

Praktikum 10

Klasifikasi KNN

Pokok/Sub Bahasan	a. <i>Klasifikasi KNN</i>
Alokasi Waktu	2 x 170 menit
Tempat	Laboratorium/virtual

1. Tujuan

- a. Mahasiswa dapat mengkonstruksi data

2. Alat dan Bahan

- a. Alat
 - 1) Laptop/Komputer dan kelengkapan standarnya
 - 2) Internet
 - 3) *Google Colaboratory*
- b. Bahan

- Data Tersedia

3. Teori Singkat

a. Pengertian Klasifikasi KNN

K-Nearest Neighbors (KNN) adalah algoritma klasifikasi yang bersifat non-parametrik dan termasuk dalam kategori supervised learning. KNN bekerja dengan cara mencari sejumlah tetangga terdekat (K) dari data baru, lalu mengklasifikasikan data tersebut berdasarkan mayoritas kelas dari tetangga-tetangganya.

Kelebihan:

- Sederhana dan mudah dipahami
- Tidak memerlukan proses pelatihan

Kekurangan:

- Waktu prediksi bisa lambat untuk dataset besar
- Sensitif terhadap skala data dan data yang tidak relevan

b. Rumus Klasifikasi KNN

Jarak antar data dihitung menggunakan rumus Euclidean Distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Dimana:

- x dan y adalah dua buah titik dalam ruang n -dimensi
- x_i dan y_i adalah fitur ke- i dari titik x dan y

c. Contoh Kasus Klasifikasi KNN

Misalkan kita memiliki data pelanggan berikut (fitur: umur dan saldo), serta label apakah mereka membeli produk deposito (y : ya/tidak):

No	Umur (x_1)	Saldo (x_2)	y
1	25	40	Tidak
2	30	60	Tidak
3	45	100	Ya
4	35	80	Ya
5	40	70	Ya

Sekarang kita ingin memprediksi **data baru** dengan fitur berikut:

Umur: 33, Saldo: 65

Gunakan **$K = 3$** (tiga tetangga terdekat), dan hitung secara manual jarak Euclidean-nya.

Hitung Jarak Euclidean ke Setiap Titik

$$d = \sqrt{(x_1 - x'_1)^2 + (x_2 - x'_2)^2}$$

Data baru: (33, 65)

No	Titik Data	Perhitungan	Jarak (d)
1	(25, 40)	$\sqrt{((33-25)^2 + (65-40)^2)} = \sqrt{(64 + 625)} = \sqrt{689}$	26.26
2	(30, 60)	$\sqrt{((33-30)^2 + (65-60)^2)} = \sqrt{(9 + 25)} = \sqrt{34}$	5.83
3	(45, 100)	$\sqrt{((33-45)^2 + (65-100)^2)} = \sqrt{(144 + 1225)} = \sqrt{1369}$	37.01
4	(35, 80)	$\sqrt{((33-35)^2 + (65-80)^2)} = \sqrt{(4 + 225)} = \sqrt{229}$	15.13
5	(40, 70)	$\sqrt{((33-40)^2 + (65-70)^2)} = \sqrt{(49 + 25)} = \sqrt{74}$	8.60

Pilih 3 Tetangga Terdekat

Dari jarak terdekat:

1. Data ke-2 (Jarak 5.83) → Tidak
2. Data ke-5 (Jarak 8.60) → Ya
3. Data ke-4 (Jarak 15.13) → Ya

Voting Mayoritas

Kelas Tetangga	Jumlah
Ya	2
Tidak	1

Maka prediksi untuk data (33, 65) adalah: *Ya*

4. Langkah Kerja

```
# 1. Import pustaka

import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.neighbors import KNeighborsClassifier
from sklearn.metrics import classification_report, confusion_matrix
import seaborn as sns
import matplotlib.pyplot as plt

# 1. Load data
df = pd.read_csv('bank-noisy-data.csv')

# 2. Isi missing value di 'balance' dengan median
df['balance'].fillna(df['balance'].median(), inplace=True)

# 3. Encode semua kolom kategorikal (object)
label_encoders = {}
for col in df.select_dtypes(include='object').columns:
    le = LabelEncoder()
    df[col] = le.fit_transform(df[col])
    label_encoders[col] = le

# 4. Pisahkan fitur dan target
X = df.drop('y', axis=1)
y = df['y'].values.ravel() # pastikan y adalah array 1 dimensi
```

```
# 5. Normalisasi fitur
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# 6. Split data
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)

# 7. Latih model KNN
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train, y_train)

# 8. Evaluasi
y_pred = knn.predict(X_test)

print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test, y_pred))

# 9. Visualisasi
plt.figure(figsize=(6,4))
sns.heatmap(confusion_matrix(y_test, y_pred), annot=True, fmt='d', cmap='Blues')
plt.title("Confusion Matrix")
plt.xlabel("Predicted")
plt.ylabel("Actual")
plt.show()
```

```
# 10. Cari nilai K terbaik (optional)
error_rate = []
for k in range(1, 21):
    knn_k = KNeighborsClassifier(n_neighbors=k)
    knn_k.fit(X_train, y_train)
    pred_k = knn_k.predict(X_test)
    error_rate.append(np.mean(pred_k != y_test))

plt.figure(figsize=(10,5))
plt.plot(range(1, 21), error_rate, marker='o', linestyle='--')
plt.title("Error Rate vs K Value")
plt.xlabel("K")
plt.ylabel("Error Rate")
plt.grid(True)
plt.show()
```

5. Rubrik Ketercapaian Praktikum

No.	Kemampuan Akhir Pembelajaran	Metode Pembelajaran	Kriteria - Indikator
1.	Mahasiswa dapat membangun model menggunakan metode yang sesuai dan mengevaluasi hasil pemodelannya	Ceramah, Latihan, <i>Case Study</i> , Pembelajaran Kooperatif	<ol style="list-style-type: none"> 1. Mahasiswa dapat menggunakan tools untuk membangun model klasifikasi dan mengevaluasinya 2. Mahasiswa dapat menyelesaikan persoalan tentang membangun model klasifikasi dan mengevaluasinya 3. Mahasiswa dapat Memahami kesesuaian metode dengan kasus

6. Tugas

Ikutilah setiap langkah yang telah dipelajari untuk data berikut baik secara manual dan secara program (pyhton). Selanjutnya dokumentasikan setiap pekerjaan anda :

Sebuah perusahaan ingin memprediksi apakah calon pelanggan akan membeli produk asuransi berdasarkan dua faktor utama: **umur** dan **penghasilan bulanan**. Data historis pelanggan sebelumnya telah dikumpulkan dan diberi label apakah mereka membeli produk atau tidak. Tugas mahasiswa adalah menggunakan algoritma **K-Nearest Neighbors (KNN)** untuk memprediksi keputusan seorang calon pelanggan baru berdasarkan data historis yang tersedia.

No	Umur (tahun)	Penghasilan (juta/bulan)	Beli Asuransi
1	25	4.0	Tidak
2	30	6.0	Tidak
3	45	10.0	Ya
4	35	8.0	Ya
5	40	7.0	Ya

Calon pelanggan baru:

- **Umur:** 33 tahun
- **Penghasilan:** 6.5 juta/bulan

7. Referensi

1. J. Han, J. Pei, and M. Kamber, Data Mining: Concepts and Techniques, 4th ed. San Francisco, CA, USA: Morgan Kaufmann, 2022.
2. C. C. Aggarwal, Data Mining: The Textbook. Cham, Switzerland: Springer, 2015.
3. Keputusan Menteri Ketenagakerjaan Republik Indonesia Nomor 299 Tahun 2020 Tentang Penetapan Standar Kompetensi Kerja Nasional Indonesia Kategori Informasi dan Komunikasi Golongan Pokok Aktivitas Pemrograman, Konsultasi Komputer dan Kegiatan Yang Berhubungan Dengan Itu (YBDI) Bidang Keahlian *Artificial Intelligence* Subbidang *Data Science*