

11422037_RobertoSamuelButar
butar-1740577460420
by Turnitin™

Submission date: 26-Feb-2025 02:45PM (UTC+0100)

Submission ID: 2599269365

File name: 11422037_RobertoSamuelButarbutar-1740577460420.docx (208.42K)

Word count: 756

Character count: 5073

Laporan Praktikum

4143203 – Data Science

Two Dimensional Data



11422037 Roberto Samuel Butarbutar

ST Teknologi Rekayasa Perangkat Lunak

Daftar Isi

1. Pendahuluan	3
2. Representasi Two-Dimensional Data	3
3. Operasi pada Two-Dimensional Data	3
4. Implementasi Two-Dimensional Data dalam Python	4
5. Visualisasi Two-Dimensional Data	6
6. Aplikasi dan Tantangan dalam Pengolahan Two-Dimensional Data	6
6.1. Aplikasi Two-Dimensional Data	6
6.2. Tantangan dalam Pengolahan Two-Dimensional Data:	6
7. Kesimpulan	7
Referensi	7

1. Pendahuluan

Two-dimensional data adalah jenis data yang disusun dalam bentuk tabel atau matriks, di mana setiap baris mewakili satu entitas atau sampel, dan setiap kolom merepresentasikan atribut atau fitur dari entitas tersebut. Struktur ini sangat umum dalam berbagai bidang, termasuk statistik, machine learning, dan analisis data.

Format two-dimensional data memungkinkan penyimpanan, pemrosesan, dan analisis data yang lebih efisien. Dengan adanya struktur yang jelas dalam bentuk baris dan kolom, data dapat dengan mudah diorganisir, dimanipulasi, dan dianalisis untuk berbagai keperluan. Data dua dimensi sering digunakan dalam berbagai industri seperti keuangan, kesehatan, pendidikan, dan teknologi untuk membantu pengambilan keputusan berdasarkan data yang tersedia.

Dalam dunia machine learning, data dua dimensi adalah bentuk dasar yang digunakan dalam banyak algoritma pembelajaran mesin. Model prediktif dan klasifikasi sering kali menggunakan data yang disajikan dalam bentuk matriks untuk melakukan analisis dan memberikan hasil yang lebih akurat. Oleh karena itu, pemahaman mengenai two-dimensional data menjadi hal yang sangat penting bagi para praktisi data dan peneliti yang bekerja di bidang data science dan kecerdasan buatan.

Selain itu, data dua dimensi juga mempermudah dalam proses eksplorasi dan visualisasi data. Dengan struktur yang terorganisir, berbagai teknik analisis data seperti pengelompokan, regresi, dan analisis statistik dapat diterapkan dengan lebih mudah. Dalam database relasional, data juga disimpan dalam bentuk tabel yang memiliki hubungan antar entitas, sehingga memungkinkan integrasi dan pengolahan data yang lebih luas.

2. Representasi Two-Dimensional Data

Data dua dimensi dapat direpresentasikan dalam berbagai format, antara lain:

- **Tabel:** Format standar dalam spreadsheet seperti Microsoft Excel atau Google Sheets.
- **Matriks:** Struktur dalam komputasi numerik, seperti dalam NumPy.
- **DataFrame:** Representasi yang digunakan dalam Pandas untuk analisis data.

3. Operasi pada Two-Dimensional Data

Beberapa operasi umum yang dilakukan pada data dua dimensi meliputi:

- **Seleksi data:** Memilih baris atau kolom tertentu untuk dianalisis lebih lanjut.
- **Filtering:** Menyaring data berdasarkan kondisi tertentu.
- **Transformasi data:** Mengubah struktur atau skala data agar lebih sesuai dengan analisis.
- **Agregasi data:** Menggunakan fungsi seperti rata-rata, median, dan total untuk mengidentifikasi pola dalam data.
- **Pembersihan data:** Menghapus atau mengisi nilai yang hilang serta menghapus duplikasi data.

Dengan memahami operasi-operasi ini, kita dapat mengelola data dua dimensi dengan lebih efisien untuk analisis yang lebih akurat.

4. Implementasi Two-Dimensional Data dalam Python

Dua pustaka utama dalam Python yang digunakan untuk menangani two-dimensional data adalah Pandas dan NumPy.

Berikut adalah contoh kode yang menggambarkan konsep two-dimensional data :

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Membuat DataFrame dengan data contoh
data = {
    'Nama': ['Ali', 'Budi', 'Citra', 'Dina', 'Eka'],
    'Usia': [25, 30, 27, 22, 35],
    'Gaji': [5000, 6000, 5500, 4500, 7000],
    'Departemen': ['IT', 'HR', 'IT', 'Finance', 'HR']
}
df = pd.DataFrame(data)

# Menampilkan DataFrame
print("DataFrame awal:")
print(df)

# Menambahkan kolom baru berdasarkan logika tertentu
df['Gaji_Tinggi'] = df['Gaji'] > 5500

# Menampilkan statistik deskriptif
print("\nStatistik Deskriptif:")
print(df.describe())

# Menampilkan jumlah karyawan per departemen
print("\nJumlah Karyawan per Departemen:")
print(df['Departemen'].value_counts())

# Visualisasi hubungan antara usia dan gaji
plt.figure(figsize=(8,5))
sns.scatterplot(x=df['Usia'], y=df['Gaji'], hue=df['Departemen'], s=100)
plt.title("Hubungan Usia dan Gaji")
plt.xlabel("Usia")
plt.ylabel("Gaji")
plt.show()
```

Output :

DataFrame awal:

	Nama	Usia	Gaji	Departemen
0	Ali	25	5000	IT
1	Budi	30	6000	HR
2	Citra	27	5500	IT
3	Dina	22	4500	Finance
4	Eka	35	7000	HR

Statistik Deskriptif:

	Usia	Gaji
count	5.000000	5.000000
mean	27.800000	5600.000000
std	4.969909	961.769203
min	22.000000	4500.000000
25%	25.000000	5000.000000
50%	27.000000	5500.000000
75%	30.000000	6000.000000
max	35.000000	7000.000000

Jumlah Karyawan per Departemen:

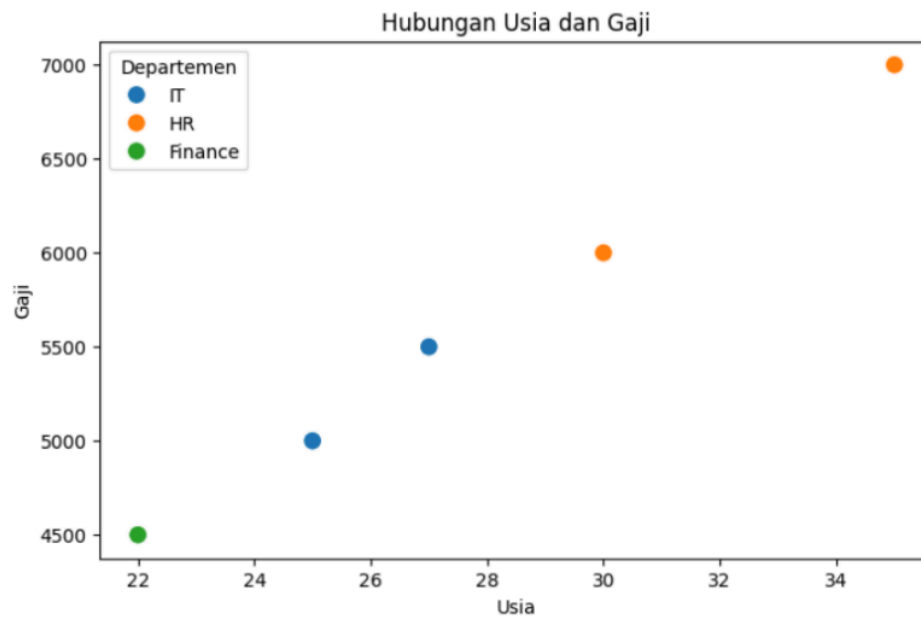
Departemen

IT 2

HR 2

Finance 1

Name: count, dtype: int64



5. Visualisasi Two-Dimensional Data

Visualisasi adalah aspek penting dalam analisis data dua dimensi. Beberapa metode umum dalam visualisasi data dua dimensi meliputi:

- **Histogram:** Menampilkan distribusi nilai dalam suatu dataset.
- **Scatter Plot:** Memvisualisasikan hubungan antara dua variabel.
- **Box Plot:** Menampilkan distribusi data dan mendeteksi outlier.
- **Heatmap:** Menunjukkan hubungan antar variabel dalam bentuk warna.

Dengan menggunakan alat visualisasi seperti Matplotlib dan Seaborn, kita dapat memahami pola dalam data dengan lebih mudah dan efektif.

6. Aplikasi dan Tantangan dalam Pengolahan Two-Dimensional Data

6.1. Aplikasi Two-Dimensional Data

Data dua dimensi digunakan dalam berbagai bidang, antara lain:

- **Analisis Keuangan:** Digunakan untuk memantau tren pasar, keuntungan perusahaan, dan pengeluaran.
- **Sistem Rekomendasi:** Data pelanggan dianalisis dalam bentuk tabel untuk memberikan rekomendasi produk.
- **Pengolahan Data Medis:** Digunakan dalam pencatatan data pasien dan diagnosis penyakit.
- **Prediksi Cuaca:** Data suhu, kelembaban, dan curah hujan disimpan dalam format tabel untuk analisis lebih lanjut.
- **Pemrosesan Citra Digital:** Gambar direpresentasikan dalam bentuk matriks piksel yang kemudian dianalisis menggunakan algoritma machine learning.

6.2. Tantangan dalam Pengolahan Two-Dimensional Data:

Meskipun sangat berguna, pengolahan data dua dimensi memiliki beberapa tantangan, seperti:

- **Data Hilang (Missing Data):** Beberapa nilai mungkin kosong dan harus ditangani dengan teknik imputasi.
- **Redundansi Data:** Duplikasi data dapat menyebabkan bias dalam analisis.
- **Kompleksitas Data:** Jumlah kolom yang terlalu banyak dapat memperumit analisis dan memerlukan teknik seleksi fitur.
- **Skalabilitas:** Pengolahan data dalam jumlah besar membutuhkan sumber daya komputasi yang lebih besar.

Dengan memahami dan mengatasi tantangan-tantangan ini, kita dapat mengelola data dua dimensi dengan lebih efektif untuk menghasilkan analisis yang lebih akurat dan bermanfaat.

7. Kesimpulan

Two-dimensional data merupakan format umum dalam machine learning dan analisis data. Dengan memahami cara merepresentasikan, memanipulasi, dan memvisualisasikan data dalam dua dimensi, kita dapat meningkatkan efisiensi dalam penerapan berbagai algoritma machine learning.

Referensi :

1. Ilavsky, J. (2012). Nika: software for two-dimensional data reduction. *Applied Crystallography*, 45(2), 324-328.

ORIGINALITY REPORT

0%

SIMILARITY INDEX

0%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

Exclude quotes On

Exclude bibliography On

Exclude matches < 3%

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7
