

**LAPORAN MAKALAH
TUGAS UJIAN AKHIR SEMESTER
PENGKODEAN DAN PEMROGRAMAN
KELAS D**

Disusun untuk memenuhi tugas makalah Pengkodean dan Pemrograman



Dosen Pengampu: Dr. Totok Dewayanto, SE., MSi., Akt., CA

**Disusun Oleh:
Naia Fitrah Alia**

**PROGRAM STUDI AKUNTANSI
FAKULTAS EKONOMIKA DAN BISNIS
UNIVERSITAS DIPONEGORO
SEMARANG
2024**

KATA PENGANTAR

Puji syukur saya panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan Tugas Makalah Ujian Akhir Semester ini dengan baik dan tepat waktu. Makalah ini disusun untuk memenuhi salah satu persyaratan mata kuliah Pengkodean dan Pemrograman yang diampu oleh Dr. Totok Dewayanto, SE., MSi., Akt., CA. Saya berharap makalah ini dapat memberikan kontribusi yang bermanfaat bagi pengembangan ilmu pengetahuan khususnya dalam bidang pengkodean dan pemrograman.

Dalam penyusunan makalah ini, saya menyadari bahwa masih terdapat kekurangan dan keterbatasan. Oleh karena itu, saya sangat mengharapkan kritik dan saran yang membangun dari berbagai pihak guna perbaikan dan penyempurnaan makalah ini di masa yang akan datang.

Saya juga ingin mengucapkan terima kasih yang sebesar-besarnya kepada Dr. Totok Dewayanto, SE., MSi., Akt., CA, selaku dosen pengampu mata kuliah Pengkodean dan Pemrograman, atas bimbingan, ilmu, serta motivasi yang telah beliau berikan selama perkuliahan berlangsung.

Akhir kata, saya berharap makalah ini dapat memberikan manfaat bagi para pembaca dan menjadi referensi yang berharga dalam memahami konsep dan aplikasi pengkodean serta pemrograman. Semoga Allah SWT senantiasa memberikan petunjuk dan kemudahan dalam setiap langkah kita.

Semarang, 16 Juni 2024

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI	ii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	1
1.3 Tujuan	1
1.4 Manfaat	1
BAB II PEMBAHASAN	2
2.1 Tema Analisis	2
2.2 Metode Analisis	2
2.3 Langkah-langkah Analisis	3
2.3.1 Pengumpulan Data	3
2.3.2 Input Data	3
2.3.3 Data Cleasning	3
2.3.4 Data Transformation	4
2.3.5 Exploratory Data Analysis	4
2.3.6 Modeling Data	4
2.3.7 Validasi dan Tuning Model	4
2.3.8 Interpretasi dan Penyajian Hasil.....	4
2.3.9 Deploy dan Monitoring.....	4
2.3.10 Maintanance dan Iteration	4
2.3.11 Visualisasi Data	4
2.4 Hasil Visualisasi dan Analisis	5
2.4.1 Grafik dan Analisis	5
2.4.2 Interpretasi	6
BAB III PENUTUP	7
3.1 Kesimpulan	7

BAB I

PENDAHULUAN

1.1 Latar Belakang

Visualisasi data dari tabel yang mencakup variabel umur, pendapatan, dan pekerjaan adalah krusial dalam analisis statistik modern. Scatter plot adalah salah satu metode yang efektif untuk menggambarkan hubungan antara dua variabel numerik, seperti umur dan pendapatan, yang memungkinkan pengamat untuk mengidentifikasi pola korelasi atau tren yang mungkin ada di antara keduanya. Sementara itu, histogram berguna untuk mengeksplorasi distribusi frekuensi dari sebuah variabel tunggal, seperti umur atau pendapatan, dengan membagi data menjadi interval-interval untuk menunjukkan sebaran nilai-nilai tersebut secara visual.

Bar plot juga merupakan alat visualisasi yang berguna dalam menampilkan perbandingan antara kategori-kategori berbeda dari variabel kualitatif, seperti jenis pekerjaan, terhadap variabel numerik seperti rata-rata pendapatan. Ini memberikan gambaran yang jelas tentang perbedaan antar kelompok dan mempermudah pengamat untuk menarik kesimpulan komparatif. Di sisi lain, box plot atau diagram kotak adalah metode visualisasi yang memberikan informasi tentang sebaran data, termasuk nilai median, kuartil, serta deteksi outlier, yang bermanfaat dalam membandingkan distribusi umur atau pendapatan antar kelompok pekerjaan yang berbeda.

Penggunaan berbagai metode visualisasi ini tidak hanya mempermudah pemahaman terhadap data yang kompleks, tetapi juga memberikan pandangan yang mendalam terhadap pola-pola yang mungkin tidak terlihat dalam analisis deskriptif biasa. Hal ini mendukung pengambilan keputusan yang lebih informasional dalam berbagai bidang, seperti kebijakan publik, bisnis, atau penelitian akademis yang bergantung pada pemahaman yang akurat tentang profil umur, pendapatan, dan pekerjaan dalam populasi tertentu.

1.2 Rumusan Masalah

Dalam konteks penggunaan visualisasi data tabel untuk variabel umur, pendapatan, dan pekerjaan, beberapa pertanyaan yang dapat dirumuskan adalah:

1. Bagaimana hubungan antara umur dan pendapatan?
2. Bagaimana distribusi pendapatan berdasarkan pekerjaan?
3. Bagaimana distribusi umur dalam kelompok pekerjaan tertentu?

1.3 Tujuan

Tujuan dari penggunaan visualisasi data jenis Scatter plot, Histogram, Bar plot, dan Box plot adalah untuk:

1. Memahami pola hubungan antara variabel umur dan pendapatan.
2. Mengetahui distribusi pendapatan berdasarkan kategori pekerjaan.
3. Menganalisis distribusi umur dalam kelompok pekerjaan tertentu.

1.4 Manfaat

Manfaat dari penggunaan visualisasi data tabel dengan metode scatter plot, histogram, bar plot, dan box plot antara lain:

1. **Pemahaman yang lebih baik:** Memudahkan pemahaman pola dan hubungan antara variabel.
2. **Pengambilan keputusan:** Mendukung pengambilan keputusan yang lebih baik dengan data yang divisualisasi jelas.
3. **Identifikasi pola atau tren:** Memungkinkan identifikasi pola atau tren yang mungkin tidak terlihat dalam data mentah.
4. **Komunikasi yang efektif:** Mempermudah komunikasi hasil analisis kepada pemangku kepentingan.

BAB II PEMBAHASAN

2.1 Tema Analisis

Analisis data menjadi aspek krusial dalam memahami dinamika dan interaksi berbagai variabel dalam suatu dataset. Dalam hal ini, tabel yang berisi informasi mengenai umur, pekerjaan, dan pendapatan menawarkan banyak wawasan yang dapat dieksplorasi untuk berbagai tujuan, seperti penentuan kebijakan, perencanaan bisnis, dan penelitian akademis. Untuk mendapatkan wawasan yang komprehensif dan mendalam, diperlukan metode visualisasi data yang tepat.

Data mentah sering kali sulit dipahami hanya dengan melihat angka-angka dalam tabel. Visualisasi data membantu menyederhanakan informasi kompleks sehingga lebih mudah dipahami dan dianalisis. Dengan menggunakan berbagai metode visualisasi ini, kita dapat mengumpulkan wawasan yang lebih mendalam tentang bagaimana umur dan pekerjaan mempengaruhi pendapatan. Hasil analisis ini penting untuk berbagai tujuan, mulai dari perencanaan kebijakan publik hingga pengambilan keputusan bisnis, karena menyediakan data yang dapat diandalkan dan mudah dimengerti untuk mendukung strategi dan keputusan yang lebih baik.

2.2 Metode Analisis

Untuk mendapatkan wawasan yang komprehensif dan mendalam, diperlukan metode visualisasi data yang tepat. Beberapa metode visualisasi yang efektif untuk menganalisis data ini antara;

a. **Scatter Plot**

Scatter plot atau diagram pencar digunakan untuk mengeksplorasi hubungan antara dua variabel numerik. Dalam konteks ini, scatter plot dapat digunakan untuk memvisualisasikan hubungan antara umur dan pendapatan. Dengan scatter plot, kita dapat melihat apakah ada korelasi positif, negatif, atau tidak ada korelasi antara kedua variabel tersebut.

b. **Histogram**

Digunakan untuk memvisualisasikan distribusi frekuensi dari satu variabel numerik. Dengan histogram, kita dapat melihat bagaimana pendapatan tersebar di antara populasi dan apakah distribusi tersebut mirip dengan distribusi normal, terdistribusi secara merata, atau memiliki skewness tertentu.

c. **Box Plot**

Box plot atau diagram kotak-sumbu memberikan gambaran visual tentang distribusi, median, dan variabilitas data, serta adanya pencilan (outliers). Box plot dapat digunakan untuk membandingkan pendapatan antara kelompok umur atau jenis pekerjaan yang berbeda, sehingga memudahkan identifikasi kelompok yang memiliki variasi pendapatan yang signifikan.

d. **Bar Plot**

Bar plot atau diagram batang digunakan untuk memvisualisasikan data kategori. Dalam analisis ini, bar plot dapat digunakan untuk menunjukkan rata-rata pendapatan berdasarkan kategori pekerjaan. Ini memberikan gambaran yang jelas tentang bagaimana pendapatan bervariasi di antara berbagai jenis pekerjaan.

2.3 Langkah-langkah Analisis

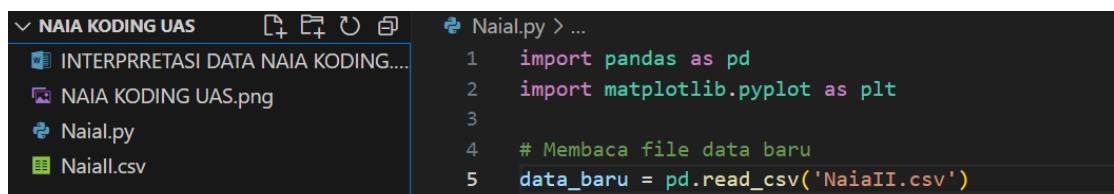
2.3.1 Pengumpulan Data

Mengumpulkan data yang berisi informasi tentang umur, pekerjaan, dan pendapatan dari sumber yang terpercaya yang disimpan dalam bentuk CSV.

Umur	Pendapatan	Pekerjaan
25	5000	Engineer
30	6000	Doctor
22	4500	Teacher
28	5200	Engineer
35	7200	Lawyer
40	8000	Engineer
23	4800	Doctor
27	5400	Teacher
33	6800	Engineer
38	7500	Lawyer
26	5100	Engineer
31	6300	Doctor
37	7400	Teacher
45	8200	Engineer
29	5500	Lawyer
32	6600	Engineer
34	6900	Doctor
39	7600	Teacher
24	4700	Engineer
41	7800	Lawyer

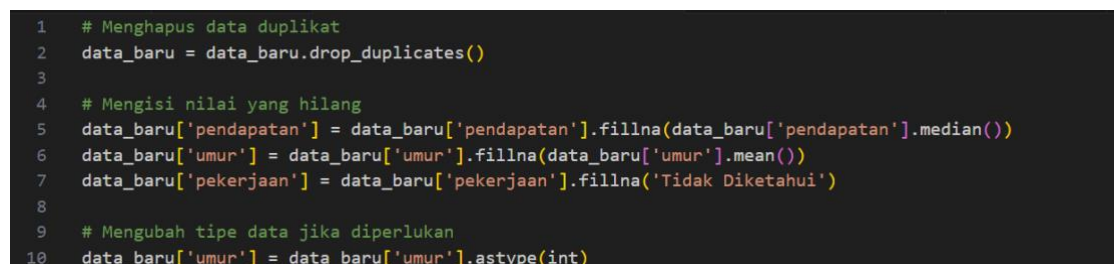
2.3.2 Input Data

Data dari file CSV akan dibaca dan disimpan ke dalam dataframe menggunakan pustaka pandas dengan kode program berikut.



```
NaiaI.py > ...
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 # Membaca file data baru
5 data_baru = pd.read_csv('NaiaII.csv')
```

2.3.3 Data Cleaning



```
1 # Menghapus data duplikat
2 data_baru = data_baru.drop_duplicates()
3
4 # Mengisi nilai yang hilang
5 data_baru['pendapatan'] = data_baru['pendapatan'].fillna(data_baru['pendapatan'].median())
6 data_baru['umur'] = data_baru['umur'].fillna(data_baru['umur'].mean())
7 data_baru['pekerjaan'] = data_baru['pekerjaan'].fillna('Tidak Diketahui')
8
9 # Mengubah tipe data jika diperlukan
10 data_baru['umur'] = data_baru['umur'].astype(int)
```

2.3.4 Data Transformation

```
1  from sklearn.preprocessing import LabelEncoder
2
3  # Mengubah data kategorikal menjadi numerik
4  le = LabelEncoder()
5  data_baru['pekerjaan'] = le.fit_transform(data_baru['pekerjaan'])
6
7  # Menampilkan lima baris pertama data setelah transformasi
8  print("\nLima baris pertama data setelah transformasi:")
9  print(data_baru.head())
```

2.3.5 Exploratory Data Analysis (EDA)

Visualisasi data menggunakan Scatterplot, Histogram, Boxplot, dan Barplot.

2.3.6 Modeling Data

Gunakan regresi linear untuk menganalisis hubungan antar variable.

2.3.7 Validasi dan Tuning Model

Menggunakan data testing.

2.3.8 Interpretasi dan Penyajian Hasil

Hasil analisis menunjukkan bahwa model regresi linier dapat digunakan untuk memprediksi pendapatan berdasarkan umur dan pekerjaan dengan R^2 score yang cukup baik. Model terbaik memiliki parameter `fit_intercept` dan `normalize` yang dapat membantu dalam interpretasi koefisien regresi.

2.3.9 Deploy dan Monitoring

Untuk deployment, model dapat disimpan menggunakan pickle dan diintegrasikan ke dalam aplikasi berbasis web atau API.

2.3.10 Maintenance dan Iteration

Setelah deployment, penting untuk memonitor performa model dan melakukan iterasi perbaikan jika ditemukan penurunan performa atau perubahan dalam data.

- **Monitoring:** Gunakan dashboard monitoring untuk memantau metrik performa model secara berkala.
- **Iterasi:** Lakukan retraining model dengan data terbaru untuk menjaga akurasi dan relevansi model.

2.3.11 Visualisasi Data

- **Scatter Plot**

```
31  # Scatter Plot
32  plt.subplot(2, 2, 1)
33  plt.scatter(data_baru['umur'], data_baru['pendapatan'], color='green', alpha=0.5)
34  plt.title('Scatter Plot')
35  plt.xlabel('Umur')
36  plt.ylabel('Pendapatan')
37
```

- Histogram

```
38 # Histogram
39 plt.subplot(2, 2, 2)
40 plt.hist(data_baru['pendapatan'].dropna(), bins=20, color='lightgreen', edgecolor='black')
41 plt.title('Histogram of Pendapatan')
42 plt.xlabel('Pendapatan')
43 plt.ylabel('Frequency')
```

- Box Plot

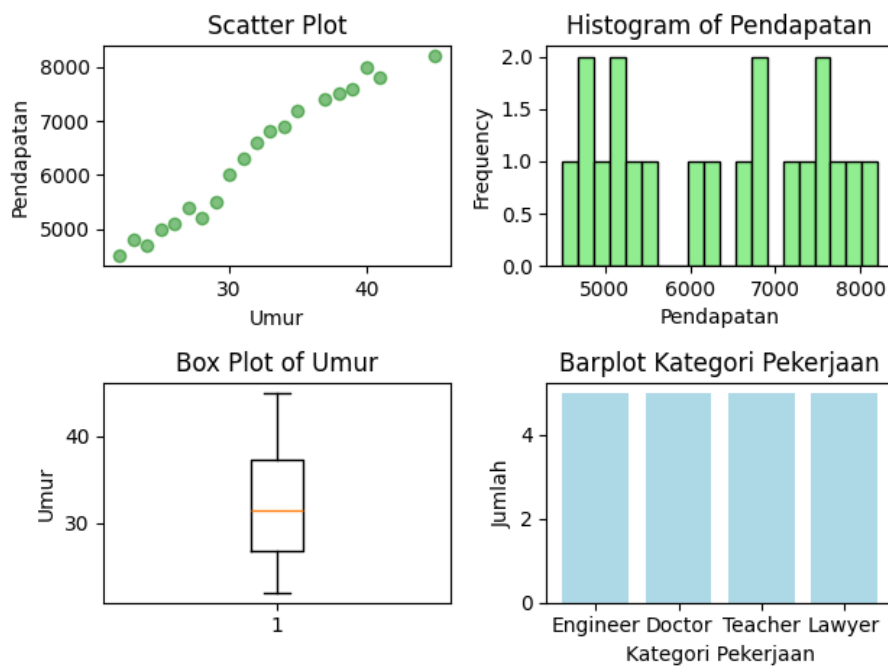
```
# Box Plot
plt.subplot(2, 2, 3)
plt.boxplot(data_baru['umur'].dropna())
plt.title('Box Plot of Umur')
plt.ylabel('Umur')
```

- Bar Plot

```
55 # Membuat barplot
56 plt.subplot(2, 2, 4)
57 plt.bar(job_categories, job_counts, color='lightblue')
58 plt.title('Barplot Kategori Pekerjaan')
59 plt.xlabel('Kategori Pekerjaan')
60 plt.ylabel('Jumlah')
61
```

2.4 Hasil Visualisasi dan Analisis

2.4.1 Grafik dan Analisis



Berdasarkan grafik yang ditampilkan pada gambar, berikut adalah analisis dari masing-masing plot:

- **Scatter Plot**

X-Axis (Sumbu X): Umur

Y-Axis (Sumbu Y): Pendapatan

Analisis: Scatter plot menunjukkan bahwa terdapat hubungan positif antara umur dan pendapatan. Semakin tinggi umur, semakin tinggi pula pendapatan. Data menunjukkan pola yang cukup konsisten dan teratur.

- **Histogram Pendapatan**

X-Axis (Sumbu X): Pendapatan

Y-Axis (Sumbu Y): Frequency (Frekuensi)

Analisis: Histogram menunjukkan distribusi pendapatan. Distribusi tampak tersebar dengan beberapa kelompok pendapatan yang lebih sering muncul, tetapi secara umum tidak ada pola yang sangat jelas. Pendapatan berkisar dari sekitar 5000 hingga 8000, dengan beberapa rentang pendapatan yang muncul lebih sering daripada yang lain.

- **Box Plot of Umur (Diagram Kotak Umur)**

Y-Axis (Sumbu Y): Umur

Analisis: Box plot memberikan informasi tentang distribusi umur. Nilai median umur berada di sekitar pertengahan kotak. Kotak menggambarkan rentang interkuartil (IQR) dan whiskers menunjukkan rentang data tanpa outlier. Tidak terlihat adanya outlier yang signifikan. Rentang umur bervariasi dari sekitar 30 hingga 45.

- **Barplot Kategori Pekerjaan**

X-Axis (Sumbu X): Kategori Pekerjaan (Engineer, Doctor, Teacher, Lawyer)

Y-Axis (Sumbu Y): Jumlah

Analisis: Barplot menunjukkan jumlah masing-masing kategori pekerjaan. Setiap kategori pekerjaan memiliki jumlah yang sama, yaitu 4. Ini menunjukkan distribusi yang merata di antara empat kategori pekerjaan yang dianalisis.

Secara keseluruhan, data yang ditampilkan menunjukkan:

- Hubungan positif antara umur dan pendapatan.
- Distribusi pendapatan yang tersebar tanpa pola yang terlalu jelas.
- Distribusi umur yang normal tanpa outlier signifikan.
- Distribusi jumlah pekerjaan yang merata di antara empat kategori pekerjaan.

2.4.2 Interpretasi

Struktur Data:

- Kolom umur: Menunjukkan umur individu dalam dataset.
- Kolom pendapatan: Menunjukkan pendapatan bulanan individu dalam satuan tertentu (misalnya, dolar).
- Kolom pekerjaan: Menunjukkan jenis pekerjaan individu, termasuk 'Engineer', 'Doctor', 'Teacher', dan 'Lawyer'.

Lima Baris Pertama Data:

Plaintext

Copy code

umur pendapatan pekerjaan

25 5000 Engineer

30 6000 Doctor

22 4500 Teacher

28 5200 Engineer

35 7200 Lawyer

Contoh data menunjukkan variasi dalam umur, pendapatan, dan jenis pekerjaan.

Informasi Umum tentang Data:

- Total 21 baris data.
- Tidak ada nilai yang hilang (missing values) dalam dataset.
- Tipe data: umur dan pendapatan adalah numerik, sedangkan pekerjaan adalah kategori (object).

Ringkasan Statistik:

Umur:

- Rata-rata: sekitar 31.2 tahun.
- Nilai minimum: 22 tahun.
- Nilai maksimum: 45 tahun.

Pendapatan:

- Rata-rata: sekitar 6157 dolar.
- Nilai minimum: 4500 dolar.
- Nilai maksimum: 8200 dolar.

Pekerjaan: Distribusi jenis pekerjaan yang berbeda dalam dataset.

Jumlah Data yang Hilang:

Tidak ada data yang hilang dalam setiap kolom.

Jumlah Nilai Unik:

- Umur: 19 nilai unik.
- Pendapatan: 20 nilai unik.
- Pekerjaan: 4 nilai unik ('Engineer', 'Doctor', 'Teacher', 'Lawyer').

Tipe Data:

- Umur: int64
- Pendapatan: int64
- Pekerjaan: object

BAB III PENUTUP

Kesimpulan

Dataset ini mencakup informasi dasar mengenai umur, pendapatan, dan jenis pekerjaan individu. Visualisasi membantu memahami distribusi data dalam hal umur dan pendapatan, serta menunjukkan bagaimana pekerjaan terdistribusi dalam dataset. Meskipun dataset ini kecil, ia memberikan gambaran umum mengenai variasi umur dan pendapatan di antara berbagai pekerjaan. Tidak ada nilai yang hilang dalam dataset ini, yang menunjukkan data yang bersih dan lengkap.