

Laporan Pengerjaan Praktikum

Proyek 1: Visualisasi Data

Disusun oleh:

Nama : Muhammad Nabil Syauqi Rasyiq
NIM : 241524018
Kelas : 1A



Sarjana Terapan Program Studi Teknik Informatika
Jurusan Teknik Komputer dan Informatika
Politeknik Negeri Bandung
2025

DAFTAR ISI

DAFTAR ISI	2
BAB I DESKRIPSI VISUALISASI DATA	3
1.1 Hubungan IPK dan Kebahagiaan Mahasiswa Informatika	3
1.2 Distribusi Kebahagiaan Mahasiswa Informatika Setiap Kategori	3
1.3 Korelasi Antar Variabel Pasien.....	4
1.4 Rata-Rata Kekayaan Pasien berdasarkan diagnose dan Gender	5
1.5 Usia vs Kebahagiaan Pasien.....	6
1.6 Rata-Rata IPK Setiap Kategori	7
1.7 Modifikasi Visual menggunakan violin	8
BAB II DATA SET CUSTOM	10
2.1 Data Set Siswa K-12	10
2.1.1 ID_Pasien	10
2.1.2 Usia	10
2.1.3 Nilai_Rapor	11
2.1.4 Gender	11
2.1.5 Prestasi	11
2.1.6 Dukungan_Orang_Tua	11
2.1.7 Durasi_Penggunaan_Gawai	11
2.1.8 Kebahagiaan.....	11
2.1.9 Prestasi_Numerik	11
2.1.10 Dukungan_Orang_Tua_Numerik:.....	11
2.2 Visual Analisis yang Digunakan	12
2.2.1 Scatter Plot untuk visualisasi korelasi durasi gawai vs Nilai Rapor	12
2.2.2.....	12
2.2.3.....	13
2.3 Mengustomisasi Visual durasi vs Nilai	14
LESSON LEARN	16

BAB I DESKRIPSI VISUALISASI DATA

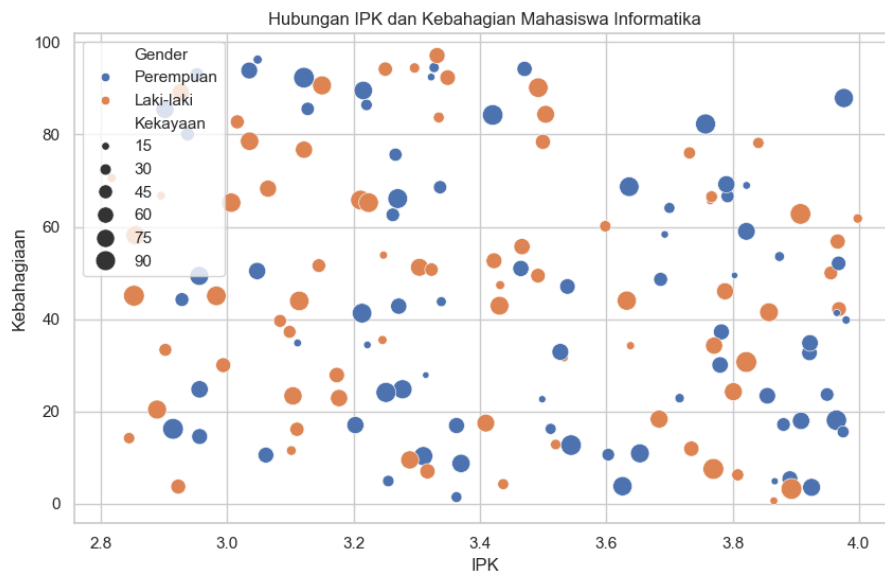
1.1 Hubungan IPK dan Kebahagiaan Mahasiswa Informatika

Secara umum Scatter Plot biasanya digunakan untuk menunjukkan hubungan atau korelasi antara dua variabel numerik.

```
# 1. Seaborn Scatter Plot
```

```
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df_informatika, x= 'IPK', y= 'Kebahagiaan',
hue='Gender', size = 'Kekayaan', sizes =(20,200))
plt.title("Hubungan IPK dan Kebahagiaan Mahasiswa Informatika")
plt.savefig("scatter_informatika_ipk_kebahagiaan.png")
plt.close()
```

Visualisasi pertama menggambarkan Hubungan antara IPK mahasiswa dengan Kebahagiaannya. Dalam penggunaan data asli kita dapat melihat hubungannya secara langsung, apakah mahasiswa dengan IPK yang tinggi memiliki Kebahagiaan yang tinggi juga? Atau sebaliknya, dengan menggunakan data visual ini juga kita dapat melihat peran Tingkat kekayaan terhadap nilai IPK ataupun Kebahagiaan serta dimana bagaimana penyebarannya secara gender.



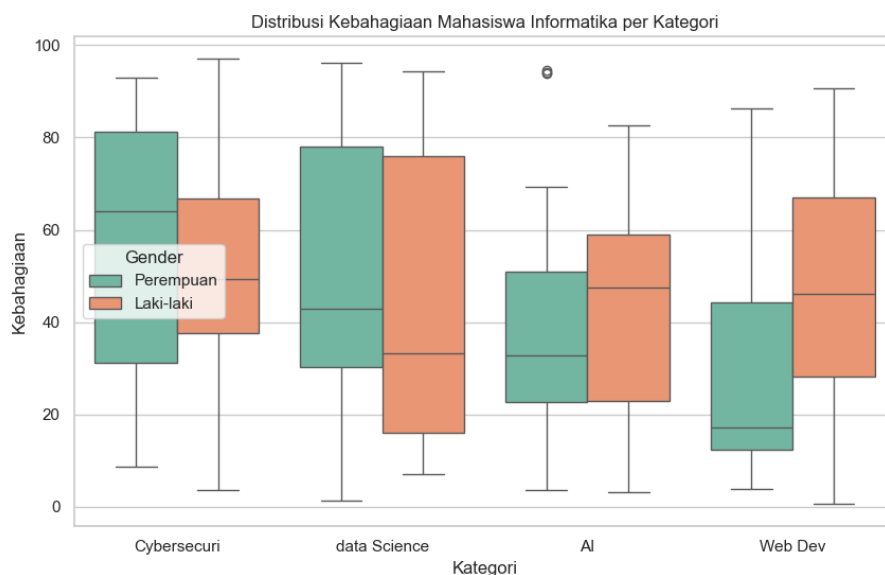
1.2 Distribusi Kebahagiaan Mahasiswa Informatika Setiap Kategori

Box Plot biasanya sangat baik untuk digunakan dalam membandingkan distribusi variabel numerik.

2. Seaborn box Plot

```
plt.figure(figsize=(10, 6))
sns.boxplot(data=df_informatika, x='Kategori', y = 'Kebahagiaan', hue
= 'Gender' , palette='Set2')
plt.title("Distribusi Kebahagiaan Mahasiswa Informatika per
Kategori")
plt.savefig('box_informatika_kebahagiaan.png')
plt.close()
```

Pada kasus ini, kita menggunakan box plot untuk melihat distribusi nilai kebahagiaan Mahasiswa setiap kategori. Dari visualisasi data ini kita dapat melihat distribusi tingkat kebahagiaan mahasiswa disetiap kategori. Dengan adanya fitur Hue juga kita dapat langsung membedakan antara gender laki-laki atau pun perempuan.



1.3 Korelasi Antar Variabel Pasien

Heatmap biasa digunakan untuk memvisualisasikan matriks korelasi

```
# 3. Seaborn Heatmap (Penyakit)
plt.figure(figsize=(10, 8))
corr_penyakit = df_penyakit[['Usia', 'Tekanan_Darah', 'Kadar_Gula',
                             'Durasi_Rawat', 'Biaya', 'IPK',
                             'Kekayaan', 'Kebahagiaan']].corr()
sns.heatmap(corr_penyakit, annot=True, cmap='coolwarm', vmin=-1,
vmax=1)
plt.title('Korelasi Antar Variabel Pasien')
```

```
plt.savefig('heatmap_penyakit_korelasi.png')
plt.close()
```

Dalam visualisasi data kali ini, kita dapat melihat bagaimana nilai korelasi antar variabel pasien. Semakin merah merah semakin variabel itu berkorelasi sebaliknya semakin biru atau pucat semakin rendah nilai korelasinya. Dari visualisasi ini idelanya kita dapat menentukan mana yang saling berhubungan atau berkorelasi dan man yang tidak.

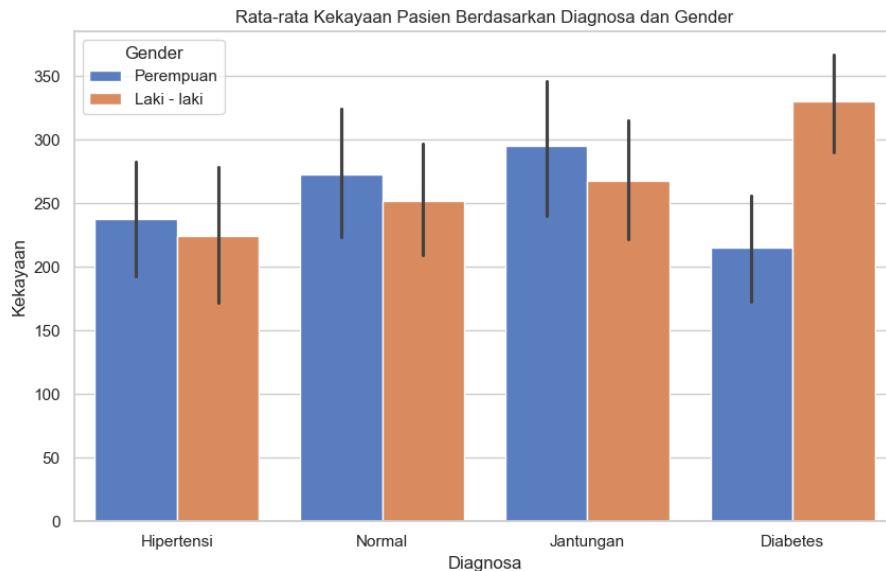


1.4 Rata-Rata Kekayaan Pasien berdasarkan diagnose dan Gender

Bar plot biasanya digunakan untuk membandingkan nilai rata-rata dari variabel numerik di berbagai kategori.

```
# 4. Seaborn Bar Plot (Penyakit)
plt.figure(figsize=(10, 6))
sns.barplot(data=df_penyakit, x='Diagnosa', y='Kekayaan',
            hue='Gender', palette='muted')
plt.title('Rata-rata Kekayaan Pasien Berdasarkan Diagnosa dan Gender')
plt.savefig('bar_penyakit_kekayaan.png')
plt.close()
```

Visualisasi ini idealnya menggambarkan penyakit yang diderita oleh orang dengan kekayaan tertentu. Pada data set yang dibuat secara random menunjukkan bahwa kekayaan tidak terlalu menentukan penyakit. Disisi lain kita dapat, membandingkan gender mana yang lebih dominan di suatu penyakit.



1.5 Usia vs Kebahagiaan Pasien

Bokeh Scatter plot serupa dengan seaborn Scatter plot dimana ia biasanya digunakan untuk menunjukkan hubungan antara 2 variabel numerik.

```
# 5. Bokeh Scatter Plot (Penyakit)
output_file('bokeh_scatter_penyakit_usia_kebahagiaan.html')
source_penyakit = ColumnDataSource(df_penyakit)

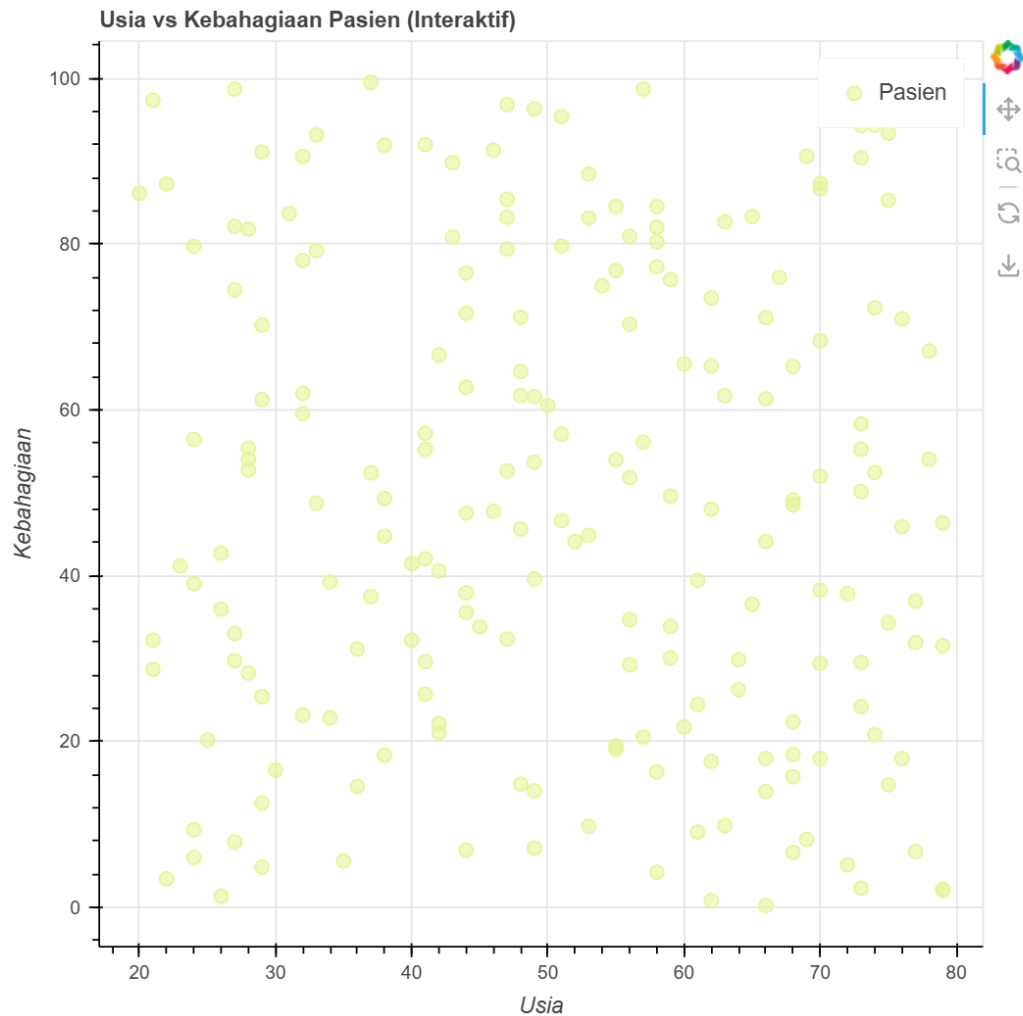
p1 = figure(title="Usia vs Kebahagiaan Pasien (Interaktif)",
            x_axis_label='Usia',
            y_axis_label='Kebahagiaan',
            tools="pan,box_zoom,reset,save")

p1.scatter('Usia', 'Kebahagiaan', source=source_penyakit,
          size=8, color=Spectral6[2],
          legend_label='Pasien', fill_alpha=0.6)

p1.legend.click_policy = "hide"
show(p1)
```

Visualisasi ini menunjukan hubungan antara usia pasien dengan kebahagiaannya. Idealnya kita dapat melihat bagaimana hubungan antara kebahagiaan pasien

dengan usianya. Meskipun saya belum menemukan fitur Hue pada jenis bokeh, saya menemukan bahwa fitur bokeh yang diexport menjadi html membuat kita dapat secara interaktif melihat visualisasi datanya.



1.6 Rata-Rata IPK Setiap Kategori

Bokeh Bar Plot serupa dengan seaborn bar plot yang biasanya digunakan untuk membandingkan nilai rata-rata variabel numerik di setiap kategori.

```
# 6. Bokeh Bar Plot (Informatika)
output_file('bokeh_bar_informatika_ipk.html')

ipk_kategori_gender = df_informatika.groupby(['Kategori',
'Gender'])['IPK'].mean().reset_index()
source_informatika = ColumnDataSource(ipk_kategori_gender)

p2 = figure(x_range=ipk_kategori_gender['Kategori'].unique(),
            title="Rata-rata IPK per Kategori dan Gender",
```

```

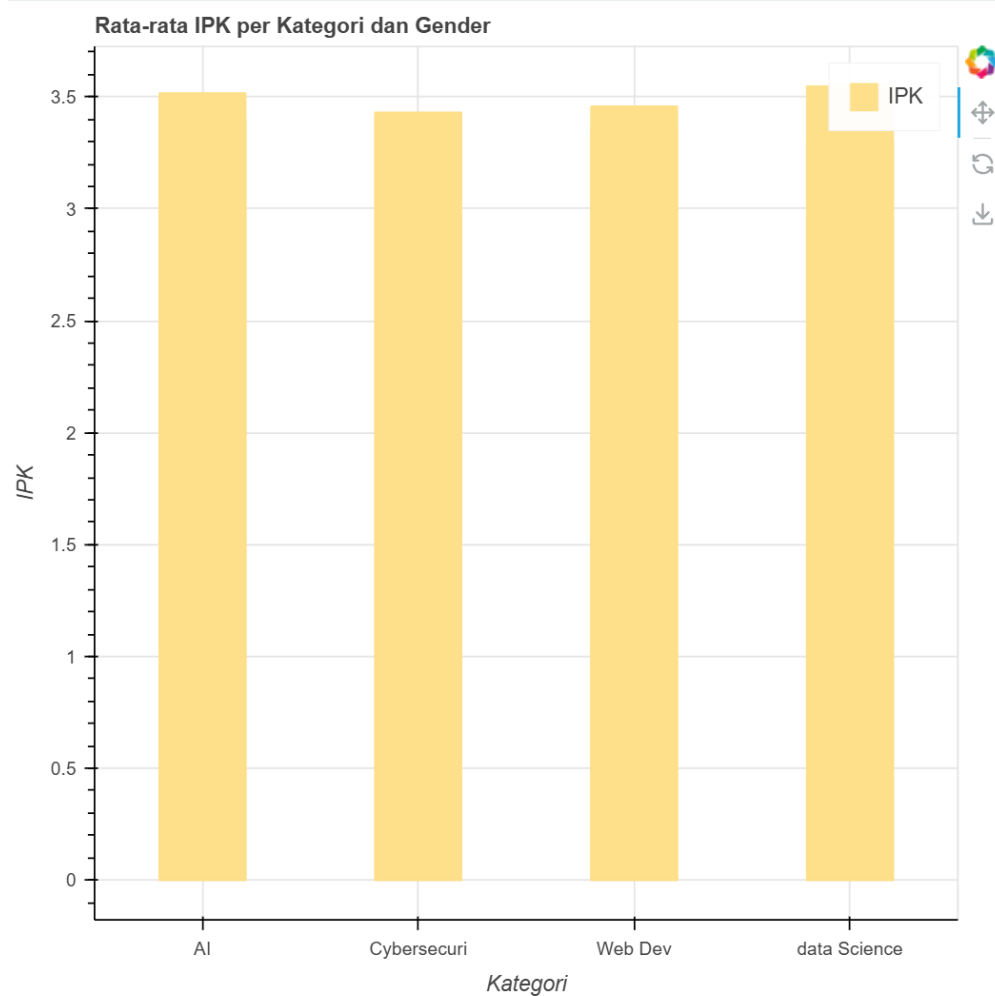
x_axis_label='Kategori', y_axis_label='IPK',
tools="pan,reset,save")

p2.vbar(x='Kategori', top='IPK', width=0.4,
        source=source_informatika, color=Spectral6[3],
        legend_label='IPK')

p2.legend.click_policy = "hide"
show(p2)

```

Visualisasi data ini idelanya kita dapat melihat rata rata IPK pada setiap kategori

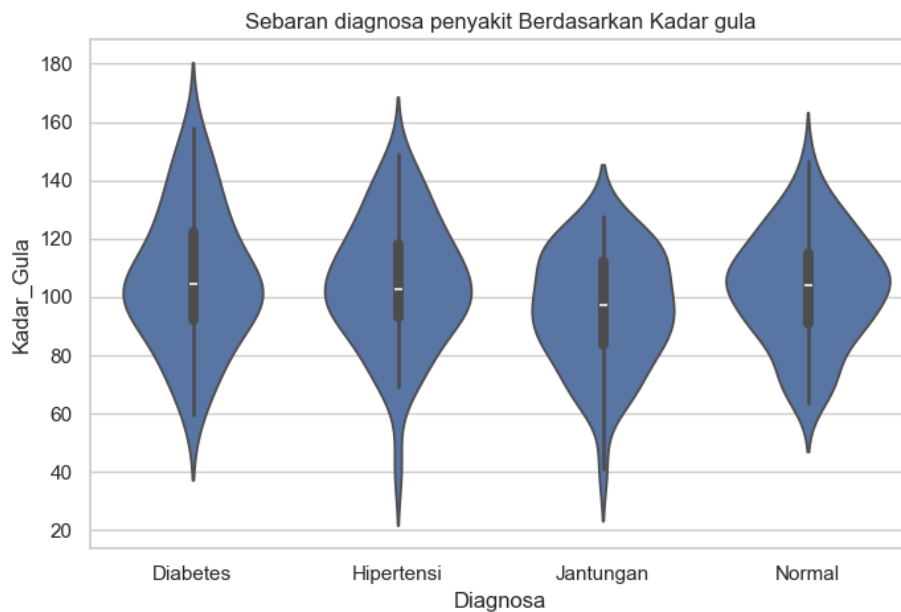


1.7 Modifikasi Visual menggunakan violin

Violin plot adalah jenis visualisasi data yang menggabungkan boxplot dan density plot untuk menunjukkan distribusi dari suatu variabel numerik dalam berbagai kategori.


```
# 7. violinplot sebagai modifikasi tampilan
plt.figure(figsize=(8, 5))
sns.violinplot(data=df_penyakit, x='Diagnosa', y='Kadar_Gula',
               order=['Diabetes', 'Hipertensi', 'Jantung',
                     'Normal'])
plt.title("Sebaran diagnosa penyakit Berdasarkan Kadar gula")
plt.savefig("violin_diagnosa_kadargula.png")
plt.close()
```

Aku menggunakan jenis visualiasi violin untuk melihat sebaran ini karena dengan menggunakan tipe visual yang ini kita dapat melihat kategori tingkatan mana yang lebih banyak menderita suatu penyakit berdasarkan kadar gulanya.



BAB II DATA SET CUSTOM

2.1 Data Set Siswa K-12

```
#Data Set Siswa K-12

n_siswa_k12 = 200
data_siswa_k12 = {
    'ID_Pasien ': [f'Pasien_{i}' for i in range ( n_siswa_k12 )],
    'Usia': np. random . randint (5 , 18, n_siswa_k12 ),
    'Nilai_Rapor': np. random . uniform (0 , 100 , n_siswa_k12 ),
    'Gender': np. random . choice (['Laki - laki' , 'Perempuan'],
n_siswa_k12 ),
    'Prestasi': np. random . choice (['Sangat Baik', 'Baik', 'Cukup',
'Kurang', 'Buruk'], n_siswa_k12 ),
    'Dukungan_Orang_Tua': np. random . choice (['Sangat
Baik', 'Baik', 'Cukup', 'Kurang', 'Buruk'], n_siswa_k12 ),
    'Durasi_Penggunaan_Gawai': np. random . uniform (0, 360 ,
n_siswa_k12 ),
    'Kebahagiaan': np. random . uniform (0, 100 , n_siswa_k12 )
}
df_siswa_k12 = pd. DataFrame(data_siswa_k12)
peta_dukungan_orang_tua = {'Sangat Baik' : 5, 'Baik': 4, 'Cukup' : 3,
'Kurang' : 2, 'Buruk':1 }
peta_prestasi = {'Sangat Baik' : 5, 'Baik': 4, 'Cukup' : 3, 'Kurang'
: 2, 'Buruk':1 }
df_siswa_k12['Prestasi_Numerik'] =
df_siswa_k12['Prestasi'].map(peta_prestasi)
df_siswa_k12['Dukungan_Orang_Tua_Numerik'] =
df_siswa_k12['Dukungan_Orang_Tua'].map(peta_dukungan_orang_tua)
```

Data set Siswa K-12 berisi data data siswa K-12 yang akan digunakan untuk analisis. Berikut penjelasan setiap variabel

2.1.1 ID_Pasien

Variabel nomor Id siswa sebagai nilai unnik dari setiap data siswa

2.1.2 Usia

Variabel Usia mencatat umur siswa dalam satuan tahun. Rentang usia yang dihasilkan (5 hingga 17 tahun) secara spesifik menargetkan populasi siswa dari jenjang pendidikan dasar (Kindergarten/Taman Kanak-Kanak) hingga pendidikan menengah atas (Sekolah Menengah Atas). Usia adalah faktor fundamental yang seringkali diidentifikasi oleh banyak peneliti dalam jurnal sebagai faktor yang

berkorelasi dengan tahap perkembangan kognitif, sosial, emosional, serta tuntutan akademis yang berbeda.

2.1.3 Nilai_Rapor

Variabel Nilai_Rapor merepresentasikan indikator kuantitatif dari performa akademik siswa secara keseluruhan, yang biasanya diukur dalam skala 0 hingga 100. Nilai ini merupakan rata-rata dari semua mata pelajaran yang diambil selama periode 1 semester.

2.1.4 Gender

Variabel Gender mengklasifikasikan siswa ke dalam kategori 'Laki-laki' atau 'Perempuan'. Ini adalah variabel yang sering kali digunakan peneliti dalam perbandingan pendidikan dan sosial.

2.1.5 Prestasi

Variabel Prestasi memberikan penilaian kualitatif terhadap pencapaian siswa, yang mencakup aspek akademik maupun non-akademik. Skala yang digunakan ('Sangat Baik', 'Baik', 'Cukup', 'Kurang', 'Buruk'), menunjukkan tingkatan pencapaian dari yang tertinggi hingga terendah.

2.1.6 Dukungan_Orang_Tua

Variabel ini menampung tingkat dukungan orang tua pada seorang anak. Variabel ini penting karena sudah banyak jurnal penelitian yang mengidentifikasi dukungan orang tua sebagai penentu prestasi anak.

2.1.7 Durasi_Penggunaan_Gawai

Variabel ini mencatat penggunaan gawai anak dalam satuan menit. Variabel ini penting karena durasi penggunaan gawai telah diidentifikasi sebagai variabel penentu tingkat kognitif anak oleh banyak penelitian.

2.1.8 Kebahagiaan

Variabel Kebahagiaan mencoba untuk mengukur tingkat kebahagiaan siswa K-12 yang nantinya akan dibandingkan dengan variabel variabel lain dan di Analisa korelasinya.

2.1.9 Prestasi_Numerik

Variabel ini merupakan transformasi pemetaan nilai dari prestasi dari kualitatif menjadi kuantitatif untuk memudahkan analisis.

2.1.10 Dukungan_Orang_Tua_Numerik:

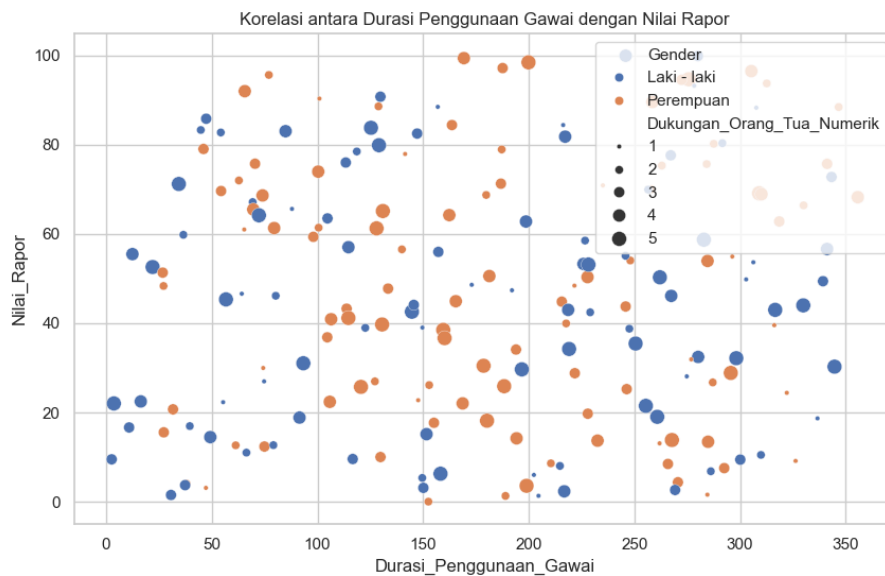
Variabel ini merupakan transformasi pemetaan nilai dari dukungan orang tua dari kualitatif menjadi kuantitatif untuk memudahkan analisis.

2.2 Visual Analisis yang Digunakan

2.2.1 Scatter Plot untuk visualisasi korelasi durasi gawai vs Nilai Rapor

```
# 1. Visualisasi Scatter plot untuk menunjukkan korelasi antara
durasi gawai vs Nilai
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df_siswa_k12, x='Durasi_Penggunaan_Gawai', y =
'Nilai_Rapor', hue = 'Gender' , size = 'Dukungan_Orang_Tua_Numerik',
sizes =(10,100))
plt.title("Korelasi antara Durasi Penggunaan Gawai dengan Nilai
Rapor")
plt.savefig("scatter_siswaK12_Rapor_DurasiGawai.png")
plt.close()
```

Saya menggunakan visualisasi Scatter plot karena visualisasi ini dapat secara langsung memperlihatkan perbedaan siswa yang memiliki durasi layer yang melebihi standar dengan siswa yang memiliki durasi layer sesuai dengan standar. Ditambah dengan visualisasi ini saya dapat menambahkan variabel dukungan orang tua sebagai variabel penentu tambahan dimana dukungan orang tua dapat diartikan sebagai penggunaan gawai yang terkontrol.

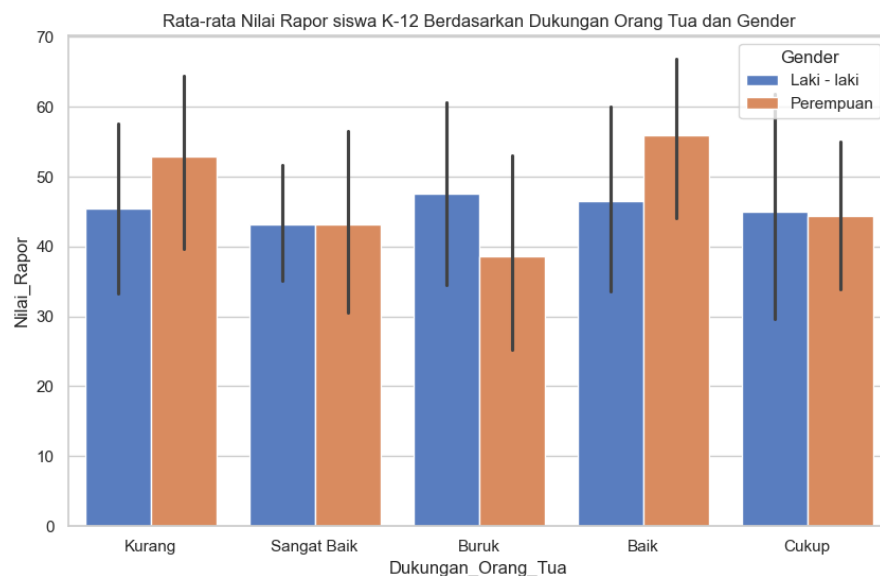


2.2.2

```
# 2. Seaborn Bar Plot untuk membandingkan nilai rata rata nilai
Rapor berdasarkan dukungan orang tua dan gender
plt.figure(figsize=(10, 6))
```

```
sns.barplot(data=df_siswa_k12, x='Dukungan_Orang_Tua',
y='Nilai_Rapor', hue='Gender', palette='muted')
plt.title('Rata-rata Nilai Rapor siswa K-12 Berdasarkan Dukungan
Orang Tua dan Gender')
plt.savefig('bar_dukunganOrtu_Gender.png')
plt.close()
```

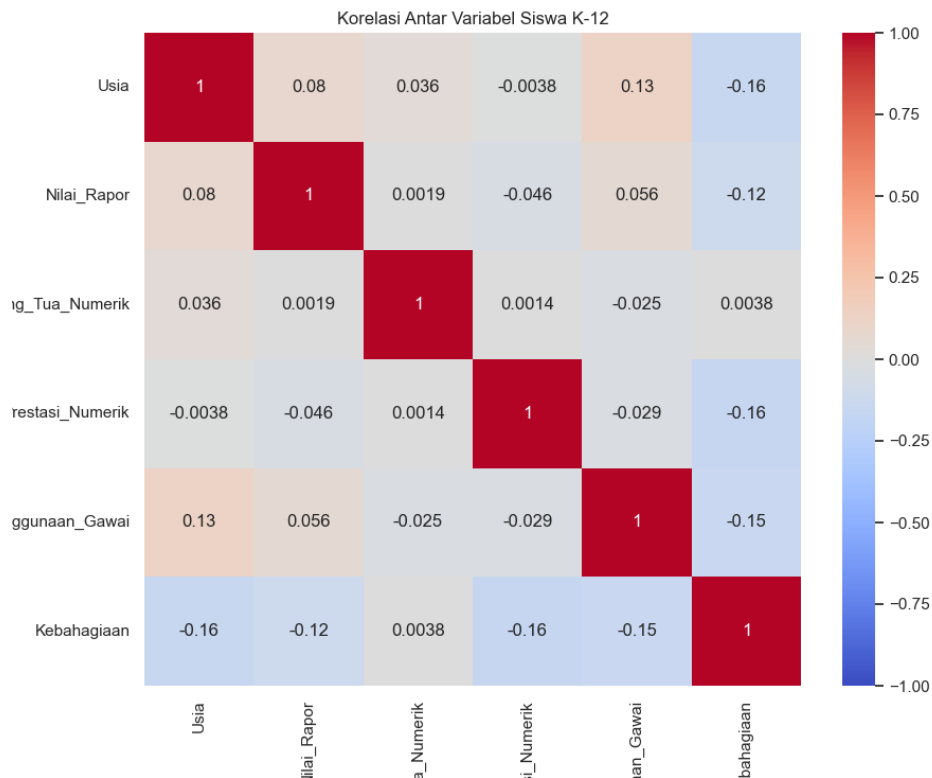
Saya menggunakan visualisasi data Bar plot untuk melihat bagaimana rata-rata nilai rapor siswa sesuai dengan dukungan orang tua. Visualisasi ini dapat secara menunjukkan nilai rata-rata secara langsung sehingga kita tidak perlu mengidentifikasinya lagi.



2.2.3

```
# 3. Seaborn Heatmap untuk korelasi antar variabel siswa K-12
plt.figure(figsize=(10, 8))
corr_siswa = df_siswa_k12[['Usia', 'Nilai_Rapor',
'Dukungan_Orang_Tua_Numerik',
'Prestasi_Numerik',
'Durasi_Penggunaan_Gawai', 'Kebahagiaan']].corr()
sns.heatmap(corr_siswa, annot=True, cmap='coolwarm', vmin=-1,
vmax=1)
plt.title('Korelasi Antar Variabel Siswa K-12')
plt.savefig('heatmap_Siswa_korelasi.png')
plt.close()
```

Saya menggunakan visualisasi ini untuk melihat variabel-variabel apa yang saling terikat atau berkorelasi sehingga dapat lebih mudah mengidentifikasi masalah dan solusi.



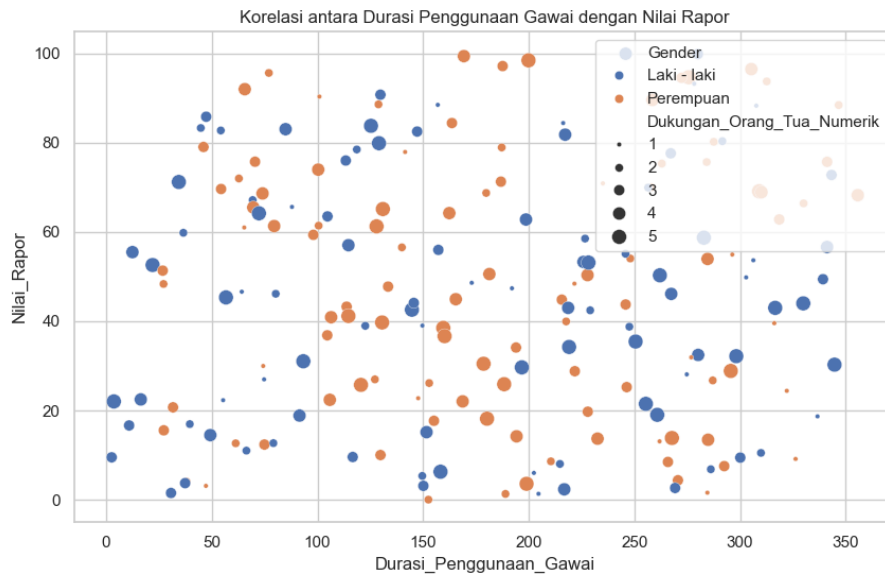
2.3 Mengustomisasi Visual durasi vs Nilai

```
# 4. Custom Visualisasi durasi vs nilai
plt.figure(figsize=(10, 6))
sns.scatterplot(data=df_siswa_k12, x='Durasi_Penggunaan_Gawai', y =
'Nilai_Rapor', hue = 'Gender' , size = 'Prestasi_Numerik', sizes
=(10,100), palette='Set2')

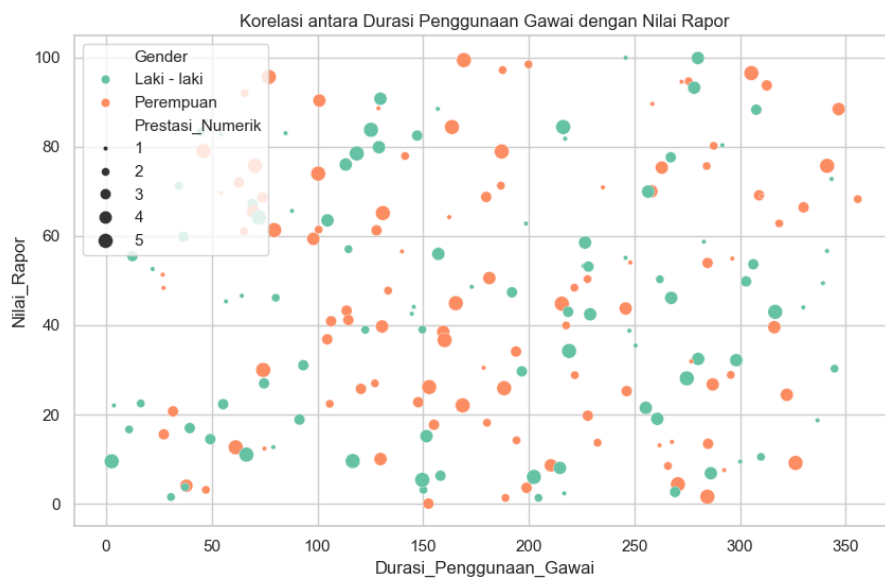
plt.title("Korelasi antara Durasi Penggunaan Gawai dengan Nilai
Rapor")
plt.savefig("scatter_siswaK12_Rapor_Durasi_Gawai_Custom.png")
plt.close()
```

Saya memodifikasi ulang visualisasi tentang durasi vs nilai rapor dengan mengubah dukungan orang tua menjadi prestasi dan menambahkan pallete warna untuk mengubah dari visual.

Sebelum:



Sesudah:



LESSON LEARN

Secara teknis praktikum kali ini membuat saya semakin paham terkait cara memvisualisasikan data dan menentukan visual yang sesuai dengan kasus tertentu.

LINK GITHUB: [https://github.com/Rasyiq603011/Visualisasi Data Python](https://github.com/Rasyiq603011/Visualisasi_Data_Python)