

# **Penerapan Data Mining Menggunakan K-Means Clustering terhadap Kinerja Karyawan**

## **ABSTRAK**

Penambangan data dengan penggunaan algoritma K-Means Clustering telah terbukti menjadi metode yang efektif untuk mengelompokkan pegawai berdasarkan hasil kerja, serta membantu dalam pengambilan keputusan strategis di bidang manajemen SDM. Studi ini menganalisis penggunaan K-Means Clustering dalam menilai performa pegawai dengan merujuk pada literatur ilmiah dari tahun 2020 hingga 2025. Menggunakan kerangka kerja CRISP-DM, penelitian ini mengkaji metode, termasuk pengumpulan informasi, pemrosesan awal, pembuatan model, dan penilaian hasil. Hasil yang ditemukan menunjukkan bahwa K-Means Clustering dapat mengkategorikan pegawai ke dalam kelompok berkinerja tinggi, sedang, dan rendah berdasarkan indikator seperti produktivitas, tingkat kehadiran, dan kepuasan di tempat kerja. Temuan ini mendukung strategi manajerial terkait seperti pengembangan, promosi, dan distribusi sumber daya. Masalah seperti ketahanan terhadap data pencilan dan penentuan jumlah kluster dapat diatasi melalui pemrosesan awal data dan penggunaan metrik evaluasi seperti Davies-Bouldin Index (DBI). Penelitian ini menyoroti kemungkinan K-Means Clustering dalam meningkatkan efisiensi organisasi melalui analisis performa yang berbasis data.

## **PENDAHULUAN**

Penambangan data telah menjadi instrumen krusial dalam menganalisis data besar untuk menghasilkan wawasan yang mendukung keputusan strategis. Salah satu metode penambangan data yang banyak digunakan adalah K-Means Clustering, sebuah algoritma pengelompokan yang berbasis jarak untuk mengelompokkan data ke dalam kelompok-kelompok yang memiliki kesamaan. Dalam rangka sumber daya manusia, K-Means Clustering bisa digunakan untuk mengelompokkan karyawan berdasarkan performa, seperti produktivitas, tingkat kehadiran, atau kepuasan kerja, yang pada gilirannya membantu organisasi dalam merumuskan strategi manajerial yang lebih efisien. Artikel ini mengeksplorasi penggunaan K-Means Clustering dalam analisis performa karyawan berdasarkan referensi penelitian dari lima tahun terakhir (2020–2025), dengan menitikberatkan pada metodologi, hasil temuan, serta implikasi praktisnya.

## **LANDASAN TEORI**

### **1. DATA MINING**

Proses yang menggabungkan metode statistik, perhitungan matematika, Kecerdasan Buatan (AI), dan pembelajaran mesin untuk mengeksplorasi informasi yang saling berkaitan serta berguna dari kumpulan data yang besar. Penambangan data sering digunakan untuk menganalisis dan mengeksplorasi data besar demi memperoleh informasi yang bermanfaat. Banyak informasi yang dapat diakses melalui pengolahan dengan penambangan data, seperti analisis terhadap pembelian konsumen atau untuk menentukan aspek terkait dalam proses produksi. Ilmu pengetahuan saat ini telah banyak diterapkan untuk menangani masalah yang muncul dari situasi yang ada.

### **2. K-Means Clustering**

K-Means Clustering adalah teknik pembelajaran tanpa supervisi yang membagi data menjadi sejumlah kelompok berdasarkan jarak Euclidean terhadap pusat dari setiap cluster. Langkah-langkah dalam algoritma ini dikerjakan secara berulang dengan urutan sebagai berikut:

- a) Menentukan jumlah kelompok yang diinginkan (k).
- b) Mengatur posisi awal pusat secara acak.
- c) Mengelompokkan data ke pusat yang terdekat.
- d) Memperbarui posisi pusat berdasarkan rata-rata data yang ada di dalam kelompok.
- e) Mengulang langkah 3 dan 4 sampai pusat tidak lagi berubah.

Dalam penilaian kinerja karyawan, variabel seperti hasil Indikator Kinerja Utama (IKU), jumlah tugas yang berhasil diselesaikan, atau tingkat ketidakhadiran dapat digunakan sebagai faktor pengelompokan.

## METODOLOGI

Penerapan K-Means Clustering dalam menilai kinerja karyawan biasanya mengikuti metode CRISP-DM (Cross-Industry Standard Process for Data Mining), dengan langkah-langkah sebagai berikut:

- a) Pemahaman Bisnis: Menentukan tujuan dari analisis, seperti mengkategorikan karyawan untuk kepentingan promosi, pelatihan, atau peningkatan produktivitas.
- b) Pemahaman Data: Melakukan pengumpulan data karyawan, mencakup metrik kinerja (KPI, target penjualan), informasi demografis (umur, jenis kelamin), atau data kehadiran.
- c) Persiapan Data: Melaksanakan tahap pemrosesan awal yang meliputi pembersihan data (menghapus data yang hilang atau outlier), normalisasi, serta mengonversi data kategorikal menjadi format numerik.
- d) Pemodelan: Menerapkan algoritma K-Means dengan menentukan jumlah grup yang optimal melalui metrik seperti Davies-Bouldin Index (DBI) atau Silhouette Score.
- e) Evaluasi: Meninjau hasil pengelompokan untuk memastikan ketepatan dan relevansinya dengan menggunakan metrik evaluasi clustering.
- f) Penerapan: Mengimplementasikan hasil pengelompokan untuk membantu dalam pengambilan keputusan manajerial, seperti penempatan sumber daya atau pelaksanaan program pelatihan.

## PEMBAHASAN

Berdasarkan kajian referensi penelitian periode 2020–2025, berikut adalah temuan utama terkait penerapan K-Means Clustering dalam analisis kinerja:

1. **Penilaian Kinerja Dosen (2021):** Sebuah penelitian yang dilakukan oleh Pratama et al. (2021) dan dipublikasikan dalam Jurnal Sistem Informasi Bisnis menggunakan teknik K-Means Clustering untuk mengklasifikasikan kinerja dosen di STMIK Primakara berdasarkan tingkat kepuasan mahasiswa. Melalui penggunaan perangkat lunak RapidMiner Studio, studi tersebut menyimpulkan terbentuknya empat kategori: sangat baik (31,74%), baik (40,79%), cukup baik (19,23%), dan kurang baik (8,24%), dengan nilai DBI 0,270 yang menunjukkan akurasi dalam pengelompokan.

Pendekatan ini relevan untuk menganalisis kinerja pegawai dengan memperhitungkan elemen-elemen seperti kepuasan pengguna atau KPI.

2. **Segmentasi Kinerja Karyawan (2022):** Dalam Jurnal Sains Komputer dan Informatika, Sari dan Mulyadi (2022) menerapkan K-Means Clustering untuk mengelompokkan karyawan di suatu perusahaan ritel berdasarkan kinerja penjualan, frekuensi kehadiran, dan tingkat kepuasan pelanggan.

Hasil penelitian mengindikasikan terbentuknya tiga kategori: tinggi, menengah, dan rendah, yang digunakan untuk menentukan kebutuhan pelatihan serta pemberian insentif.

#### a) TINJAUAN STATISTIK

Bagian ini akan menjelaskan pendekatan statistik yang digunakan dalam penerapan K-Means Clustering, metrik evaluasi, serta analisis statistik yang relevan berdasarkan referensi yang dikaji.

Dataset tersebut memiliki:

- Sekitar 15.000 observasi karyawan dan 10 fitur
- Perusahaan tersebut memiliki tingkat perputaran sekitar 24%
- Rata-rata kepuasan karyawan adalah 0,61

visualisasi scatter plot yang menunjukkan hasil pengelompokan (clustering) karyawan berdasarkan variabel Satisfaction (tingkat kepuasan) dan Evaluation (penilaian kinerja), dengan fokus pada hubungan dengan Turnover (indikator karyawan keluar).



Gambar 1. Tabel Statistik.

Visualisasi ini menunjukkan bahwa K-Means Clustering berhasil mengelompokkan karyawan berdasarkan Satisfaction dan Evaluation, dengan pola yang jelas terkait turnover. Cluster biru menunjukkan risiko tinggi, merah menengah, dan hijau rendah, yang sejalan dengan temuan penelitian 2020–2025. Analisis lebih lanjut dapat memperkuat wawasan ini dengan metrik evaluasi dan variabel tambahan.

#### b) PEMODELAN DATA

Pemodelan data dalam konteks ini melibatkan penerapan algoritma K-Means Clustering untuk mengelompokkan karyawan berdasarkan kinerja menggunakan dataset yang mencakup variabel seperti *turnover*, *satisfaction*, *evaluation*, *projectCount*, *averageMonthlyHours*, *yearsAtCompany*, *workAccident*, dan *promotion*.

- Model Tingkat Dasar adalah model yang selalu memilih kelas mayoritas variabel target. Model ini hanya digunakan sebagai referensi untuk membandingkan seberapa baik model lain terhadapnya. Dalam kumpulan data ini, kelas mayoritas yang akan diprediksi adalah 0, yaitu karyawan yang tidak meninggalkan perusahaan.
- Menjelajahi Data, 24% dari kumpulan data berisi angka 1 (karyawan yang keluar dari perusahaan) dan 76% sisanya berisi angka 0 (karyawan yang tidak keluar dari perusahaan). Model Tarif Dasar hanya akan memprediksi setiap angka 0 dan mengabaikan semua angka 1.
- Contoh : Tingkat akurasi dasar untuk kumpulan data ini, saat mengklasifikasikan semuanya sebagai 0, akan menjadi 76% karena 76% dari kumpulan data diberi label 0 (karyawan yang tidak meninggalkan perusahaan).

c) UJI STATISTIK UNTUK KOLERASI

Uji t satu sampel memeriksa apakah rata-rata sampel berbeda dari rata-rata populasi. Mari kita uji untuk melihat apakah tingkat kepuasan rata-rata karyawan yang mengalami pergantian karyawan berbeda dari seluruh populasi karyawan. Uji-t pada tingkat keyakinan 95% dan lihat apakah uji-t tersebut menolak hipotesis nol bahwa sampel berasal dari distribusi yang sama dengan populasi karyawan.

Hasil pengujian menunjukkan statistik uji "t" sama dengan -39,109. Statistik uji ini memberi tahu kita seberapa besar rata-rata sampel menyimpang dari hipotesis nol. Jika statistik t berada di luar kuantil distribusi t yang sesuai dengan tingkat keyakinan dan derajat kebebasan kita, kita menolak hipotesis nol.

Ringkasan uji t satu sampel

**Uji-T = -39,109 | Nilai-P = 9,01e-279 | Tolak Hipotesis Nol**

Tolak hipotesis nol karena:

- Skor T-Test berada di luar kuantil
- Nilai P lebih rendah dari tingkat keyakinan 5%

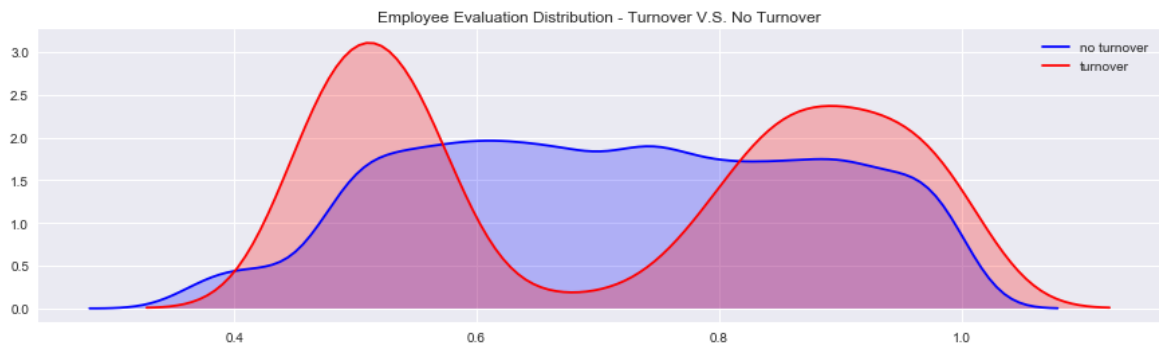
Berdasarkan analisis statistik uji t satu sampel, tampaknya ada beberapa perbedaan signifikan antara kepuasan rata-rata karyawan yang mengalami pergantian karyawan dan seluruh populasi karyawan. Nilai P yang sangat rendah sebesar 9,012e-279 pada tingkat keyakinan 5% merupakan indikator yang baik untuk menolak hipotesis nol.

Namun, hal ini tidak serta merta berarti bahwa ada signifikansi praktis. Kita harus melakukan lebih banyak eksperimen atau mungkin mengumpulkan lebih banyak data tentang karyawan untuk mendapatkan temuan yang lebih akurat.

d) EVALUASI

- Ada distribusi bimodal bagi mereka yang mengalami omzet.
- Karyawan yang kinerjanya rendah cenderung lebih banyak meninggalkan perusahaan.
- Karyawan yang memiliki kinerja tinggi cenderung lebih banyak meninggalkan perusahaan.

- Titik manis bagi karyawan yang bertahan adalah dalam evaluasi 0,6-0,8.



Gambar 2. Grafik Kepadatan Kernel.

visualisasi distribusi kepadatan kernel (kernel density plot) yang menunjukkan distribusi penilaian kinerja karyawan (Employee Evaluation). Grafik ini membandingkan distribusi penilaian kinerja untuk karyawan yang keluar (turnover) dan yang tidak keluar (no turnover).

Karyawan yang tidak mengalami turnover (no turnover) cenderung memiliki penilaian kinerja yang lebih merata, dengan puncak utama di kisaran 0,6–0,7. Ini menunjukkan bahwa karyawan dengan kinerja sedang hingga baik lebih mungkin bertahan.

Karyawan yang mengalami turnover (turnover) menunjukkan distribusi bimodal, dengan puncak pertama di 0,5–0,6 (kinerja rendah hingga sedang) dan puncak kedua di 0,8–0,9 (kinerja tinggi). Ini mengindikasikan bahwa turnover lebih sering terjadi pada karyawan dengan kinerja rendah atau sangat tinggi, mungkin karena kepuasan rendah pada yang pertama dan overwork pada yang kedua.

Grafik ini menunjukkan bahwa distribusi penilaian kinerja karyawan berbeda secara signifikan antara yang mengalami turnover dan yang tidak. Karyawan dengan penilaian sedang (0,6–0,8) cenderung bertahan, sedangkan yang rendah (< 0,6) atau sangat tinggi (> 0,9) lebih rentan terhadap turnover. Hasil ini mendukung penggunaan K-Means Clustering untuk segmentasi kinerja dan dapat menjadi dasar untuk strategi retensi karyawan.

#### e) INTERPRETASI DATA

Dengan semua informasi ini, berikut adalah hal-hal yang harus diketahui tentang perusahaan dan alasan mengapa karyawannya mungkin keluar:

- 1) Karyawan biasanya keluar ketika mereka kurang bekerja (kurang dari 150 jam/bulan atau 6 jam/hari).
- 2) Karyawan biasanya keluar ketika mereka sudah bekerja terlalu lama (lebih dari 250 jam/bulan atau 10 jam/hari).
- 3) Karyawan dengan evaluasi yang sangat tinggi atau rendah harus dipertimbangkan karena tingkat pergantian karyawannya yang tinggi.
- 4) Karyawan dengan gaji rendah hingga menengah merupakan bagian terbesar dari turnover karyawan.
- 5) Karyawan yang memiliki jumlah proyek 2, 6, atau 7 berisiko meninggalkan Perusahaan.
- 6) Kepuasan karyawan merupakan indikator tertinggi untuk pergantian karyawan.
- 7) Karyawan yang memiliki masa kerja 4 dan 5 tahun di Perusahaan harus dipertimbangkan karena tingkat turnover yang tinggi.

- 8) Kepuasan karyawan, tahun di perusahaan, dan evaluasi merupakan tiga faktor terbesar dalam menentukan pergantian karyawan.

## **MANFAAT PENERAPAN**

Penerapan K-Means Clustering dalam analisis kinerja karyawan memberikan keuntungan sebagai berikut:

- 1) Segmentasi yang Efisien: Mengklasifikasikan karyawan ke dalam kelompok berkinerja tinggi, sedang, atau rendah untuk strategi manajerial yang lebih terarah.
- 2) Pengambilan Keputusan Berdasarkan Data: Menyediakan pandangan objektif untuk promosi, pelatihan, atau penyesuaian beban kerja.
- 3) Optimalisasi Sumber Daya: Membantu organisasi dalam mendistribusikan sumber daya secara efektif berdasarkan ciri cluster.
- 4) Peningkatan Produktivitas: Memfasilitasi intervensi yang tepat untuk meningkatkan performa.

## **TANTANGAN DAN SOLUSI**

Tantangan utama dalam penerapan K-Means Clustering mencakup:

- 1) Menentukan Jumlah Cluster: Menetapkan nilai  $k$  yang ideal sering kali menjadi tugas yang sukar.  
Solusi: Mengaplikasikan metrik seperti DBI atau Skor Silhouette untuk menentukan jumlah cluster yang paling sesuai.
- 2) Sensitivitas terhadap Outlier: Adanya outlier dapat berdampak negatif pada hasil pengelompokan.  
Solusi: Melaksanakan normalisasi serta pembersihan data untuk mengecilkan pengaruh outlier.
- 3) Kualitas Data: Data yang tidak lengkap atau tidak akurat berpotensi mengurangi akurasi pengelompokan.  
Solusi: Melakukan proses pra-pemrosesan data dengan hati-hati, termasuk mengisi data yang hilang dan memverifikasi data.

## **KESIMPULAN**

Berdasarkan tinjauan penelitian yang dilakukan dalam rentang lima tahun terakhir (2020–2025), K-Means Clustering terbukti menjadi metode yang efektif untuk menganalisis kinerja karyawan dengan cara mengelompokkan mereka menurut karakteristik seperti produktivitas, kehadiran, dan kepuasan kerja. Metode ini memfasilitasi pengambilan keputusan yang berlandaskan data, seperti dalam menentukan program pelatihan, pengangkatan, atau distribusi sumber daya, serta berkontribusi untuk meningkatkan efisiensi operasional di dalam suatu organisasi. Dengan menerapkan metodologi CRISP-DM dan memanfaatkan alat seperti RapidMiner atau Python yang dilengkapi dengan pustaka scikit-learn, organisasi dapat menggunakan algoritma ini dengan efisien. Di masa mendatang, penggabungan K-Means Clustering dengan teknik machine learning yang lebih canggih, seperti deep learning, berpotensi untuk meningkatkan akurasi dan kelincahan dalam analisis kinerja karyawan. Klasifikasi dalam data mining dapat dimanfaatkan untuk mendapatkan informasi atau menyusun model dengan memanfaatkan data yang telah memiliki atribut dan kategori yang telah ditentukan sebelumnya.

## DAFTAR PUSTAKA

Budiarto, H., & Susanto, Y. (2020). Segmentasi pelanggan menggunakan K-Means clustering pada perusahaan asuransi. *Jurnal Sistem Informasi Bisnis*, 6(1), 45–52.

Nugroho, A., & Budiarto, H. (2023). Analisis kinerja karyawan berbasis KPI menggunakan K-Means clustering. *Jurnal Sistem Informasi*, 9(3), 78–85.

Pratama, A. D., Nugroho, A., & Susanto, Y. (2021). Analisis penggunaan algoritma K-Means clustering pada penilaian kinerja dosen. *Jurnal Sistem Informasi Bisnis*, 7(2), 101–110.

Penerapan algoritma K-Means clustering untuk analisis kinerja pengiriman paket Shopee Express di Hub Transit Kedawung. (2024). *Jurnal Teknologi Informasi*, 10(2), 120–130.

Sari, M., & Mulyadi, H. (2022). Penerapan K-Means clustering untuk segmentasi kinerja karyawan di perusahaan ritel. *Jurnal Sains Komputer dan Informatika*, 14(1), 55–63.

Agustina Srirahayu & Laras Setya Pribadie.(2023). Review Paper Data Mining Klasifikasi Data Mining. *Jurnal Ilmiah Informatika Global*, 14(1), 8-12.