ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

KLASTERISASI PEMETAAN KEDISIPLINAN PEGAWAI BERDASARKAN REKAP KEHADIRAN MENGGUNAKAN ALGORITMA CLUSTERING K-MEANS

Imam Ahmad Ashari✉, Purwono, Jatmiko Indriyanto, Arif Setia Sandi
Fakultas Sains dan Teknologi, Universitas Harapan Bangsa, Purwokerto, Indonesia
Email: imamahmadashari@uhb.ac.id

DOI: https://doi.org/10.46880/jmika.Vol9No1.pp12-18

ABSTRACT

Employee discipline is one of the key success factors in a company. Work discipline has an important role in the formation of a positive work environment. One of the things that shows employee discipline is the time of attendance. Attendance time is usually recorded at the time the employee enters and leaves. Disciplinary information can be mapped into several groupings so that it is easy for decision makers to read. One of the computational methods that can perform data mapping is the K-Means Clustering method. The K-Means Clustering method can group data based on their characteristics. In this study, attendance data were analyzed using the K-Means method to obtain disciplinary groupings. The number of Clusters is calculated using the elbow method, 3 Clusters are obtained which are the best Cluster choices, namely Clusters 0, 1, and 2. The data analysis process shows Cluster 2 is the Cluster with the best level of discipline. From the analysis, it shows that the K-Means Clustering method can classify data based on employee discipline. Based on these results, decision makers can be helped in assessing employee discipline at Universita Harapan Bangsa using the disciplinary data grouping that has been made.

Keyword: Attendance Recap, Discipline, Data Mining, Clustering, K-Means Method.

ABSTRAK

Kedisiplinan pegawai merupakan salah satu faktor kunci keberhasilan di suatu perusahaan. Kedisiplinan kerja mempunyai peranan penting dalam terbentuknya lingkungan kerja yang positif. Salah satu hal yang menunjukkan sikap disiplin pegawai adalah waktu kehadiran. Waktu kehadiran biasanya dicatat pada waktu masuk dan pulang pegawai. Informasi kedisiplinan dapat dipetakkan ke dalam beberapa pengelompokkan sehingga mudah dibaca oleh pengambil keputusan. Salah satu metode komputasi yang dapat melakukan pemetaan data adalah metode K-Means Clustering. Metode K-Means Clustering dapat mengelompokkan data berdasarkan karakteristiknya. Pada penelitian ini data kehadiran dianalisis menggunakan metode K-Means untuk mendapatkan pengelompokkan kedisiplinan. Jumlah Cluster dihitung menggunakan metode elbow, didapatkan 3 Cluster yang menjadi pilihan Cluster terbaik yaitu Cluster 0, 1, dan 2. Proses analisis data menunjukkan Cluster 2 adalah Cluster dengan tingkat kedisiplinan terbaik. Dari analisis yang dilakukan menunjukkan bahwa metode K-Means Clustering dapat mengelompokkan data berdasarkan kedisiplinan pegawai. Berdasarkan hasil ini pengambil keputusan dapat terbantu dalam menilai kedisiplinan pegawai di Universita Harapan Bangsa menggunakan pengelompokkan data kedisiplinan yang telah dibuat.

Kata Kunci: Rekap Kehadiran, Kedisiplinan, Data Mining, Clustering, Metode K-Means.

PENDAHULUAN

Rekap kehadiran pegawai merupakan laporan wajib yang dibutuhkan oleh setiap instansi atau perusahaan. Manfaat dari rekap kehadiran salah satunya adalah sebagai acuan untuk penggajian dan monitoring kinerja pegawai. Pencatatan kehadiran yang baik dan *real-time* akan mempermudah instansi atau perusahaan dalam mengevaluasi kinerja dari seorang pegawai. Pegawai yang baik adalah pegawai yang hadir dan pulang tepat waktu sesuai dengan SOP (*Standart operating procedur*) yang telah ditetapkan oleh perusahaan. Kehadiran pegawai juga dapat memberikan kontribusi kemajuan kepada perusahaan untuk masa-masa yang akan datang (H Kara, 2014). Salah satu apresiasi yang diberikan untuk pegawai salah satunya adalah bonus kedisiplinan.

Disiplin adalah rasa taat dan patuh terhadap nilai yang dipercaya serta tanggung jawab yang diberikan kepada seseorang. Disipilinisasi di perusahaan berarti bahwa pegawai patuh terhadap peraturan atau tunduk pada pengawasan dan pengendalian sesuai dengan SOP yang berlaku. Perusahaan melakukan tindakan

kedisplinan dengan tujuan untuk memperkuat aturan yang telah ditetapkan, meningkatkan moral, dan mempertahankan standar perusahaan. Disiplin kerja juga merupakan salah satu karakteriristik dari sumber daya manusia yang berkualitas (Saputra, 2019). Selain itu disiplin kerja di dalam perusahaan mampu berpengaruh positif pada lingkungan kerja terhadap kinerja pegawai (Putri et al., 2019). Kebermanfaatan yang disebutkan di atas menandakan bahwa begitu pentingnya perusahaan harus selalu menjaga karakter disiplin oleh setiap pegawainya, maka dari itu perlu dilakukan monitoring dan evalusi sehingga karakter kedisiplinan bisa tetap terjaga. Monitoring dan evaluasi data kehadiran akan lebih mudah dilakukan ketika data kehadiran pegawai tersajikan dalam bentuk pemetaan. Pemetaan kehadiran dapat dikelompokkan ke dalam beberapa sub sesuai dengan karakteristik kedisiplinan. Salah satu metode komputasi yang dapat digunakan untuk melakukan pengelompokkan sebuah data adalah metode *Data Mining*.

Dengan ketersediaan data yang luas dan kebutuhan untuk mengubahnya menjadi informasi dan pengetahuan yang bermanfaat, *data mining* telah menarik perhatian industri informasi dalam beberapa tahun terakhir (Ramasamy & Nirmala, 2020). *Data mining* adalah teknik yang digunakan untuk menemukan struktur data dari kumpulan data yang besar. Hasil dari proses ini memungkinkan para pengambil keputusan untuk membuat keputusan yang tepat tentang bagaimana mengembangkan bisnis ke depannya (Hossain et al., 2019). *Clustering* adalah salah satu metode *Data Mining*.

Teknik pengelompokan data *clustering* menggunakan label kelas tanpa mendefinisikan kelas data pengetahuan (Rodriguez et al., 2019). Tujuan utama melakukan pengelompokan data pada data besar adalah untuk mengurutkan kumpulan data terkait dan kemudian menemukan pola dari data yang dikelompokkan (Chen et al., 2019). Pada teknik *Clustering* sendiri pertama, *Clustering* diterapkan di ruang desain dan hasilnya kemudian divisualisasikan di ruang objektif. Setelah pengelompokkan, detail fitur di setiap *Cluster* dianalisis berdasarkan konsep analisis aturan asosiasi, sehingga substruktur karakteristik dapat diekstraksi dari setiap *Cluster* solusi (Sato et al., 2019). Salah satu metode *Clustering* yang paling populer dan tercepat adalah metode *K-Means* (Manochandar et al., 2020).

Metode *K-Means* merupakan metode yang terkenal karena efisiensinya dalam mengelompokkan kumpulan data yang besar, bekerja hanya pada data numerik bukan diterapkan untuk mengelompokkan data kategorikal (Nguyen et al., 2019). Algoritma *clustering* yang terkenal yaitu K-Means memperoleh partisi k objek melalui perbaikan berulang sehingga jumlah jarak kuadrat antara objek dan pusat cluster diminimalkan (Jothi et al., 2019). Cara kerja *K-Means* dimulai dengan k pusat awal yang dipilih secara acak, *K-Means* menetapkan setiap objek ke sebuah *Cluster* yang memiliki pusat terdekat. Kemudian, pusat *Cluster* dihitung ulang menggunakan keanggotaan objek saat ini. Sekali lagi objek dipindahkan ke pusat terdekat. Proses ini berulang sampai tidak ada perubahan pada pusat *Cluster*.

Penelitian terdahulu tentang *K-Means Clustering* adalah tentang penyajian pendekatan pengelompokkan untuk mempartisi siswa ke dalam kelompok yang berbeda berdasarkan perilaku belajar siswa dengan metode *Data Mining* (Kausar et al., 2018). Pada penelitian bidang kesehatan metode *K-Means* dikombinasikan dengan *neural network* untuk deteksi objek dan identifikasi kelainan tumor otak (Arunkumar et al., 2019). Pada penelitian lain metode *K-Means* digunakan untuk mengklasifikasikan pegawai di sebuah perusahaan, kemudian setiap jenis algoritma pohon keputusan digunakan untuk melakukan prediksi dan analisis turnover (Yunmeng & Chengyi, 2019).

Kontribusi kami dalam penelitian ini ialah melakukan *Cluster*isasi terhadap rekap kehadiran pada daftar hadir pegawai di Universitas Harapan Bangsa. Hasil yang diharapkan adalah adanya pengelompokkan data sesuai dengan kedisiplinan kerja pegawai. Keputusan ini akan menjadi bahan evaluasi bagi instansi untuk membuat kebijakan SOP rekap kehadiran ke depannya. Sehingga ke depan tradisi menjunjung tinggi nilai kedisiplinan di kampus Universitas Harapan Bangsa dapat lebih ditingkatkan.

KAJIAN LITERATUR

Terdapat penelitian tentang Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma *K-Means Clustering* (Virgo et al., 2020). Penelitian ini bertujuan untuk mengimplementasikan sebuah aplikasi yang dapat mencatat jumlah pertemuan yang dilakukan oleh dosen selama proses belajar mengajar, dengan tujuan untuk menggunakan data pertemuan tersebut sebagai penilaian terhadap kinerja mereka. Penelitian ini berhasil mengelompokkan dosen menjadi tiga kelompok: kelompok jarang melakukan pertemuan (4.7650%), kelompok sedang dalam melakukan pertemuan (4.5665%), dan kelompok rajin melakukan pertemuan (90.6684%). Pendekatan *Knowledge Discovery in Database* (KDD) dan Algoritma *K-Means*

Clustering dalam *Data Mining* digunakan. Hasil penelitian menunjukkan bahwa dosen pengampu matakuliah tertentu cenderung rajin menghadiri setiap pertemuan pada tahun akademik 2017/2018 semester gasal dan genap, dengan kehadiran rata-rata antara delapan belas hingga dua belas kali pertemuan per semester.

Penelitian lain yaitu tentang Penerapan *Data Mining* untuk Memprediksi Kinerja Akademik Menggunakan *K-means Clustering* dan Klasifikasi *Naïve Bayes* (Ali et al., 2020). Penelitian ini bertujuan untuk meningkatkan kualitas sistem pendidikan di perguruan tinggi dengan menggunakan proses penambangan data pendidikan menggunakan *algoritma k-means clustering* dan klasifikasi *Naïve Bayes*. Penelitian ini berfokus pada mengeksplorasi data evaluasi mahasiswa terkait kinerja pengajar untuk memahami atribut utama yang dapat memengaruhi kinerja pendidikan di berbagai mata kuliah. Sistem yang diusulkan ini membantu mengidentifikasi mahasiswa yang berhenti belajar dan memberikan saran atau konseling yang sesuai untuk manajemen pendidikan dalam mengambil keputusan yang berpengetahuan.

Penelitian lainnya tentang Analisis Perilaku Mahasiswa Berdasarkan Algoritma *Fusion K-Means Clustering* (Chang et al., 2020). Penelitian ini bertujuan untuk meningkatkan efisiensi dan efek klasterisasi dalam analisis data dengan mengusulkan algoritma berbasis *K-Means* dan K-CFSFDP. Hasil analisis klaster menunjukkan bahwa mahasiswa dari kategori yang berbeda di empat universitas memiliki performa yang berbeda dalam kebiasaan hidup dan performa belajar, sehingga universitas dapat memahami perilaku mahasiswa dari berbagai kategori dan memberikan layanan personalisasi yang sesuai, yang memiliki signifikansi praktis tertentu.

Penelitian terkait lainnya yaitu tentang Strategi Marketing Penerimaan Mahasiswa Baru Menggunakan Machine Learning dengan Teknik Clustering (Dana et al., 2019). Dengan menggunakan teknologi *machine learning* dan metode *clustering*, penelitian ini bertujuan untuk meningkatkan efektivitas kegiatan penerimaan mahasiswa baru di perguruan tinggi. Hasilnya, pendaftar dikelompokkan menjadi tiga kelompok, yaitu kelompok 1 sebesar 11%, kelompok 2 sebesar 56%, dan kelompok 3 sebesar 33%. Ini memudahkan penentuan strategi dan pola pemasaran yang sesuai dengan karakteristik masing-masing kelompok pendaftar, yang diharapkan akan meningkatkan minat dan keberhasilan penerimaan mahasiswa baru.

METODE PENELITIAN

Metode penelitian dilakukan mulai dari pengumpulan data, pemodelan data, penentuan jumlah *Cluster*, analisis data sampai dengan menghasilkan output yang diharapkan. Data dikumpulkan untuk mendapatkan informasi yang diperlukan untuk mencapai tujuan penelitian. Pemodelan data digunakan untuk menyeleksi data sehingga mudah dipahami dan memastikan bahwa data tersebut akurat untuk proses analisis. Penentuan jumlah *Cluster* dihitung menggunakan metode *elbow*. Analisis data menggunakan metode *K-Means*, digunakan untuk melakukan pemetaan kedisiplinan pegawai sesuai dengan data rekap kehadiran yang sudah ditransformasi sesuai dengan kebutuhan.

Metode Pengumpulan Data

Data kehadiran pegawai di Universitas Harapan Bangsa digunakan sebagai sumber penelitian ini. Data tersebut dianalisis untuk menghasilkan pengelompokkan kedisiplinan pegawai. Data diambil dari tanggal 08 Juni 2023 sampai dengan 21 September 2023. Jumlah keseluruhan rekap kehadiran yang tersimpan adalah 20.935 data, dengan jumlah pegawai yaitu 121 orang. Sebelum memulai proses analisis dengan menggunakan metode *K-Means Clustering* data yang telah ditransformasi, beberapa fitur yang tidak diperlukan dihilangkan. Data rekap kehadiran yang sudah disesuaikan dapat dilihat pada Tabel 1.

Tabel I. Data Rekap Kehadiran

Nama	Jam Berangkat	Jam Pulang
Pep	07:30:03	16:03:35
Gal	07:04:23	16:05:03
Dan	07:11:51	16:01:17
Had	07:13:17	16:06:25
Des	07:14:13	16:01:26
Hes	07:14:14	16:01:36
Fua	07:14:24	16:02:16
Est	07:15:28	16:04:40
Adi	07:15:34	16:03:38
Arl	07:16:50	16:10:01
Ala	07:17:04	16:00:17

Langkah-langkah prosedur menggunakan k melibatkan metode pengelompokan yang terdiri dari beberapa tahapan. Pertama, hitung jumlah *Cluster* yang diinginkan dengan nilai k. Kemudian, pilih *centroid*, pusat *Cluster* awal, sejumlah k nilai tersebut. Selanjutnya, gunakan rumus jarak *Euclidean* untuk

menghitung jarak dari setiap data input ke *centroid*, dan cari *centroid* terdekat dari setiap data. Langkah selanjutnya adalah mengklasifikasikan atau mengelompokkan setiap item data berdasarkan kedekatannya, di mana data akan ditempatkan dalam *Cluster* dengan jarak minimum ke *centroid*. Terakhir, perbarui nilai pusat dengan mengambil *mean Cluster* tersebut menggunakan rumus yang sesuai.

Metode *Elbow*

Metode *Elbow* digunakan untuk menjelaskan dan memverifikasi konsistensi analisis *Clustering*, yang bertujuan untuk membantu menemukan jumlah *Cluster* yang sesuai dalam kumpulan data (Purwono et al., 2021). Metode *Elbow* membandingkan jumlah *Cluster* yang akan membentuk siku dengan persentase hasil. Langkah pertama metode *elbow*, derajat distorsi yang diperoleh dengan metode *Elbow* dinormalisasi dengan kisaran 0 sampai 10. Kedua, hasil normalisasi digunakan untuk menghitung kosinus sudut perpotongan antar titik siku. Ketiga, perhitungan kosinus sudut perpotongan dan teorema *arccosinus* ini digunakan untuk menghitung sudut perpotongan antara titik siku. Akhirnya, indeks sudut perpotongan minimal yang dihitung di atas antara titik siku digunakan sebagai perkiraan jumlah *Cluster* optimal potensial (Liu & Deng, 2021).

Untuk mengetahui persentase hasil perbandingan antara jumlah *Cluster* adalah dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai *Cluster*. Nilai SSE akan berkurang seiring dengan jumlah *Cluster* K. Rumus SSE pada *K-Means* dapat dilihat pada persamaan (1).

$$\text{SSE} = \sum_{k=1}^{k} \sum_{Xi \epsilon Sk} ||x_i - C_k||_2^2 \qquad (1)$$

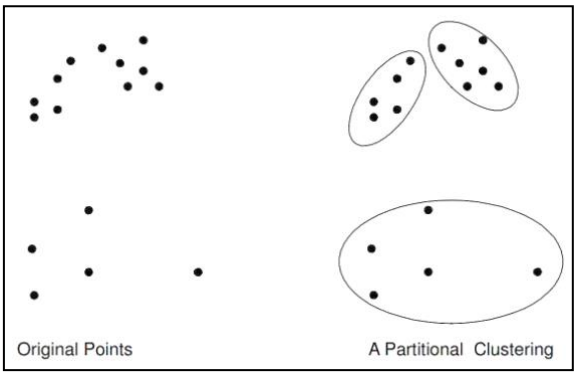
Keterangan dari persamaan berikut adalah k merupakan *Cluster* ke-c, x_i merupakan jarak data obyek ke-I, dan C_k merupakan pusat *Cluster* ke-i. Contoh hasil dari *K-Means Clustering* menggunakan metode *elbow* dapat dilihat pada tabel 2.

Tabel 2. Data Clustering dengan metode *elbow*

Nama	Jam Berangkat	Jam Pulang	*Cluster*
Pep	7:30:03	16:03:35	0
Ang	7:22:58	17:03:49	1
Noo	7:08:14	16:16:26	2

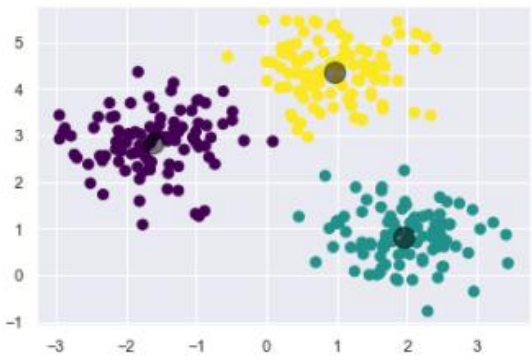
Metode K-Means Clustering

Metode *K-Means* adalah metode *clustering* yang paling umum dan populer (Sinaga & Yang, 2020). Metode penganalisa data *K-Means Clustering* melakukan proses pemodelan tanpa supervisi dan termasuk dalam kelompokan data dengan sistem partisi. Dua jenis model pengelompokan data yang paling umum digunakan dalam proses pengelompokan data adalah *hierarchical* dan *non-hierarchical.* Metode pengelompokan data *K-Means* menggunakan model *non-hierarchical,* juga dikenal sebagai partisional pengelompokan. Model partitional *Clustering* dapat dilihat pada gambar 1.



Gambar 1. *Partitional Clustering*

Metode *K-Means Clustering* mengelompokkan data ke dalam beberapa kelompok, masing-masing dengan karakteristik yang sama dan berbeda satu sama lain. Tujuan dari metode ini adalah untuk mengurangi fungsi tujuan yang ditetapkan selama proses *clustering* dengan mengurangi variasi antara data dalam kelompok dan meningkatkan variasi antara data dalam kelompok. Contoh Cluster dengan jumlah 3 dapat dilihat pada gambar 2.



Gambar 2. Contoh *Cluster*

Tahapan proses *Clustering* menggunakan metode *K-Means Clustering* umumnya mengikuti algoritma dasar sebagai berikut: Pertama, tentukan jumlah *Cluster* yang diinginkan. Selanjutnya, alokasikan data secara acak ke dalam *Cluster*. Kemudian, hitung *centroid* atau rata-rata dari data yang

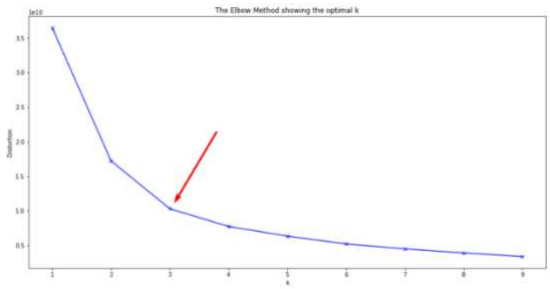
ISSN: 2598-8565 (media cetak)
ISSN: 2620-4339 (media online)

terdapat di setiap *Cluster*. Langkah berikutnya adalah mengalokasikan setiap data ke *centroid* atau rata-rata terdekat. Proses tersebut dilanjutkan dengan kembali ke langkah ketiga, jika masih ada data yang berpindah *Cluster*, jika terdapat perubahan nilai *centroid* di atas nilai *threshold* yang ditentukan, atau jika terdapat perubahan nilai pada fungsi tujuan di atas nilai *threshold* yang ditentukan.

HASIL DAN PEMBAHASAN

Menentukan Jumlah Cluster dengan Metode Elbow

Percobaan dilakukan menggunakan data presensi yang diambil dari tanggal 08 Juni 2023 sampai dengan 21 September 2023. Jumlah *Cluster* yang dipakai pada penelitian ini yaitu sebanyak 3 *Cluster*. Jumlah *Cluster* diambil berdasarkan perhitungan metode *Elbow*. Metode *Elbow* menghitung jumlah Cluster yang tepat dengan membandingkan persentase hasil antara jumlah Cluster yang akan membentuk siku pada titik tertentu. Hasil perhitungan metode ini dapat dilihat pada gambar 3.



Gambar 3. Hasil Perhitungan Metode *Elbow*

Menentukan Jumlah Cluster dengan Metode Elbow

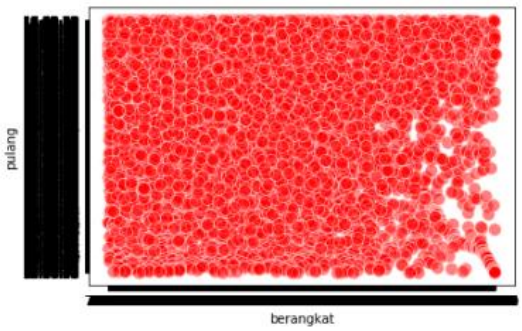
Setelah mendapatkan jumlah *cluster* yang ideal, langkah selanjutnya adalah menggabungkan data kehadiran menggunakan metode *K-Means*. Metode *K-Means* memulai langkahnya dengan memilih secara acak kelompok pertama *centroid* yang digunakan sebagai titik awal untuk setiap *Cluster*. Langah berikutnya adalah melakukan perhitungan berulang ulang untuk mengoptimalkan posisi *centroid*. Proses *K-Means* berhenti ketika *centroid* telah stabil atau jumlah iterasi yang ditentukan telah tercapai.

Variabel dari data kehadiran yang akan di *Cluster* adalah data waktu berangkat dan data waktu pulang, sehingga transformasi data berubah seperti terlihat pada gambar 4.

	berangkat	pulang	kluster
0	7:30:03	16:03:35	0
1	7:04:23	16:05:03	2
2	7:11:51	16:01:17	2
3	7:13:17	16:06:25	2
4	7:14:13	16:01:26	2
5	7:14:14	16:01:36	2
6	7:14:24	16:02:16	2
7	7:15:28	16:04:40	0
8	7:15:34	16:03:38	0
9	7:16:50	16:10:01	0

Gambar 4. Transformasi Data

Transformasi data yang telah dibuat selanjutnya di cek rentang nilainya. Rentang nilai yang terlalu jauh akan menyebabkan plot tidak muncul dengan sempurna. Persebaran dari transformasi data rekap kehadiran yang sudah disesuaikan dapat dilihat pada Gambar 5.



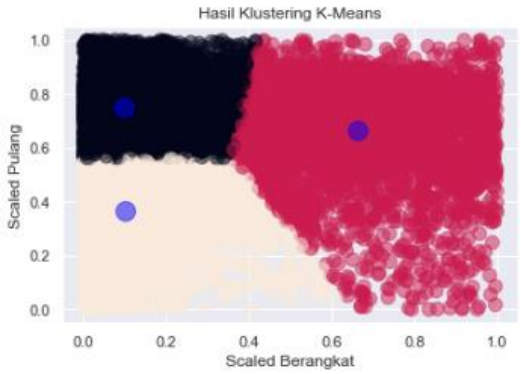
Gambar 5. Persebaran Data Kehadiran

Ukuran dari rentang persebaran data sudah sangat baik sehingga tidak perlu dilakukan standarisasi data lagi. Ukuran persebaran data yang merata akan membuat plot muncul dengan sempurna. Langkah terakhir adalah menambahkan kolom *Cluster* kedalam data *frame*. Hasil akhir klasterisasi dengan metode *K-Means* dapat dilihat pada Gambar 6.

berangkat	pulang	kluster
7:30:03	16:03:35	0
7:04:23	16:05:03	2
7:11:51	16:01:17	2
7:13:17	16:06:25	2
7:14:13	16:01:26	2
7:14:14	16:01:36	2
7:14:24	16:02:16	2
7:15:28	16:04:40	0
7:15:34	16:03:38	0
7:16:50	16:10:01	0

Gambar 6. Hasil *Cluster*isasi dengan Metode *K-Means*

Untuk memudahkan membaca persebaran *Cluster*, data di visualisasikan ke dalam grafik plot. Visualisasi hasil *Cluster* dalam bentuk grafik plot dapat dilihat pada gambar Gambar 7.



Gambar 7. Visualisasi Hasil *Cluster* Grafik Plot

Dari hasil analsis *Cluster* didapat data presensi sejumlah 12.659 berada pada *Cluster* 0, 3.552 berada pada *Cluster* 1, dan 4.724 berada pada *Cluster* 2. Bentuk karakteristik dari *Cluster* 0 adalah waktu berangkat mendekati jam batas waktu masuk kerja, sedangkan waktu pulang lebih sedikit dari ketentuan waktu pulang kerja. Karakteristik dari *Cluster* 1 adalah waktu berangkat mendekati jam batas waktu masuk kerja, sedangkan waktu pulang lebih lama dari ketentuan waktu pulang kerja. Karakteristik dari *Cluster* 2 adalah waktu berangkat lebih pagi dari jam batas waktu masuk kerja, sedangkan waktu pulang lebih sedikit dari ketentuan waktu pulang kerja. Dari karakteristik tersebut dapat diambil kesimpulan bahwa data presensi yang berada pada *Cluster* 2 adalah presensi dengan karakteristik terbaik. Dari hasil percobaan didapat 5 pegawai dengan nilai *Cluster* 0 terbanyak dapat dilihat pada tabel 3.

Tabel 3. Data Pegawai dengan Nilai Cluster 0 terbanyak

No	Nama	Total *Cluster* 0
1	Gal	227
2	Bas	221
3	Evi	214
4	Ari	173
5	Sun	169

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, metode *K-Means* dapat digunakan untuk pengelompokkan kedisiplinan pegawai menggunakan data rekap kehadiran. Metode *K-Means* mengelompokkan karakter kedisplinan kedalam 3 *Cluster*, yaitu 0, 1, dan 2. Analisis *Cluster* menunjukkan bahwa *Cluster* 2 merupakan *Cluster* yang mempunyai tingkat kedisiplinan lebih baik dari *Cluster* lainnya. Kedisiplinan terbaik diambil dari pegawai yang mempunyai waktu berangkat lebih pagi dari jam batas waktu masuk kerja, sedangkan waktu pulang lebih sedikit dari ketentuan waktu pulang kerja. Pengelompokkan kedisiplinan dengan rekap kehadiran ini mudah dibaca sehingga akan membantu pihak pengambil keputusan untuk menilai kedisiplinan pegawai di dalam perusahannya. Kekurangan dalam penelitian ini adalah masih perlunya indikator lain untuk menentukan kedisiplinan kerja, seperti ketaatan dalam mematuhi semua peraturan dan tanggung jawab terhadap sebuah pekerjaan. Harapan pengembangan penelitian ke depan adalah adanya indikator lain yang digunakan dalam menilai kedisiplinan pegawai, sehingga hasil yang didapat akan lebih akurat.

DAFTAR PUSTAKA

Ali, Z. M., Hassoon, N. H., Ahmed, W. S., & Abed, H. N. (2020). The Application of Data Mining for Predicting Academic Performance Using K-means Clustering and Naïve Bayes Classification. *International Journal of Psychosocial Rehabilitation*, *24*(03), 2143–2151. https://doi.org/10.37200/ijpr/v24i3/pr200962

Arunkumar, N., Mohammed, M. A., Abd Ghani, M. K., Ibrahim, D. A., Abdulhay, E., Ramirez-Gonzalez, G., & de Albuquerque, V. H. C. (2019). K-Means clustering and neural network for object detecting and identifying abnormality of brain tumor. *Soft Computing*, *23*(19), 9083–9096. https://doi.org/10.1007/s00500-018-3618-7

Chang, W., Ji, X., Liu, Y., Xiao, Y., Chen, B., Liu, H., & Zhou, S. (2020). Analysis of university students' behavior based on a fusion K-means clustering algorithm. *Applied Sciences (Switzerland)*, *10*(18). https://doi.org/10.3390/APP10186566

Chen, W., Oliverio, J., Kim, J. H., & Shen, J. (2019). The Modeling and Simulation of Data Clustering Algorithms in Data Mining with Big Data. *Journal of Industrial Integration and Management*, *04*(01), 1850017. https://doi.org/10.1142/s2424862218500173

Dana, R. D., Rohmat, C. L., & Rinaldi, A. R. (2019). Strategi Marketing Penerimaan Mahasiswa Baru Menggunakan Machine Learning dengan Teknik Clustering. *Jurnal Informatika: Jurnal Pengembangan IT*, *4*(2–2), 201–204. https://doi.org/10.30591/jpit.v4i2-2.1879

H Kara, O. A. M. A. (2014). *Paper Knowledge. Toward a Media History of Documents*, *7*(2), 107–115.

Hossain, M. Z., Akhtar, M. N., Ahmad, R. B., & Rahman, M. (2019). A dynamic K-means clustering for data mining. *Indonesian Journal of Electrical Engineering and Computer Science*, *13*(2), 521–526. https://doi.org/10.11591/ijeecs.v13.i2.pp521-526

Jothi, R., Mohanty, S. K., & Ojha, A. (2019). DK-means: a deterministic K-means clustering algorithm for gene expression analysis. *Pattern Analysis and Applications*, *22*(2), 649–667. https://doi.org/10.1007/s10044-017-0673-0

Kausar, S., Huahu, X., Hussain, I., Wenhao, Z., & Zahid, M. (2018). Integration of Data Mining Clustering Approach in the Personalized E-Learning System. *IEEE Access*, *6*, 72724–72734. https://doi.org/10.1109/ACCESS.2018.2882240

Liu, F., & Deng, Y. (2021). Determine the Number of Unknown Targets in Open World Based on Elbow Method. *IEEE Transactions on Fuzzy Systems*, *29*(5), 986–995. https://doi.org/10.1109/TFUZZ.2020.2966182

Manochandar, S., Punniyamoorthy, M., & Jeyachitra, R. K. (2020). Development of new seed with modified validity measures for k-means clustering. *Computers and Industrial Engineering*, *141*(July 2018), 106290. https://doi.org/10.1016/j.cie.2020.106290

Nguyen, T. H. T., Dinh, D. T., Sriboonchitta, S., & Huynh, V. N. (2019). A method for k-means-like clustering of categorical data. *Journal of Ambient Intelligence and Humanized Computing*, *Berkhin 2002*. https://doi.org/10.1007/s12652-019-01445-5

Purwono, P., Ma'arif, A., Mangku Negara, I. S., Rahmaniar, W., & Rahmawan, J. (2021). Linkage Detection of Features that Cause Stroke using Feyn Qlattice Machine Learning Model. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, *7*(3), 423. https://doi.org/10.26555/jiteki.v7i3.22237

Putri, E. M., Ekowati, V. M., Supriyanto, A. S., & Mukaffi, Z. (2019). the Effect of Work Environment on Employee Performance Through Work Discipline. *International Journal of Research -GRANTHAALAYAH*, *7*(4), 132–140. https://doi.org/10.29121/granthaalayah.v7.i4.2019.882

Ramasamy, S., & Nirmala, K. (2020). Disease prediction in data mining using association rule mining and keyword-based clustering algorithms. *International Journal of Computers and Applications*, *42*(1), 1–8. https://doi.org/10.1080/1206212X.2017.1396415

Rodriguez, M. Z., Comin, C. H., Casanova, D., Bruno, O. M., Amancio, D. R., Costa, L. da F., & Rodrigues, F. A. (2019). Clustering algorithms: A comparative approach. In *PLoS ONE* (Vol. 14, Issue 1). https://doi.org/10.1371/journal.pone.0210236

Saputra, T. (2019). Pengaruh Motivasi Kerja Terhadap Disiplin Kerja Karyawan Pada Hotel Permai Pekanbaru. *Jurnal Benefita*, *4*(2), 316. https://doi.org/10.22216/jbe.v4i2.1548

Sato, Y., Izui, K., Yamada, T., & Nishiwaki, S. (2019). Data mining based on clustering and association rule analysis for knowledge discovery in multiobjective topology optimization. *Expert Systems with Applications*, *119*, 247–261. https://doi.org/10.1016/j.eswa.2018.10.047

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Virgo, I., Defit, S., & Yuhandri, Y. (2020). Klasterisasi Tingkat Kehadiran Dosen Menggunakan Algoritma K-Means Clustering. *Jurnal Sistim Informasi Dan Teknologi*, *2*, 23–28. https://doi.org/10.37034/jsisfotek.v2i1.17

Yunmeng, Z., & Chengyi, Z. (2019). The Application of the Decision Tree Algorithm Based on K-means in Employee Turnover Prediction. *Journal of Physics: Conference Series*, *1325*(1). https://doi.org/10.1088/1742-6596/1325/1/012123