

PROJECT MACHINE LEARNING

Kelompok 5:

Amelinda Renjani (20220040163)

Rahayu Amaliyah Kamis (20220040248)

Selpia Meilani (20220040148)

Rayhansyah (20220040283)

Ayang Marliani F (20220040246)

Syachra Indyra Puteri (20220040282)

K-Nearest Neighbor (KNN) adalah salah satu algoritma machine learning yang digunakan untuk klasifikasi dan regresi. KNN adalah metode berbasis instance (instance-based learning) di mana klasifikasi sebuah data baru ditentukan berdasarkan kedekatannya dengan sejumlah $\backslash(k\backslash)$ tetangga terdekat dalam data pelatihan.



Konsep Dasar KNN

Definisi KNN

-KNN adalah algoritma sederhana yang menyimpan semua kasus yang tersedia dan mengklasifikasikan kasus baru berdasarkan kemiripan.

- Dalam klasifikasi, sebuah objek diklasifikasikan dengan mayoritas suara dari tetangganya.
- Dalam regresi, output adalah nilai rata-rata dari tetangganya.

Parameter Penting

-K (jumlah tetangga) Jumlah tetangga terdekat yang dipertimbangkan untuk menentukan kelas atau nilai objek baru.

-Metrik Jarak Cara mengukur kedekatan antara titik data. Umumnya menggunakan Euclidean Distance, tetapi bisa juga menggunakan Manhattan Distance, Minkowski Distance, dll.

Algoritma KNN

1. Menentukan parameter $\backslash(k\backslash)$
2. Menghitung jarak antara data baru dengan semua data dalam data pelatihan menggunakan metrik jarak yang dipilih.
3. Mengurutkan data pelatihan berdasarkan jarak terdekat.
4. Memilih $\backslash(k\backslash)$ tetangga terdekat
5. Klasifikasi (untuk tugas klasifikasi): Menentukan kelas dengan memilih kelas mayoritas dari $\backslash(k\backslash)$ tetangga terdekat.
6. Regresi (untuk tugas regresi) Menghitung nilai rata-rata dari $\backslash(k\backslash)$ tetangga terdekat.



Optimasi dan Variasi KNN

01

Scaling Fitur: Gunakan normalisasi atau standardisasi fitur untuk memastikan semua fitur berkontribusi secara proporsional terhadap jarak

02

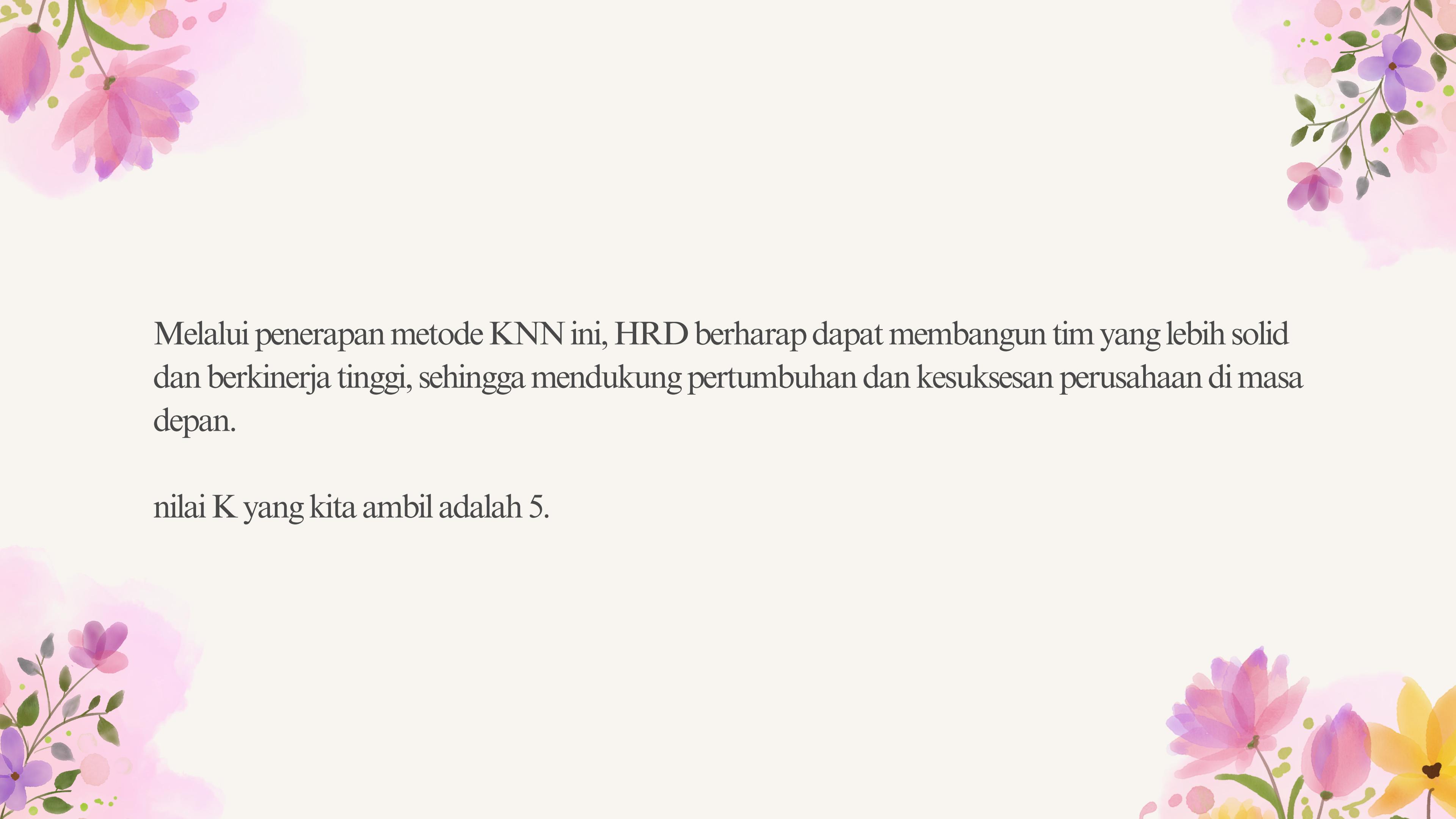
KNN dengan Pemberian Bobot: Memberikan bobot lebih pada tetangga yang lebih dekat.

03

Cross-validation: Untuk menentukan nilai $\backslash(k \backslash)$ yang optimal dengan membagi dataset menjadi subset pelatihan dan validasi.

Contoh studikasus :

Seorang HRD di sebuah perusahaan besar sedang dalam proses perekrutan karyawan baru untuk mengisi berbagai posisi penting. Oleh karena itu, HRD memutuskan untuk mengimplementasikan metode yang lebih objektif dan terstruktur untuk mengevaluasi kandidat menggunakan teknologi machine learning, yaitu metode K-Nearest Neighbors (KNN). Sebagai langkah awal, HRD mulai mengumpulkan data dari hasil penilaian keseluruhan proses perekrutan karyawan sebelumnya. Data ini kemudian diolah menjadi sebuah dataset yang mencakup empat atribut utama yang dianggap paling berpengaruh dalam menentukan kualitas seorang kandidat: Usia (X_1), IPK (X_2), nilai wawancara (X_3), Pengalaman kerja (X_4).



Melalui penerapan metode KNN ini, HRD berharap dapat membangun tim yang lebih solid dan berkinerja tinggi, sehingga mendukung pertumbuhan dan kesuksesan perusahaan di masa depan.

nilai K yang kita ambil adalah 5.

No	Usia	IPK	Skor Wawancara	Pengalaman Kerja	Y
1	22	3.8	80	0	Lolos
2	25	3.5	75	1 tahun	Lolos
3	27	3.2	65	2 tahun	Tidak Lolos
4	23	3.9	85	0	Lolos
5	26	3.6	70	1 tahun	Lolos
6	28	3.3	60	2 tahun	Tidak Lolos
7	24	3.8	82	0	Lolos
8	21	3.4	72	0	Tidak Lolos
9	29	3.1	68	2 tahun	Tidak Lolos
10	25	3.7	77	1 tahun	Lolos
11	30	3.9	84	1 tahun	?

Kesimpulan

K-Nearest Neighbor adalah algoritma yang sederhana namun efektif untuk berbagai tugas klasifikasi dan regresi. Meski memiliki beberapa kelemahan, dengan penanganan yang tepat, KNN bisa menjadi alat yang kuat dalam toolkit machine learning Anda.

TERIMA
KASIH