting in the best clustering into three clusters: Cluster 0 with 48 data points, Cluster 1 with 74 data points, and Cluster 2 with 6 data points. The evaluation using the Silhouette Coefficient method yielded a score of 0.593. The outcome of this research is a visualization of the employee performance grouping results in the form of a performance clustering dashboard using

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186

Tableau tools. It is expected that the employee performance grouping will provide recommendations for decision-making related to the development of human resource management strategies at the East Java Provincial Water Resources Department.

.

I. PENDAHULUAN

KINERJA merupakan suatu pengukuran terhadap hasil yang dicapai oleh pegawai terhadap pekerjaannya yang diharapkan berupa sesuatu yang optimal [1]. Kinerja pegawai sendiri mencakup prestasi seseorang dalam hal kualitas maupun kuantitas yang telah mereka capai ketika menjalankan tugas sesuai dengan tanggung jawab yang telah diberikan [2]. Prestasi kerja pegawai yang optimal tidak hanya berkontribusi dalam peningkatan produktivitas dan efisiensi, namun juga pada pencapaian tujuan strategis perusahaan. Oleh karena itu, evaluasi dan penilaian prestasi kerja pegawai merupakan hal yang penting untuk menilai sebaik apa pegawai telah mencapai target mereka.

Dinas Pekerjaan Umum Sumber Daya Air Provinsi Jawa Timur atau yang biasa disingkat Dinas PUSDA Jatim merupakan sebuah unsur yang melaksanakan otonomi daerah dalam menjalankan tugas pemerintahan yang menjadi wewenang pemerintah provinsi pada bidang Sumber Daya Air serta tugas pembantuan. Dinas yang dipimpin oleh Kepala Dinas ini berada di bawah naungan serta pertanggungjawaban Gubernur melalui Sekretaris Daerah [3]. Dinas PUSDA memiliki sekitar 128 pegawai dalam menjalankan tugas serta wewenangnya. Dalam menjalankan tugasnya, dinas PUSDA perlu melakukan pemantauan dan evalu asi kinerja pegawai untuk melihat apakah para pegawai yang bekerja telah melakukan tugas serta tanggung jawabnya denggan baik sesuai dengan aturan yang telah ditetapkan. Hal ini memiliki tujuan agar Dinas PUSDA Jatim dapat mengembangkan karir pegawai.

Setiap bulannya, instansi akan selalu melakukan penilaian serta evaluasi untuk para pegawai Dinas PUSDA Jatim untuk melihat bagaimana performa dari para pegawai yang dimilikinya. Hasil penilaian pegawai telah dikelompokkan dengan kuadran kinerja ASN, yaitu di bawah ekspektasi, sesuai ekspektasi, dan di atas ekspektasi. Sehingga, hasil akhir penilaian kinerja pegawai akan berada dalam kinerja yang sangat baik, baik, butuh perbaikan, kurang, atau sangat kurang. Meskipun, Dinas PUSDA telah melakukan penilaian dan evaluasi kinerja pegawai, akan tetapi instansi tersebut belum memiliki visualisasi pengelompokan kinerja pegawai untuk melihat kinerja pegawai secara komprehensif. Sehingga diperlukannya visualisasi untuk memantau kinerja pegawai agar dapat mengembangkan strategi pengelolaan SDM dan pemberian rekomendasi tunjangan kinerja pegawai di Dinas PU Sumber Daya Air Provinsi Jawa Timur.

Pengelompokan kinerja pegawai pada penelitian ini akan menerapkan data mining serta algoritma *K-Means Clustering* dalam pengolahan data-data pegawai yang dimaksud. Data pegawai akan diolah dan kemudian akan dikelompokan berdasarkan kelompok atau *cluster* masing-masing. *K-Means Clustering* digunakan dalam metode ini dikarenakan metode ini sesuai dengan karakteristik atribut penelitian dan datanya yang bersifat *unsupervised*. Algoritma *K-Means Clustering* sangat cocok digunakan untuk pengolahan dengan data yang besar dan digunakan untuk data numerik yang tidak memiliki label seperti data kinerja pegawai. Selain itu, terdapat penelitian yang membandingkan algoritma K-Means, K-Medoid, dan X-Means untuk penentuan data kinerja pegawai terbaik karena memiliki hasil evaluasi Davies Bouldin Index paling kecil diantara ketiganya [4]. Oleh karena itu, metode K-Means Clustering cocok untuk penelitian ini daripada metode lainnya.

Hasil dari penelitian ini adalah visualisasi *dashboard* dari hasil pengelompokan data kinerja pegawai untuk memantau kinerja pegawai agar dapat mengembangkan strategi pengelolaan SDM di Dinas PU Sumber Daya Air Provinsi Jawa Timur. Hasil K-Means Clustering pada akhirnya akann dibuat dan dikembangkan menjadi dashboard pengelompokan kinerja pegawai, hal ini dapat membantu instansi untuk membaca hasil dari pengelompokan kinerja pegawai dengan *K-Means Clustering*. Dengan *dashboard*, instansi dapat melihat persebaran data berdasarkan cluster masing-masing sehingga dari hal tersebut pihak instansi dapat merancang kebijakan tunjangan kinerja dari hasil evaluasi K-Means Clustering serta melakukan pengembangan strategi pengelolaan SDM instansi.

Terdapat banyak penelitian yang menggunakan *K-Means Clustering* sebagai algoritma untuk mengelompokkan suatu data. Beberapa di antaranya adalah penelitian dari S. Regina dan kawan-kawan yang berjudul ”Clustering Kualitas Kinerja Karyawan Pada Perusahaan Bahan Kimia Menggunakan Algoritma K-Means” [5]. Dalam penelitian tersebut dijelaskan bahwa para peneliti mengelompokkan hasil penilaian kinerja pegawai dengan metode *K-Means Clustering* berdasarkan variabel seperti kualitas kerja, tanggung jawab, kerjasama, kehadiran, dan disiplin. Penilaian kinerja itu nantinya akan dikelompokkan menjadi 3 cluster (sangat produktif, cukup produktif, dan kurang produktif). Dalam pengolahan K-Means nya, penelitian ini menggunakan perhitungan manual tanpa menggunakan alat untuk pengolahannya. Lalu, penelitian lainnya dari D. Aulia dan kawan-kawan yang berjudul ”Penerapan Algoritma K-Means dalam Proses Clustering Penilaian Kinerja Aparatur Sipil Negara di Sekretariat DPRD Pematangsiantar” [6]. Penelitian ini membahas mengenai pengelompokan

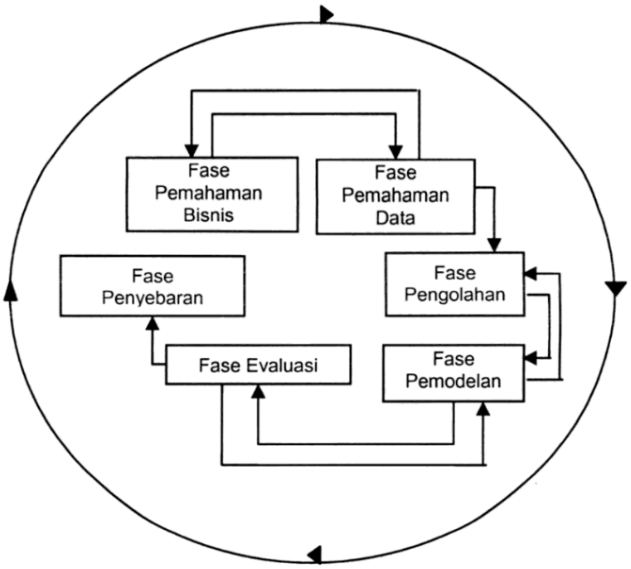
JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186

pegawai dengan algoritma K-Means Clustering berdasarkan nilai SKP (Sasaran Kinerja Pegawai), sehingga dapat mengetahui apakah kinerja para pegawai masuk ke dalam kinerja yang sangat baik, cukup, atau kurang. Dalam pengolahan *K-Means Clustering*-nya menggunakan tools *RapidMiner*.

Penelitian di atas hanya menunjukkan hasil berapa jumlah data yang masuk dalam masing-masing cluster. Seperti hasil dari penelitian pertama menunjukkan jika cluster 1 dengan kategori sangat produktif terdapat 16 data karyawan, cluster 2 dengan kategori cukup produktif terdapat 18 data karyawan, dan cluster 3 dengan kategori kurang produktif terdapat 4 data karyawan. Penelitian kedua menunjukkan hasil 3 cluster, yaitu yang pertama cluster tinggi memiliki satu data, cluster sedang memiliki 9 data, dan cluster rendah memiliki 13 data. Kedua penelitian tersebut hanya menampilkan hasil cluster saja, sedangkan penelitian ini akan menunjukkan bagaimana persebaran tiap variabel per cluster-nya dan menampilkan *dashboard* yang juga memiliki visualisasi persebaran datanya. Hal ini dapat memperluas wawasan instansi serta pengguna untuk lebih memahami mengenai pengelompokan data kinerja pegawai. Sehingga, dengan adanya visualisasi *dashboard* akan memudahkan pihak instansi untuk memantau hasil pengelompokan pegawai yang bekerja pada instansi tersebut.

II. METODE PENELITIAN

Data mining merupakan proses pengumpulan beberapa informasi dari suatu data yang besar untuk kemudian diolah sehingga menghasilkan suatu informasi penting seperti menemukan suatu pola dari sekumpulan data, memprediksi suatu hal dari sekumpulan data, serta *knowledge extraction* [7], [8]. CRISP-DM (*Cross-Industry Standard Process for Data Mining*) merupakan sebuah standar dari proses data mining untuk strategi dalam memecahkan masalah secara umum dari unit penelitian atau bisnis. Menurut Larose, CRISP-DM memiliki enam tahapan di dalamnya, yaitu [7]:



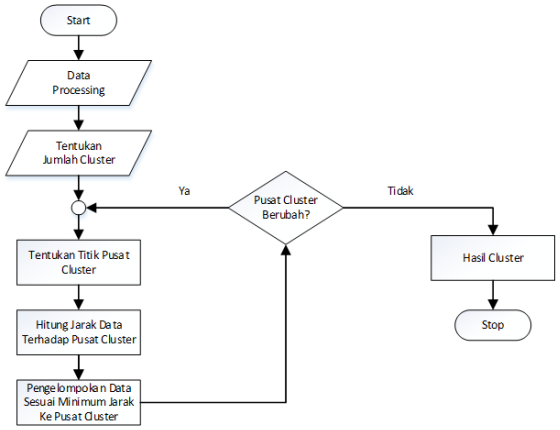
Gambar 1. Tahapan dalam CRISP-DM

1. Tahapan *Business Understanding* (Pemahaman Bisnis)
   Tahapan ini merupakan tahapan untuk menentukan tujuan proyek dan kebutuhan secara rinci apa yang dibutuhkan dalam lingkup bisnis atau lingkup penelitian secara menyeluruh.
2. Tahapan *Data Understanding* (Pemahaman Data)
   Tahapan ini merupakan tahapan pengumpulan data dan menggunakan analisis identifikasi data untuk memahami lebih dalam data serta pencarian pengetahuan awal.
3. Tahapan *Data Preparation* (Pengolahan Data)
   Tahapan ini merupakan tahapan dilakukan pemilihan kasus dan variabel yang ingin diteliti lebih lanjut yang sesuai dengan analisis yang akan dilakukan. Kemudian melakukan perubahan pada beberapa variabel apabila diperlukan dan menyiapkan data awal sampai siap untuk pemodelan.
4. Tahapan *Modeling* (Pemodelan)
   Tahapan ini dilakukan pengaplikasian teknik pemodelan yang sesuai serta dilakukan kalibrasi aturan model untuk mengoptimalkan hasil. Hal ini perlu memperhatikan beberapa teknik mungkin digunakan untuk permasalahan data mining yang sama.

   K-Means Clustering merupakan salah satu algoritma pengelompokan non hirarki yang mempartisi objek ke dalam satu cluster atau lebih berdasarkan karakteristik kemiripannya [8]. Pengelompokan kinerja pegawai

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186

pada penelitian ini akan menerapkan data mining serta algoritma *K-Means Clustering* dalam pengolahan data-data pegawai yang dimaksud. Data pegawai akan diolah dan kemudian akan dikelompokan berdasarkan kelompok atau cluster masing-masing. *K-Means Clustering* digunakan dalam metode ini dikarenakan metode ini sesuai dengan karakteristik atribut penelitian dan datanya yang bersifat *unsupervised*.



Gambar 2. Tahapan Algoritma K-Means Clustering

Beberapa tahapan dalam metode *K-Means Clustering* yaitu sebagai berikut [8], [9]:

a) Menentukan jumlah k (*cluster*) yang diinginkan untuk menjadi acuan dalam mengelompokkan data.
b) Menentukan titik pusat (*centroid*) dari setiap cluster yang terbentuk.
c) Menghitung jarak setiap data ke pusat *centroid* terdekat, dengan menggunakan persamaan jarak seperti *Euclidean Distance*. Berdasarkan beberapa penelitian seperti pada penelitian Nishom dan juga dari Pangestu dkk, metode *Euclidean Distance* merupakan metode perhitungan distance/jarak yang paling akurat diantara distance lainnya dalam pengklusteran [10], [11].
Euclidean Distance

$$d(x_1, x_2) = ||x_1 - x_2|| = \sqrt{\sum_{i=1}^{n}(x_{1i} - x_{2i})^2} \quad (1)$$

dimana:
d = jarak antara data dengan centroid
x_1 = data latih
x_2= data uji
i = data ke-i
n = jumlah data
d) Menentukan nilai centroid baru di setiap cluster dengan menghitung rata-rata (means) dari data tiap cluster menggunakan rumus persamaan (2).

$$Cluster\ center = \sum \frac{a_i}{n} \quad (2)$$

e) Mengulangi langkah nomor 3 sampai nomor 4 hingga tidak ada data yang berpindah *cluster*. Iterasi akan terus dilakukan hingga perubahan posisi centroid antara dua iterasi terakhir tidak signifikan. Jika perubahan lebih kecil dari ambang batas yang telah ditetapkan sebelumnya (toleransi), maka iterasi akan berhenti. Iterasi juga dapat berhenti ketika jumlah iterasi maksimum yang sebelumnya ditetapkan telah tercapai. Bahkan apabila centroid belum sepenuhnya konvergen. Setelah salah satu antara kedua kemungkinan bagaimana iterasi berhenti tercapai, centroid tidak berubah lagi secara signifikan, dan titik data telah diberi label cluster akhir. Hasil akhir dari pengolahan K-Means adalah pembagian data ke dalam cluster (k), dengan setiap cluster direpresentasikan oleh centroid terakhir yang dihitung.

5. Tahapan *Evaluation* (Evaluasi)
Tahapan ini digunakan mengevaluasi satu model atau lebih yang digunakan pada tahap pemodelan guna mendapatkan kualitas dan efektivitas sebelum disebarkan.
6. Tahapan *Deployment* (Penyebaran)
Tahapan ini merupakan tahap penyebaran dengan menggunakan model yang dihasilkan. Model yang terbentuk bukan menjadi tanda bahwa proyek telah selesai.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186

III. HASIL DAN PEMBAHASAN

A. *Business Understanding*

Business Understanding atau pemahaman bisnis adalah proses untuk mengidentifikasi tujuan dari dilakukannya penelitian ini dimana pengklusteran data evaluasi dan penilaian kinerja pegawai Dinas PU Sumber Daya Air Provinsi Jawa Timur, serta memahami kebutuhan dan permasalahan yang terjadi pada instansi terkait yaitu Dinas PU Sumber Daya Air Provinsi Jawa Timur. Tujuan bisnis dari penelitian ini adalah instansi terkait telah memiliki sistem evaluasi dan penilaian kinerja pegawai yang setiap bulan diadakan serta telah dikelompokan menggunakan kuadran ASN. Akan tetapi, instansi ini belum memiliki *sistem monitoring* yang dapat membantu untuk mengevaluasi dan memonitoring kinerja pegawai secara komprehensif. Adapun, tujuan dari penelitian ini adalah membuat *sistem monitoring* berupa *dashboard* untuk memvisualisasikan hasil dari evaluasi dan penilaian kinerja pegawai di Dinas PU Sumber Daya Air Provinsi Jawa Timur dengan menggunakan metode *K-Means Clustering*.

Berdasarkan wawancara dari pihak SDM Dinas PU Sumber Daya Air Provinsi Jawa Timur, didapatkan bahwa terdapat beberapa evaluasi dan penilaian kinerja pegawai yang telah dilakukan oleh instansi yaitu penilaian kepada pegawai. Dimana penilaian kinerja ini biasanya dilakukan setiap satu bulan sekali untuk menilai bagaimana kinerja para pegawai yang ada pada instansi tersebut. Akan tetapi, instansi belum memiliki visualisasi pengelompokan kinerja pegawai yang dapat digunakan untuk memantau kinerja pegawai yang bekerja disana. Oleh karena itu, pada penelitian ini akan dibuatlah visualisasi *dashboard* pengelompokan hasil kinerja pegawai Dinas PU Sumber Daya Air Provinsi Jawa Timur.

B. *Data Understanding*

Data Understanding atau pemahaman data adalah proses untuk memahami lebih lanjut data evaluasi dan penilaian kinerja pegawai yang akan diolah dan dianalisis. Data yang digunakan dalam penelitian ini yaitu data Hasil Kinerja pegawai Dinas PU Sumber Daya Air Provinsi Jawa Timur tahun 2023. Adapun data hasil kinerja pegawai yang digunakan yaitu seluruh pegawai ASN yang bekerja di Dinas PU Sumber Daya Air Provinsi Jawa Timur. Data Hasil Kinerja pegawai tersebut memiliki beberapa aspek penilaian yaitu aspek berorientasi pelayanan, akuntabel, kompeten, harmonis, loyal, adaptif, kolaboratif, dan hasil kerja. Data pegawai yang didapat akan diolah dan dikelompokkan berdasarkan delapan variabel tersebut.

C. *Data Preparation*

Data Preparation atau pengolahan data adalah proses penyiapan data yang digunakan dalam penelitian ini. Dimana data yang akan digunakan perlu disiapkan untuk diidentifikasi dan ditangani untuk kemudian diolah. Pada tahap ini, perlu dilakukan *preprocessing* data dengan penanganan *missing value* dan data duplikat, serta deteksi data *outlier*.

```
Jumlah data sebelum pembersihan:  128
Jumlah data setelah pembersihan:  128
Jumlah data yang hilang:  0
Jumlah data yang duplikat:  0
Jumlah data yang ditemukan:  0
```

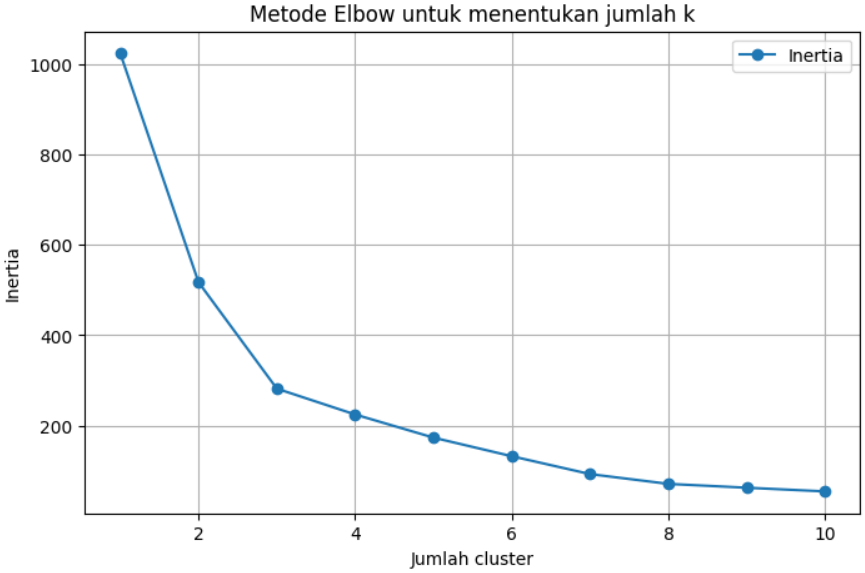
Gambar 3. Hasil Cleaning Data

Hasil *cleaning data* di atas menunjukkan jumlah data yang hilang adalah 0 dari 128 data dan jumlah data yang duplikat sebanyak 0 dari 128. Sehingga, didapatkan bahwa tidak terdapat data yang hilang (*missing value*) dan data duplikat/ganda.

D. *Modeling*

Dalam proses pemodelannya, algoritma *K-Means Clustering* perlu menentukan k (*cluster*) optimal supaya data yang akan dikelompokkan dapat dikelompokkan dengan optimal. Oleh karena itu, untuk penentuan k optimal, algoritma *K-Means Clustering* perlu menentukannya dengan menggunakan metode *Elbow*. Di bawah ini merupakan hasil penentuan cluster optimal pada *K-Means Clustering* dengan metode *Elbow*.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186



Gambar 4. Grafik Metode Elbow

TABEL I
SELISIH INERSIA

Cluster	Hasil Inersia	Selisih Inersia
1	1024	0
2	517.17	506.83
3	281.68	235.49
4	224.47	57.22
5	176.24	51.18
6	131.41	41.17
7	92.55	39.58
8	70.88	21.67
9	60.99	8.47
10	52.52	8.20

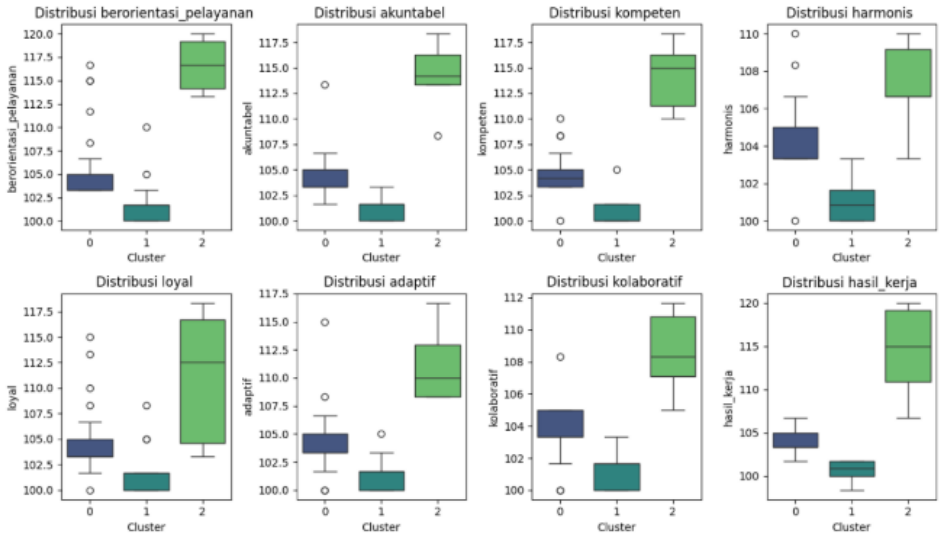
Hasil cluster di atas menunjukkan jika k optimal yang dihasilkan oleh metode *elbow* adalah 3 cluster. Dapat dilihat dari grafik *elbow* diatas titik penurunan inersia mulai melambat pada cluster ke 3, dimana setelah cluster 3 titik penurunan inersia mulai turun dengan konstan dan mulai membentuk sudut siku-siku. Hal ini ditunjukkan oleh hasil selisih inersia antar cluster, selisih inersia cluster 2 dan 3 sangat drastis yaitu 509.83 dan 235.49. Kemudian mulai turun bertahap dan dapat dikatakan tidak signifikan setelah cluster ke 3 yaitu mulai cluster 4 selisih inersia tidak memiliki perbedaan yg signifikan seperti cluster sebelumnya yaitu dengan selisih inersia 57.22, 51.18, 41.17, dst. Sehingga, pada metode ini cluster optimal untuk penelitian ini adalah 3 cluster.

Pada tahap *modeling* ini, dilakukan pengolahan model K-Means Clustering dengan 8 variabel, yaitu berorientasi pelayanan, akuntabel, kompeten, harmonis, loyal, adaptif, kolaboratif, serta hasil kerja. Data penilaian dari delapan variabel tersebut akan diolah serta dikelompokkan menjadi 3 cluster. Sehingga didapatkan hasil seperti pada gambar di bawah.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186

Gambar 5. Pairplot Hasil K-Means Clustering

Gambar 6 merupakan hasil pengolahan dari K-Means Clustering dengan menggunakan 8 variabel sehingga didapatkan visualisasi dengan menggunakan pairplot seperti yang terlihat pada gambar diatas. Pada pairplot di atas dapat dilihat bahwa data-data telah dikelompokkan ke dalam 3 cluster berdasarkan karakteristik masing-masing data. Pairplot diatas menampailkan kepadatan dan persebaran data pada setiap cluster. Persebaran tersebut dapat dilihat dari perbedaan warna antar cluster. Seperti cluster 0 yang ditunjukkan oleh warna ungu, cluster 1 ditunjukkan oleh warna hijau, dan cluster 2 yang ditunjukkan oleh warna kuning. Cluster 0 memiliki persebaran data berada di tengah-tengah cluster 1 dan 2, menunjukkan cluster ini memiliki nilai penilaian yang moderat. Cluster 1 berada di bagian bawah kiri yang ditunjukkan oleh warna hijau, menunjukkan bahwa penilaian kinerja pegawai dalam cluster ini lebih rendah dibandingkan dengan cluster lain. Cluster 2 berada di bagian paling atas sebelah kanan yang menunjukkan cluster ini memiliki nilai penilaian yang lebih tinggi dibandingkan cluster lain. Distribusi atau persebaran data mempengaruhi bagaimana hasil cluster yang didapat nantinya.



Gambar 6. Box Plot Persebaran Data Setiap Cluster

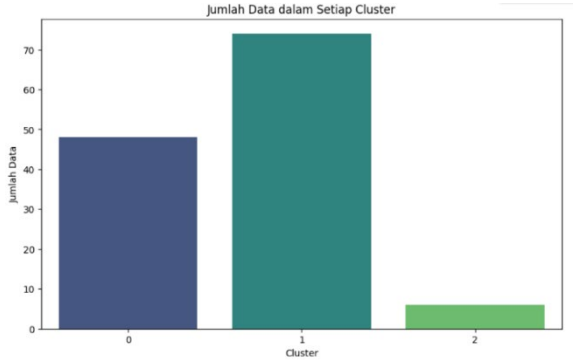
Dari hasil persebaran cluster per variabel di atas menunjukkan bagaimana data terdistribusi dalam setiap cluster, termasuk rentang penilaian dan outlier. Cluster yang memiliki rentang penilaian lebih sempit menunjukkan konsistensi kinerja tinggi, dan rentang penilaian yang lebih lebar menunjukkan variasi kinerja lebih besar. Outlier

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186

dapat mempengaruhi hasil dari *K-Means Clustering*, namun hasil outlier yang didapatkan tetap digunakan karena pada penelitian ini data yang digunakan adalah data penilaian sehingga semua data akan diclusterkan. Hasil diatas juga menunjukkan bahwa cluster 0 memiliki ciri-ciri data penilaian dari rentang 103-105 dengan beberapa data outlier diluar nilai 103-105, cluster 1 memiliki ciri-ciri data penilaian dari rentang 100-102, dan cluster 2 memiliki ciri-ciri data penilaian dari rentang 105-120. Sehingga, dapat disimpulkan bahwa cluster 0 memiliki ciri-ciri data menengah, cluster 1 memiliki ciri-ciri data paling bawah, dan cluster 2 memiliki ciri-ciri data paling atas. Sehingga dapat disimpulkan sebagai berikut.

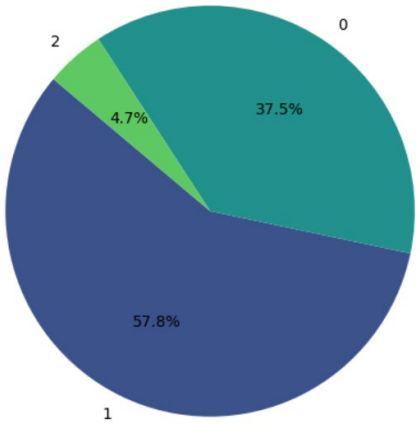
Tabel II Keterangan Cluster

Cluster	Keterangan
0	Baik
1	Butuh Perbaikan
2	Sangat Baik



Gambar 7. Histogram Jumlah Data Setiap Cluster

Jumlah data yang dihasilkan dari kode diatas adalah sebanyak 48 data masuk ke dalam cluster 0, sebanyak 74 data masuk ke dalam cluster 1, dan sebanyak 6 data masuk ke dalam cluster 2. Sehingga didapatkan persentase jumlah data pegawai berdasarkan cluster adalah sebagai berikut. Pada hasil cluster tersebut didapatkan bahwa persentase cluster 0 yaitu 37.5%, cluster 1 yaitu 57.813%, dan cluster 2 yaitu 4.687%.



Gambar 8. Persentase Jumlah Cluster

Persentase jumlah cluster di atas dapat membantu pihak instansi untuk memahami bagaimana kinerja pegawai terdistribusi. Pegawai dengan cluster 2 yaitu kinerja paling tinggi memiliki performa yang lebih baik pada tiap variabel seperti berorientasi pelayanan, akuntabel, kompetensi, adaptasi, dan variabel lainnya dibandingkan dengan pegawai di cluster lain. Sehingga analisis ini menjadi dapat menjadi dasar instansi untuk mengatur kebijakan tunjangan kinerja, sehingga pegawai yang memiliki kinerja paling tinggi mendapatkan tunjangan yang tepat untuk dirinya. Pegawai dengan cluster 1 yaitu kinerja paling rendah memiliki performa yang cukup baik namun dapat ditingkatkan untuk lebih baik lagi. Apalagi melihat persentase jumlah cluster 1 memiliki jumlah paling besar, jadi

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186

instansi dapat mengatur dan merancang bagaimana pengelolaan pengembangan pegawai agar dapat lebih baik lagi.

E. Evaluation

Pada tahap evaluasi, evaluasi model dilakukan dengan menggunakan metode *Silhouette Coefficient*. Dimana apabila nilai yang dihasilkan semakin mendekati nilai 1 maka *cluster* yang dihasilkan semakin bagus. Menurut Monshizadeh, hasil *Silhouette Score* yang memiliki nilai lebih dari 0.7 mengindikasikan bahwa cluster tersebut adalah cluster yang sangat baik (kuat), nilai lebih dari 0.5 mengindikasikan cluster tersebut cukup baik (cukup), serta nilai lebih dari 0.25 mengindikasikan cluster tersebut adalah cluster yang lemah [12].

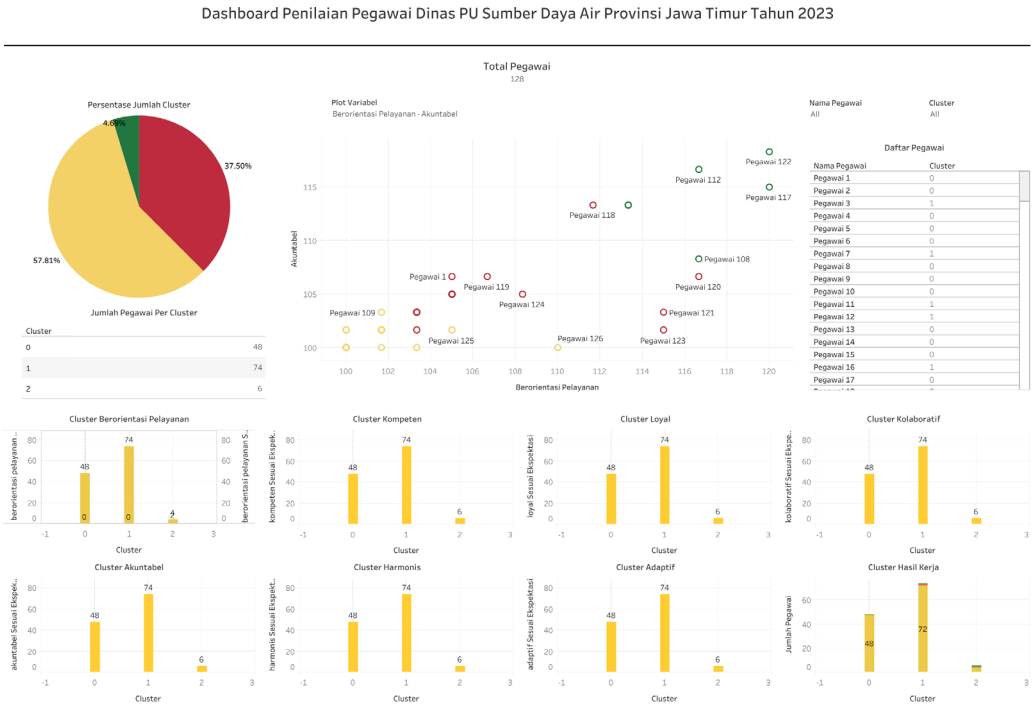
```
Nilai Silhouette Score:
0.5932286324588552
```

Gambar 9. Nilai *Silhouette Coefficient*

Nilai *Silhouette Coefficient* yang dihasilkan adalah 0.593, dimana hal tersebut menunjukkan akurasi dan kualitas cluster cukup baik karena angka tersebut mendekati nilai 1 [13] dan juga memiliki nilai lebih dari 0.5 yang mana mengindikasikan bahwa 3 cluster adalah cluster yang cukup baik. Hasil *silhouette coefficient* ini memiliki keterkaitan dengan outlier yang tidak dihapus karena algoritma *K-Means* sangat sensitif terhadap outlier dan outlier dapat mempengaruhi bagaimana hasil cluster karena keberadaan dirinya yang ikut diolah (karena distribusi data outlier yang tersebar).

F. Deployment

Tahap selanjutnya yaitu, tahap deployment, tahap penyebaran menggunakan model yang dihasilkan. *Deployment* yang dilakukan pada penelitian ini yaitu pembuatan *dashboard*. Pembuatan *dashboard* ini dilakukan untuk menunjukkan hasil dari model yang telah dibuat dan pengguna dapat melihat secara keseluruhan karakteristik dari setiap cluster [13]. Dashboard ini diharapkan dapat menjelaskan secara keseluruhan hasil apa yang didapat dari pengelompokan penilaian pegawai Dinas PUSDA. Di bawah ini merupakan hasil *dashboard* yang telah dilakukan oleh peneliti.

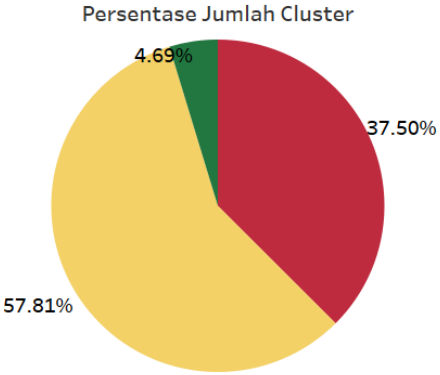


Gambar 10. Dashboard Penilaian Pegawai Dinas PUSDA Provinsi Jatim Tahun 2023

Dashboard yang ditampilkan memiliki beberapa diagram dan plot di dalamnya, berikut adalah penjelasan mengenai masing-masing diagram yang ditampilkan:

1. **Persentase jumlah cluster**

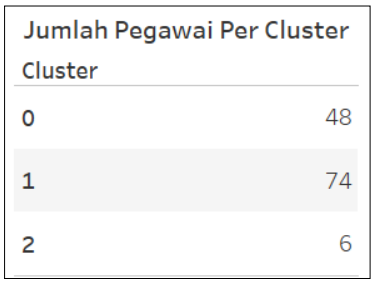
JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186



Gambar 11. Dashboard Persentase Jumlah Cluster

Pie Chart yang ditampilkan merupakan diagram untuk menunjukkan persentase jumlah pegawai per cluster. Diagram lingkaran yang ditampilkan menunjukkan persentase cluster 0 sebesar 37.50%, cluster 1 sebesar 57.81%, dan cluster 2 sebesar 4.69% dengan tujuan supaya instansi dapat melihat jumlah persentase masing-masing cluster yang didapatkan.

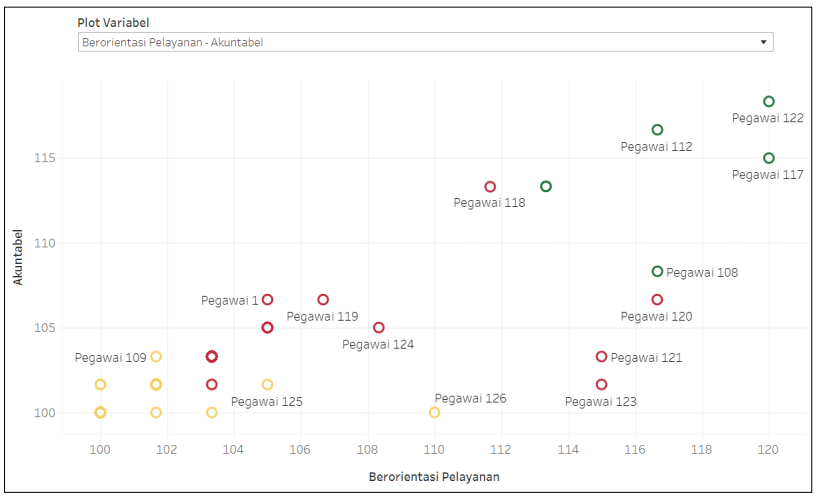
2. **Jumlah pegawai per cluster**



Gambar 12. Dashboard Jumlah Pegawai Per Cluster

Diagram yang ditampilkan diatas merupakan diagram untuk menunjukkan jumlah pegawai dari masing-masing cluster, cluster 0 sebanyak 48 pegawai, cluster 1 sebanyak 74 Pegawai, Cluster 2 sebanyak 6 pegawai. Sehingga jumlah pegawai dari masing-masing cluster dapat dilihat langsung oleh pihak instansi.

3. **Plot cluster 2 variabel**

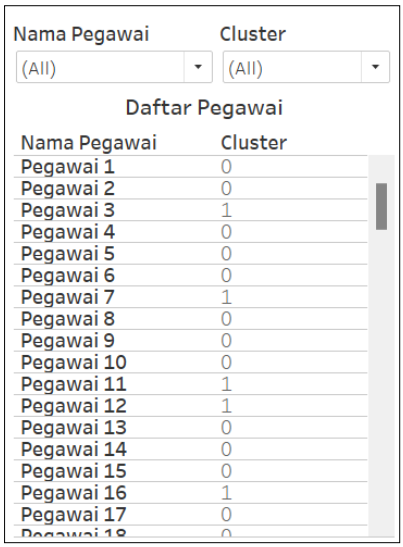


Gambar 13. Plot Cluster 2 Variabel

Plot diatas menunjukkan hasil cluster 2 variabel yang digunakan dalam penelitian. **Gambar 14** menunjukkan hasil cluster variabel berorientasi pelayanan dan akuntabel, akan tetapi hasil diatas tidak hanya akan menunjukkan kedua variabel itu saja namun juga variabel-variabel lain yang digunakan dalam penelitian yang divisualisasikan dengan sumbu x dan y. Oleh karena itu, hasil yang ditampilkan adalah persebaran cluster dua variabel. Di atas plot cluster, terdapat filter yang dapat memilih variabel mana yang akan ditampilkan. Sehingga, pihak instansi dapat melihat persebaran cluster tiap variabel.

JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)
Journal homepage: https://jurnal.stkippgritulungagung.ac.id/index.php/jipi
ISSN: 2540-8984
Vol. 10, No. 3, September 2025, Pp. 2175-2186

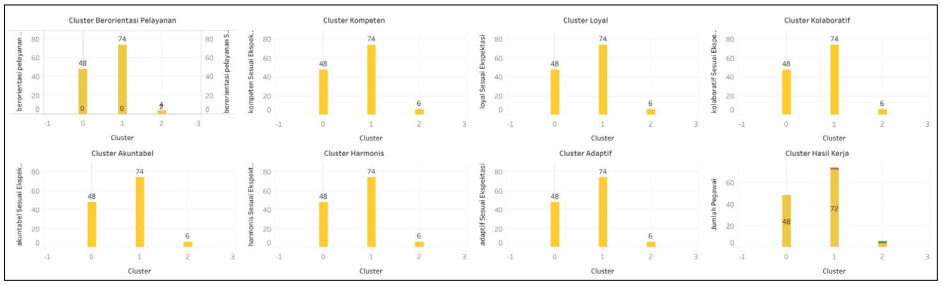
4. **Daftar pegawai**



Gambar 14. Dashboard Daftar Pegawai

Tabel yang ditampilkan seperti pada **Gambar 15** adalah tampilan untuk menunjukkan setiap pegawai berdasarkan cluster yang telah didapatkan. Terdapat fitur filter untuk memilih nama pegawai yang dicari dan cluster yang ingin ditampilkan.

5. **Persebaran cluster per variabel**



Gambar 15. Dashboard Histogram Peresebaran Cluster per Variabel

Histogram persebaran cluster tiap variabel ditampilkan untuk dapat menunjukkan persebaran data dalam cluster pada setiap variabel yang ada. Dengan diagram batang seperti pada **Gambar 16** dapat menunjukkan analisis jumlah tiap cluster serta variabelnya.

Perbedaan penelitian ini dari penelitian sebelum-sebelumnya yaitu dari S. Regina [5] dan D. Aulia [6] adalah pada tahap persebaran (*deployment*) dan juga pada tools yang digunakan untuk memodelkan *K-Means Clustering*, yaitu menggunakan bahasa pemrograman Python. Pada penelitian sebelumnya mereka menggunakan tools *RapidMiner* dan juga perhitungan secara manual, akan tetapi pada penelitian ini tools yang digunakan adalah bahasa pemrograman Python dengan Google Colab sebagai platform untuk menjalankan programnya. Pemrograman Python jelas lebih unggul daripada perhitungan manual terutama pada perhitungannya karena jika manual bisa terjadi *human error* atau salah perhitungan, sedangkan Python mengurangi *human error* karena sistem tersebut yang mengolahnya. Python juga dapat melakukan perhitungan data yang banyak dengan lebih cepat dibandingkan perhitungan manual. *RapidMiner* sendiri memiliki antarmuka yang intuitif dan mudah digunakan namun Python lebih fleksibel dengan pengkustoman kode-kode tertentu, juga dapat memodifikasi algoritma yang ada. Python juga memiliki library yang luas sehingga memungkinkan untuk integrasi dengan alat dan teknik lain. Oleh karena itu, Python lebih unggul dibandingkan kedua metode lainnya.

IV. KESIMPULAN

Berdasarkan hasil di atas, dapat disimpulkan bahwa penilaian pegawai Dinas PU Sumber Daya Air Provinsi Jawa Timur Tahun 2023 dikelompokkan menggunakan algoritma *K-Means Clustering* dan menghasilkan 3 cluster berdasarkan metode *elbow*. Dengan 48 pegawai masuk ke dalam cluster 0, dimana dalam analisis diatas cluster 0 termasuk ke dalam cluster yang memiliki data menengah yaitu 103-105. Kemudian, 74 pegawai masuk ke dalam

cluster 1, dengan analisis cluster tersebut masuk ke dalam cluster yang memiliki data paling bawah. Serta, 6 pegawai masuk ke dalam cluster 2, dengan analisis cluster yang memiliki data paling atas. Hasil pengujian evaluasi menggunakan *Silhouette Coefficient* menunjukkan angka 0.593, dimana angka tersebut sudah bisa dikatakan sebagai model yang cukup baik. Hasil *Clustering* penilaian pegawai dibuat dalam visualisasi *dashboard* yang telah dibuat menggunakan tools Tableau berisi daftar nama pegawai yang telah dikelompokkan berdasarkan masing-masing cluster, serta terdapat persebaran data tiap variabel dan total jumlah pegawai berdasarkan cluster.

DAFTAR PUSTAKA

[1] Stephen P., *Perilaku Organisasi*. Jakarta: PT. Indeks Kelompok Gramedia, 2006.
[2] Anwar Prabu Mangkunegara, *Evaluasi Kinerja Sumber Daya Manusia*. Bandung: Refika Aditama, 2005.
[3] PU SDA Provinsi Jawa Timur, "Selamat Datang."
[4] G. B. Kaligis and S. Yulianto, "ANALISA PERBANDINGAN ALGORITMA K-MEANS, K-MEDOIDS, DAN X-MEANS UNTUK PENGELOMPOKKAN KINERJA PEGAWAI (Studi Kasus: Sekretariat DPRD Provinsi Sulawesi Utara)," *IT-EXPLORE*, vol. 1, no. 3, pp. 179–193, 2022.
[5] 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.
[6] Della Aulia, M Safii, and Dedi Suhendro, "Penerapan Algoritma K-Means dalam Proses Clustering Penilaian Kinerja ASN," *Jurnal Riset Sistem Informasi dan Teknik Informatika (JURASIK)*, vol. 6, no. 1, pp. 47–60, Feb. 2021.
[7] Kusrini and Emha Taufiq Luthfi, *Algoritma Data Mining*, 1st ed. Yogyakarta: ANDI, 2009.
[8] Agus Nur Khomarudin, "Teknik Data Mining : Algoritma K-Means Clustering," *IlmuKomputer.Com*, 2016.
[9] S. Hidayati, A. T. Darmaliana, and R. Riski, "Comparison of K-Means, Fuzzy C-Means, Fuzzy Gustafson Kessel, and DBSCAN for Village Grouping in Surabaya Based on Poverty Indicators," *Jurnal Pendidikan Matematika (Kudus)*, vol. 5, no. 2, p. 185, Dec. 2022, doi: 10.21043/jpmk.v5i2.16552.
[10] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 4, no. 1, pp. 20–24, Jan. 2019, doi: 10.30591/jpit.v4i1.1253.
[11] M. S. Pangestu and M. A. Fitriani, "Perbandingan Perhitungan Jarak Euclidean Distance, Manhattan Distance, dan Cosine Similarity dalam Pengelompokan Data Bibit Padi Menggunakan Algoritma K-Means," *Sainteks*, vol. 19, no. 2, p. 141, Oct. 2022, doi: 10.30595/sainteks.v19i2.14495.
[12] M. Monshizadeh, V. Khatri, R. Kantola, and Z. Yan, "A deep density based and self-determining clustering approach to label unknown traffic," *Journal of Network and Computer Applications*, vol. 207, p. 103513, Nov. 2022, doi: 10.1016/j.jnca.2022.103513.
[13] Dyah Putri Rahmawati, Sri Hidayati, Paramaditya Arismawati, and Ahmad Wali Satria Bahari Johan, "Prospective New College Student Dashboard: Insights from K-Means Clustering with Principal Component Analysis," *Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 9, no. 2, pp. 137–144, Jul. 2024.
[14] S. Sahibu, R. Bambang, and I. Taufik, "Penerapan Data Mining Dalam Analisis Penilaian Kinerja Pegawai Menerapkan Metode K-Means," *Jurnal Media Informatika Budidarma*, 2023.
[15] Sonia Sundari, Relita Buaton, and Rusmin Saragih, "Clustering Kepuasan Layanan Pengguna Bus Trans Binjai Dengan Metode Cluster Data Mining Studi Kasus Dinas Perhubungan Kota Binjai," *Seminar Nasional Informatika (SENATIKA)*, 2021.
[16] A. Septiarini, I. A. Thaher, and N. Puspitasari, "Pengelompokan Kualitas Kinerja Pegawai Menggunakan Metode K-Means Clustering," *Komputika : Jurnal Sistem Komputer*, vol. 11, no. 2, pp. 131–141, Jul. 2022, doi: 10.34010/komputika.v11i2.5518.
[17] EKO PRASETYO, *DATA MINING : KONSEP DAN APLIKASI MENGGUNAKAN MATLAB*, 2nd ed. Yogyakarta: ANDI, 2012.
[18] D. Alfiandi, E. Ernawati, and E. P. Purwandari, "Implementasi K-Means Clustering dan Pemetaan Pemukiman Kumuh di Kota Bengkulu Berbasis Web," *Rekursif J. Informatika*, vol. 6, no. 2, Nov. 2018.
[19] A. T. R. Saragih, A. S. Sembiring, and M. Sayuthi, "Penerapan Metode Clustering KMeans untuk Proses Seleksi Calon Peserta Lomba MTQ," *Pelita Inform*, vol. 17, pp. 117–122, 2018.
[20] Solmin Paembonan, Hisma Abduh, and K. Kunci, "Penerapan Metode Silhouette Coeficient Untuk Evaluasi Clutering Obat Clustering; K-means; Silhouette coeficient," *Pena Teknik*, vol. 6, no. 2, pp. 48–54, Sep. 2021.
[21] Annisa Jamilatul, "5 Cara MendetekApa Itu Outlier dan Bagaimana Cara Mendeteksinyasi Outlier dalam Data," Pacmann.
[22] Bunga Dea Laraswati, "5 Cara Mendeteksi Outlier dalam Data," algoritma.
[23] A. Chusyairi and P. Ramadar Noor Saputra, "Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering," *Telematika*, vol. 12, no. 2, pp. 139–148, Aug. 2019, doi: 10.35671/telematika.v12i2.848.
[24] Andre Oliver, "Mengenal Google Colab: Mulai dari Definisi, Cara Menggunakan, hingga Manfaatnya," glints.