

**TUGAS MANDIRI
FUNDAMENTALS OF DATA MINING**

**SISTEM ANALISIS KINERJA KARYAWAN BERBASIS PYHTON
DENGAN METODE DECISION TREE**



Nama : Aviva Nur

NPM : 231510065

Dosen : Erlin Elisa, S.Kom., M.Kom.

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK KOMPUTER
UNIVERSITAS PUTERA BATAM
2025**

KATA PENGANTAR

Laporan ini merupakan hasil dokumentasi dari proyek mandiri yang berjudul "Sistem Analisis Kinerja Karyawan Berbasis Python dengan Metode Decision Tree". Proyek ini bertujuan untuk mengeksplorasi bagaimana algoritma pembelajaran mesin dapat membantu departemen SDM dalam mengidentifikasi faktor-faktor kunci yang memengaruhi produktivitas kerja.

Pengerjaan proyek ini didasarkan pada dataset "Kinerja Karyawan" yang bersumber dari Kaggle (kontribusi oleh Ahmad Fawzan). Penggunaan metode Decision Tree dipilih karena kemampuannya dalam menyajikan alur logika keputusan yang transparan dan mudah dipahami, layaknya pola pikir manusia dalam mengambil kebijakan.

Penulis berharap dokumentasi ini tidak hanya menjadi arsip pribadi, tetapi juga menjadi bukti proses belajar dalam mengolah data mentah menjadi wawasan yang bernilai strategis. Kritik dan saran pengembangan sangat diharapkan untuk menyempurnakan analisis ini di masa depan.

Batam, 19 Desember 2025

AVIVA NUR

LISIS & HASIL PENGOLAHAN DATA (PYTHON + DATA MINING)

Sistem Analisis Kinerja Karyawan Berbasis Python dengan Metode Decision Tree

Deskripsi Dataset

- Sumber dataset : Kaggle.com <https://www.kaggle.com/datasets/ahmadfawzan/kinerja-karyawan>
- Jumlah record : 100 baris data
- Jumlah atribut : 6 kolom
- Tipe data

Kolom	Tipe Data	Jenis	Keterangan
ID	Integer	Identitas	Tidak digunakan untuk model
Disiplin	Integer	Numerik Ordinal	Fitur
Produktivitas	Integer	Numerik Ordinal	Fitur
Kerja Sama	Integer	Numerik Ordinal	Fitur
Inisiatif	Integer	Numerik Ordinal	Fitur
Kinerja Akhir	Kategorikal → Numerik (Encoding)	Label Klasifikasi	Target

- Target/label (jika supervised) : Dataset ini menggunakan atribut “Kinerja Akhir” sebagai label klasifikasi dalam proses supervised learning.
- Permasalahan yang ingin diselesaikan

✧ Contoh kalimat:

Permasalahan utama yang ingin diselesaikan dalam penelitian ini adalah memprediksi tingkat kinerja karyawan berdasarkan data penilaian disiplin kerja, produktivitas, kemampuan kerja sama, dan inisiatif. Dengan menggunakan algoritma Decision Tree, sistem ini bertujuan untuk membantu perusahaan dalam mengklasifikasikan kinerja karyawan secara otomatis, sehingga pengambilan keputusan terkait evaluasi, pelatihan, dan pengembangan sumber daya manusia dapat dilakukan dengan lebih cepat, objektif, dan akurat.

Persiapan Data & Preprocessing

Tahap persiapan data dan preprocessing dilakukan untuk memastikan bahwa dataset berada dalam kondisi siap digunakan dalam proses pemodelan klasifikasi menggunakan algoritma Decision Tree. Dataset yang digunakan terdiri dari 100 baris data dan 6 atribut, dengan empat atribut numerik ordinal sebagai fitur dan satu atribut sebagai label klasifikasi.

- Data Cleaning (Missing Value dan Outlier)

Berdasarkan hasil eksplorasi awal terhadap dataset, tidak ditemukan missing value pada seluruh atribut. Oleh karena itu, tidak diperlukan proses imputasi data. Setiap baris data memiliki nilai lengkap pada atribut Disiplin, Produktivitas, Kerja Sama, Inisiatif, dan Kinerja Akhir.

Pemeriksaan terhadap distribusi data menunjukkan adanya beberapa nilai ekstrem pada beberapa atribut penilaian, seperti nilai sangat rendah (1) atau sangat tinggi (5). Namun, nilai-nilai tersebut tidak dianggap sebagai outlier yang harus dihapus, karena masih berada dalam rentang penilaian yang wajar dan merepresentasikan kondisi nyata dalam evaluasi kinerja karyawan.

- Encoding Data Kategorikal

Dataset memiliki satu variabel kategorikal, yaitu Kinerja Akhir, yang berfungsi sebagai label klasifikasi. Agar dapat diproses oleh algoritma Decision Tree, variabel ini dikonversi ke dalam bentuk numerik menggunakan teknik Label Encoding.

Contoh proses encoding:

- “Baik” → 1
- “Cukup” → 2
- “Kurang” → 3

Proses encoding ini bertujuan untuk mengubah label berbentuk teks menjadi representasi numerik tanpa mengubah makna kategorinya.

- Normalisasi / Scaling

Seluruh fitur input pada dataset memiliki skala nilai yang sama, yaitu 1–5, dan bersifat numerik ordinal. Selain itu, algoritma Decision Tree tidak bergantung pada skala data, karena proses pemisahan node didasarkan pada aturan dan threshold, bukan pada jarak numerik.

Oleh karena itu, normalisasi atau scaling data tidak diwajibkan dan tidak memberikan pengaruh signifikan terhadap kinerja model. Dataset digunakan dalam bentuk aslinya tanpa proses scaling tambahan agar tetap mencerminkan nilai penilaian kinerja yang sesungguhnya.

- Feature Selectio

Proses pemilihan fitur dilakukan untuk menentukan atribut mana yang relevan dalam proses pemodelan. Dari enam atribut yang tersedia, kolom ID dihapus karena hanya berfungsi sebagai identitas dan tidak memiliki pengaruh terhadap penentuan kinerja karyawan.

Empat atribut yang digunakan sebagai fitur input adalah:

- Disiplin
- Produktivitas
- Kerja Sama
- Inisiatif

Atribut Kinerja Akhir digunakan sebagai target klasifikasi. Tidak dilakukan feature engineering tambahan karena seluruh fitur yang digunakan sudah merepresentasikan indikator utama dalam penilaian kinerja karyawan.

- Pembagian Dataset (Train–Test Split)

Dataset dibagi menjadi dua bagian menggunakan metode train–test split dengan proporsi 80% data latih **dan** 20% data uji. Pembagian ini dilakukan untuk mengevaluasi kemampuan model dalam memprediksi data yang belum pernah dilihat sebelumnya.

Rincian pembagian dataset:

- 80 data untuk pelatihan (*training*)
- 20 data untuk pengujian (*testing*)

Proses pembagian dilakukan dengan nilai `random_state` tetap untuk memastikan hasil eksperimen dapat direplikasi.

✂ Sertakan tabel ringkasan:

- Tabel Ringkasan Preprocessing Data

Keterangan	Sebelum Preprocessing	Sesudah Preprocessing
Total Data	100 baris	100 baris
Missing Value	Tidak ada	Tidak ada
Outlier	Ada sedikit	Dalam batas wajar (tidak dihapus)
Kolom Kategorikal	1 kolom (Kinerja Akhir)	Sudah diencoding
Fitur Tidak Relevan	Kolom ID	Dihapus

Skala Nilai	1–5	Digunakan tanpa scaling
-------------	-----	-------------------------

- Tabel Distribusi Data Train dan Test

Dataset	Jumlah Data	Persentase
Train	80	80%
Test	20	20%

Analisis Statistik & Visualisasi

- Statistik deskriptif dataset

Berdasarkan output fungsi `describe()`, diperoleh ringkasan statistik berupa nilai rata-rata (mean), median (50% quantile), nilai minimum, maksimum, serta kuartil untuk empat fitur numerik ordinal, yaitu Disiplin, Produktivitas, Kerja Sama, dan Inisiatif.

Hasil statistik menunjukkan bahwa rata-rata nilai setiap fitur berada di kisaran tengah skala penilaian (1–5), yaitu sekitar nilai 3. Distribusi kuartil juga menunjukkan bahwa sebagian besar data berada pada rentang nilai 2 hingga 4, yang mengindikasikan bahwa nilai ekstrem (sangat rendah atau sangat tinggi) relatif sedikit.

Temuan ini menunjukkan bahwa distribusi data cukup seimbang dan tidak didominasi oleh nilai ekstrem, sehingga dataset berada dalam kondisi yang baik untuk digunakan dalam proses pemodelan klasifikasi. Kondisi ini penting agar model tidak terlalu bias terhadap outlier dan mampu mempelajari pola data secara lebih stabil.

- Distribusi target/label

Grafik distribusi label “Kinerja Akhir” menunjukkan bahwa terdapat ketidakseimbangan jumlah data antar kelas, di mana satu kelas (misalnya kategori *Baik* atau kelas mayoritas) memiliki jumlah data lebih banyak dibandingkan kelas lainnya (seperti *Cukup* atau *Kurang*).

Ketidakseimbangan distribusi label ini perlu diperhatikan karena dapat memengaruhi performa model klasifikasi. Model cenderung lebih mudah mempelajari pola dari kelas mayoritas dan berpotensi mengabaikan kelas minoritas apabila tidak dilakukan penanganan khusus.

Meskipun demikian, pada penelitian ini tidak dilakukan teknik penyeimbangan kelas, sehingga kondisi distribusi data dibiarkan apa adanya agar mencerminkan kondisi nyata dalam penilaian kinerja karyawan.

- Korelasi antar fitur (heatmap)

Hasil matriks korelasi memperlihatkan:

- Terdapat korelasi positif cukup kuat antara fitur Produktivitas dan Inisiatif → menandakan karyawan yang inisiatif cenderung lebih produktif.
- Korelasi antara Disiplin dengan fitur lain relatif moderat → disiplin memberikan kontribusi, tapi tidak terlalu dominan dibanding aspek produktivitas atau inisiatif.
- Korelasi antar sebagian fitur rendah → artinya tiap fitur membawa informasi berbeda dan tidak redundant. Ini bagus untuk model klasifikasi karena variasi fitur tetap menjaga keunikan informasi.

Kondisi ini sangat menguntungkan untuk proses klasifikasi karena variasi informasi antar fitur tetap terjaga dan membantu model dalam membentuk aturan keputusan yang lebih baik.

- Visualisasi pendukung (histogram, boxplot, pairplot)

Histogram tiap fitur menunjukkan distribusi yang relatif merata, tanpa lonjakan ekstrem pada satu nilai saja. Misalnya, distribusi “Kerja Sama” mungkin agak condong ke nilai tengah (3–4), menunjukkan sebagian besar staf memiliki kerja sama rata-rata.

Boxplot memperlihatkan sedikit outlier — beberapa karyawan mungkin memiliki skor sangat rendah atau sangat tinggi pada fitur tertentu. Namun jumlah outlier sedikit dan tidak mendominasi, sehingga tidak perlu dihapus secara otomatis.

Hasil visualisasi ini menunjukkan bahwa data dalam kondisi bersih, distribusi wajar, dan memiliki variasi yang cukup, sehingga layak digunakan dalam analisis klasifikasi.

- Hubungan Antar Fitur & Label (Pairplot)

Berdasarkan hasil pairplot antara fitur dan label Kinerja Akhir, diperoleh beberapa pola sebagai berikut:

- Data dengan label Kinerja Akhir tinggi (misalnya *Baik*) cenderung berada pada area dengan nilai Produktivitas dan Inisiatif yang tinggi, serta Disiplin dan Kerja Sama pada tingkat menengah hingga tinggi.
- Data dengan label Kinerja Akhir rendah (misalnya *Kurang*) lebih banyak berada pada area dengan nilai rendah pada sebagian besar fitur.

- Tidak terdapat satu fitur tunggal yang sepenuhnya menentukan label, melainkan kombinasi nilai dari seluruh fitur yang memengaruhi hasil klasifikasi.

Temuan ini menunjukkan bahwa penilaian kinerja karyawan bersifat multidimensi, sehingga algoritma Decision Tree sangat sesuai digunakan karena mampu menangkap pola keputusan berdasarkan kombinasi beberapa atribut sekaligus.

✧ Sertakan insight dari grafik, bukan sekadar menampilkan.

Visualisasi Decision Tree menampilkan bagaimana model melakukan proses klasifikasi Kinerja Akhir karyawan berdasarkan empat atribut utama, yaitu Disiplin, Produktivitas, Kerja Sama, dan Inisiatif. Dari struktur pohon keputusan yang dihasilkan, diperoleh beberapa insight penting terkait cara model mengambil keputusan.

1. Fitur Paling Berpengaruh dalam Pemisahan Awal

Dari struktur pohon terlihat bahwa model memulai pemisahan (split awal) menggunakan fitur seperti Produktivitas dan Disiplin. Hal ini menunjukkan bahwa kedua fitur tersebut dianggap sebagai indikator paling kuat dalam membedakan kategori kinerja karyawan.

Insight:

Produktivitas dan disiplin memiliki peran dominan dalam menentukan kinerja akhir karyawan dibandingkan fitur lainnya.

2. Nilai Gini Semakin Rendah di Level Bawah Pohon

Pada node-node di bagian bawah, banyak terlihat nilai:

- $gini = 0.0$
- $samples =$ jumlah sedikit
- $class =$ kategori tertentu

Hal ini menandakan bahwa kondisi di node tersebut sudah sangat homogen.

Insight:

Model mampu menemukan kelompok karyawan dengan karakteristik yang sangat jelas, sehingga prediksinya lebih pasti.

3. Pohon yang Kompleks Menunjukkan Variasi Tinggi dalam Data

Struktur pohon keputusan tampak bercabang cukup banyak dengan beberapa node kecil. Hal ini menunjukkan bahwa karakteristik karyawan dalam dataset cukup beragam, sehingga model membentuk aturan yang relatif detail untuk setiap variasi data.

Insight:

Dataset memiliki distribusi yang tidak terlalu sederhana, dan model berusaha mempelajari pola spesifik dari kombinasi nilai fitur yang berbeda.

4. Kombinasi Fitur Menentukan Keputusan Akhir

Model tidak hanya menggunakan satu atribut dalam menentukan hasil klasifikasi, melainkan menggunakan kombinasi beberapa fitur secara bertahap, seperti:

- tingkat Disiplin,
- nilai Produktivitas,
- Kerja Sama,
- serta Inisiatif.

Keputusan akhir baru ditentukan setelah beberapa aturan terpenuhi secara berurutan.

Insight:

Prediksi kinerja karyawan sangat dipengaruhi oleh kombinasi nilai dari seluruh fitur, bukan oleh satu faktor tunggal.

5. Indikasi Overfitting Ringan

Dari visualisasi pohon keputusan, terlihat adanya beberapa node dengan jumlah sampel yang sangat kecil (misalnya 1–2 data). Hal ini mengindikasikan bahwa model cenderung mempelajari data training secara sangat detail.

Insight:

Model Decision Tree berpotensi mengalami overfitting ringan, sehingga diperlukan pengaturan parameter seperti `max_depth`, `min_samples_split`, atau `min_samples_leaf` untuk meningkatkan kemampuan generalisasi model.

6. Jalur Keputusan Transparan dan Mudah Dipahami

Setiap node dalam pohon keputusan menampilkan informasi berupa aturan keputusan, nilai gini, jumlah sampel, distribusi kelas, serta hasil prediksi. Hal ini memudahkan penelusuran jalur keputusan dari awal hingga akhir.

Insight:

Decision Tree sangat sesuai untuk analisis kinerja karyawan karena memberikan penjelasan yang jelas dan mudah dipahami oleh pihak HRD atau manajemen.

7. Pola Kinerja Sesuai Logika Penilaian di Perusahaan

Pola keputusan yang terbentuk pada pohon menunjukkan bahwa karyawan dengan nilai Disiplin, Produktivitas, Kerja Sama, dan Inisiatif yang lebih tinggi cenderung diklasifikasikan ke dalam kategori kinerja yang lebih baik. Sebaliknya, nilai rendah pada sebagian besar fitur mengarah pada kategori kinerja yang lebih rendah.

Insight:

Model mampu menangkap pola penilaian kinerja yang secara praktis sesuai dengan proses evaluasi karyawan di dunia nyata.

Pemilihan dan Penerapan Algoritma

- Nama algoritma

Algoritma yang digunakan dalam penelitian ini adalah Decision Tree Classifier yang diimplementasikan menggunakan library scikit-learn. Model ini menggunakan kriteria Entropy dalam proses pemilihan atribut terbaik pada setiap node, sehingga pendekatan yang digunakan serupa dengan konsep algoritma Decision Tree C4.5, meskipun bukan merupakan implementasi C4.5 secara murni.

- Alasan pemilihan

Pemilihan algoritma Decision Tree C4.5 didasarkan pada beberapa pertimbangan berikut:

1. Sesuai untuk Data Klasifikasi

Dataset penelitian memiliki target berupa kategori “Kinerja Akhir” (Baik, Cukup, Kurang), sehingga algoritma klasifikasi seperti C4.5 sangat tepat digunakan.

2. Mampu Mengolah Data Numerik dan Kategorikal

Fitur input berupa nilai numerik ordinal (skala 1–5), sedangkan label bersifat kategorikal. C4.5 mampu menangani kedua jenis data tanpa perlu konversi kompleks.

3. Cocok untuk Data Non-Linear

Hubungan antar fitur tidak bersifat linear. Pohon keputusan efektif untuk data non-linear karena membentuk aturan keputusan yang bercabang.

4. Model Mudah Dipahami dan Dijelaskan

Decision Tree menghasilkan visualisasi pohon yang transparan, sehingga memudahkan interpretasi oleh HRD atau manajemen terkait alasan setiap keputusan klasifikasi.

5. Mendukung Pruning untuk Mengurangi Overfitting

Algoritma C4.5 memiliki mekanisme pemangkasan (pruning) yang berguna untuk meningkatkan generalisasi model dalam memprediksi data baru.

- Parameter utama yang digunakan

Implementasi Decision Tree dalam penelitian ini dilakukan menggunakan fungsi DecisionTreeClassifier dari library scikit-learn, dengan parameter utama sebagai berikut:

Parameter	Nilai	Penjelasan
criterion	"entropy"	Menggunakan perhitungan Entropy untuk pemilihan atribut, sesuai karakteristik algoritma C4.5.
splitter	"best"	Memilih pemisahan terbaik di setiap node.
max_depth	None (default)	Pohon dibiarkan tumbuh secara penuh agar model dapat mempelajari pola secara maksimal.
min_samples_split	2	Jumlah minimum sampel untuk memecah node.
min_samples_leaf	1	Jumlah minimum sampel pada node daun.
random_state	42	Menghasilkan model yang konsisten dan dapat direplikasi.

✎ Daftarkan algoritma yang diuji

Algoritma	Library Python	Tujuan Penggunaan
Decision Tree (Entropy-based)	sklearn.tree.DecisionTreeClassifier	Mengklasifikasikan kinerja karyawan berdasarkan atribut Disiplin, Produktivitas, Kerja Sama, dan Inisiatif dengan model yang mudah diinterpretasikan.

Pengujian dan Evaluasi Model

Tabel tersebut mengklasifikasikan tipe masalah machine learning (jenis tugas) dan metode evaluasi yang sesuai untuk tiap tipe. Berikut penjelasan tiap baris:

Jenis Tugas	Metrics Evaluasi	Artinya / Kapan Digunakan
Klasifikasi	Accuracy, Precision, Recall, F1-Score, Confusion Matrix, ROC-AUC	Digunakan ketika target/model menghasilkan kategori / label (misalnya “Baik”, “Cukup”, “Kurang”). Evaluasi ini mengukur seberapa baik model memprediksi kelas yang benar dibanding kelas yang salah.

Regresi	MAE, MSE, RMSE, R^2	Digunakan ketika target otomatis berupa nilai kontinyu (angka), bukan kategori. Metrik ini mengukur kesalahan prediksi model terhadap nilai sesungguhnya.
Clustering	Silhouette Score, Inertia, Davies–Bouldin (dan bisa juga Calinski-Harabasz)	Digunakan untuk evaluasi hasil <i>unsupervised clustering</i> ketika tidak ada label ground-truth. Metrik ini menilai seberapa baik kluster terpisah dan seberapa kompak kluster internal.

🔗 Sertakan tabel perbandingan hasil:

- Perbandingan Metrik per Kelas

Kelas	Precision	Recall	F1-Score	Support
1	0.00	0.00	0.00	3
2	0.25	0.33	0.29	3
3	1.00	0.33	0.50	6
4	0.29	0.25	0.27	8

- Perbandingan Rata-Rata Kinerja Model

Jenis Rata-rata	Precision	Recall	F1-Score	Total Data
Macro Average	0.38	0.23	0.26	20
Weighted Average	0.45	0.25	0.30	20
Accuracy	–	–	0.25	20

- Ringkasan Confusion Matrix

Kelas Aktual ↓ / Prediksi →	1	2	3	4
1	0	0	0	3
2	1	1	0	1
3	1	2	2	1
4	5	1	0	2

Analisis & Interpretasi Hasil

Berikan jawaban dari analisis data mining:

- Algoritma mana yang paling optimal? Kenapa?

Dalam penelitian ini, algoritma yang digunakan adalah Decision Tree berbasis entropy. Berdasarkan hasil evaluasi menggunakan metrik Accuracy, Precision, Recall, dan F1-Score, diperoleh nilai akurasi sebesar **25%**, yang menunjukkan bahwa performa model klasifikasi masih tergolong rendah.

Meskipun demikian, pemilihan Decision Tree tetap relevan karena algoritma ini sesuai dengan karakteristik dataset yang memiliki fitur numerik ordinal dan hubungan non-linear antar atribut. Selain itu, struktur pohon keputusan memberikan kelebihan dari sisi interpretabilitas, meskipun performa prediksi belum optimal.

- Fitur apa yang paling berpengaruh?

Berdasarkan visualisasi pohon keputusan, dapat diidentifikasi bahwa:

- Disiplin dan Produktivitas sering muncul pada level awal percabangan, yang menunjukkan bahwa kedua fitur tersebut memiliki pengaruh relatif lebih besar dalam menentukan Kinerja Akhir.
- Kerja Sama dan Inisiatif berperan sebagai fitur pendukung yang muncul pada node-level lebih dalam untuk memperjelas klasifikasi.

Interpretasi ini sesuai dengan logika penilaian SDM, di mana karyawan dengan disiplin dan produktivitas tinggi cenderung memiliki kinerja yang lebih baik.

- Apakah model sudah baik? Apa kekurangannya?

Kelebihan Model:

- Model Decision Tree mudah diinterpretasikan karena struktur pohonnya transparan.
- Model mampu menangkap hubungan non-linear antar fitur penilaian kinerja.
- Dapat digunakan untuk menggali insight faktor dominan dalam penilaian kinerja karyawan.

Kekurangan Model:

- Nilai akurasi dan metrik evaluasi lainnya masih rendah, sehingga kemampuan prediksi model belum optimal.
- Beberapa kelas tidak terprediksi dengan baik, terlihat dari nilai precision dan recall yang mendekati nol.
- Model sensitif terhadap variasi data dan berpotensi mengalami overfitting.

- Apakah overfitting/underfitting terjadi?

Berdasarkan struktur pohon keputusan dan hasil evaluasi model, terdapat indikasi overfitting ringan. Hal ini disebabkan oleh:

- Jumlah dataset yang relatif kecil (100 data).
- Struktur pohon yang cukup kompleks dengan banyak node dan cabang detail.
- Adanya node dengan jumlah sampel yang sangat kecil.

Namun, tidak terdapat indikasi underfitting, karena model masih mampu mempelajari pola dari data training meskipun generalisasinya rendah. Risiko overfitting dapat dikurangi dengan pengaturan parameter seperti `max_depth`, `min_samples_split`, dan `min_samples_leaf`.

- Insight terhadap domain dataset

Dari keseluruhan analisis, diperoleh beberapa insight penting terkait domain penilaian kinerja karyawan, antara lain:

- Disiplin dan Produktivitas merupakan indikator paling berpengaruh dalam menentukan kinerja karyawan.
- Kerja Sama dan Inisiatif berperan sebagai faktor tambahan yang memperkuat hasil klasifikasi.
- Model Decision Tree dapat membantu pihak HRD dalam:
 - mengidentifikasi karyawan dengan potensi kinerja rendah,
 - menentukan kebutuhan pelatihan,
 - melakukan intervensi berbasis data secara lebih objektif.

Insight ini menunjukkan bahwa meskipun performa prediksi model masih perlu ditingkatkan, pendekatan data mining berbasis Decision Tree tetap memiliki nilai praktis dalam mendukung pengambilan keputusan berbasis data (data-driven HR).

Contoh narasi

Berdasarkan hasil pengujian model, algoritma Decision Tree berbasis entropy menghasilkan nilai akurasi sebesar 25% dalam mengklasifikasikan data kinerja karyawan. Hasil ini menunjukkan bahwa performa prediksi model masih tergolong rendah, sehingga model belum mampu melakukan klasifikasi dengan tingkat ketepatan yang optimal pada data uji.

Nilai Precision, Recall, dan F1-Score yang bervariasi antar kelas menunjukkan bahwa model mengalami kesulitan dalam membedakan beberapa kategori kinerja, terutama pada kelas dengan jumlah data yang relatif sedikit. Kondisi ini dipengaruhi oleh keterbatasan jumlah dataset, distribusi kelas yang tidak seimbang, serta kompleksitas pola data yang dipelajari oleh model.

Meskipun demikian, penggunaan algoritma Decision Tree tetap memberikan nilai tambah dari sisi interpretabilitas. Model mampu membentuk aturan keputusan yang menggambarkan hubungan antara variabel penilaian kinerja, seperti Disiplin, Produktivitas, Kerja Sama, dan Inisiatif, meskipun hubungan tersebut belum sepenuhnya terkonversi menjadi prediksi yang akurat.

Hasil visualisasi pohon keputusan menunjukkan bahwa kombinasi nilai Produktivitas dan Inisiatif yang lebih tinggi cenderung mengarah pada kategori kinerja yang lebih baik, sedangkan nilai rendah pada sebagian besar fitur mengarah pada kategori kinerja yang lebih rendah. Dengan demikian, meskipun performa prediksi model masih perlu ditingkatkan, pendekatan Decision Tree tetap mampu memberikan insight yang relevan bagi pihak manajemen dalam memahami faktor-faktor utama yang memengaruhi penilaian kinerja karyawan secara lebih objektif.

Kesimpulan & Rekomendasi

- Kesimpulan

Penelitian ini bertujuan untuk memprediksi tingkat kinerja karyawan berdasarkan data penilaian Disiplin, Produktivitas, Kerja Sama, dan Inisiatif menggunakan pendekatan data mining berbasis algoritma Decision Tree.

Berdasarkan hasil eksperimen dan evaluasi model, algoritma Decision Tree berbasis entropy mampu mempelajari pola hubungan antar fitur penilaian kinerja, namun performa prediksi yang dihasilkan masih tergolong rendah, dengan nilai akurasi sebesar 25%. Hal ini menunjukkan bahwa model belum mampu melakukan klasifikasi kinerja karyawan secara optimal pada data uji.

Meskipun demikian, struktur pohon keputusan yang dihasilkan bersifat transparan dan mudah diinterpretasikan, sehingga proses pengambilan keputusan tidak bersifat *black-box*. Model dapat memberikan gambaran faktor-faktor dominan yang memengaruhi penilaian kinerja karyawan, yang bermanfaat sebagai **alat bantu analisis**, bukan sebagai sistem prediksi akhir yang siap digunakan.

Dengan demikian, tujuan penelitian tercapai dari sisi eksplorasi pola data dan interpretasi, namun masih diperlukan pengembangan lebih lanjut agar model memiliki performa prediksi yang lebih baik.

- Link repository (GitHub/Drive/Colab) :

- 1) Colab:

- https://colab.research.google.com/drive/1_In4f_7DYBYX9DanBKy0UsrBDmeH3ldN#scrollTo=97862ed1

