

MODUL PRAKTIKUM

IS411 – DATA MODELING
PROGRAM SARJANA S1 SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

Gedung B Lantai 5, Kampus UMN
Jl. Scientia Boulevard, Gading Serpong, Tangerang, Banten-15811 Indonesia
Telp: +62-21.5422.0808 (ext. 1803), email: ict.lab@umn.ac.id, web: umn.ac.id

MODUL 2

INTRODUCTION TO DATA MINING



DESKRIPSI TEMA

1. Data Storytelling case study.
2. Using Matplotlib in Python to generate basic visualizations.

CAPAIAN PEMBELAJARAN MINGGUAN (SUB-CAPAIAN PEMBELAJARAN)

Student are able to explain the concept of data mining (C2).

PENUNJANG PRAKTIKUM

1. Anaconda Navigator
 2. Jupyter Notebook
- (+ Perlengkapan Apd/Alat Pelindung Diri Yang Harus Digunakan, Jika Ada)

LANGKAH-LANGKAH PRAKTIKUM

Import Library

1. Lakukan langkah yang sama seperti minggu sebelumnya untuk meng-import library Numpy dan Pandas.

Membaca Data

2. Gunakan file pendukung yang diberikan, lalu masukkan menjadi dataframe. Kali ini sumber data adalah file excel.

```
1 dataSales = pd.read_excel('sales_data_sample.xlsx', engine = 'openpyxl', header=0)
2 #engine = 'openpyxl' ditambahkan jika muncul error "Excel xlsx file; not supported"
3 dataSales
```

Membuat Visualisasi Data dengan Library Matplotlib

3. Pilih data yang digunakan untuk menampilkan *chart* pada contoh pembuatan *chart* sederhana, data yang digunakan adalah data penjualan per tahun.
Lakukan grouping jika diperlukan.

```
1 #mengambil data penjualan per tahun
2 SalesperYr = dataSales[['YEAR_ID', 'SALES']].groupby(['YEAR_ID'])['SALES'].sum().reset_index()
3 SalesperYr['YEAR_ID'] = SalesperYr['YEAR_ID'].astype('string')
```

4. Untuk membuat *chart* sederhana membutuhkan library **Matplotlib**. Matplotlib adalah library yang memiliki function khusus untuk membuat *chart* atau diagram. Untuk menggunakannya perlu meng-import library tersebut.

```
1 import matplotlib.pyplot as plt
```

5. Salah satu *chart* yang bisa digunakan adalah *bar chart*. Berikut adalah contoh penggunaan *bar chart* untuk menampilkan data penjualan setiap tahun.

```
1 #menampilkan data penjualan per tahun dengan bar chart
2 plt.bar(SalesperYr['YEAR_ID'],SalesperYr['SALES'],color=('blue','red','yellow'))
3 plt.title('Sales per Year')
4 plt.ylabel('Sales')
5 plt.show()
```

6. Contoh penggunaan *chart* lainnya jika menggunakan data sales per negara menggunakan *bar chart*.

```
1 #menampilkan data penjualan per negara
2
3 SalesperCnt = dataSales[['COUNTRY', 'SALES']].groupby(['COUNTRY'])['SALES'].sum().reset_index()
4
5 plt.bar(SalesperCnt['COUNTRY'],SalesperCnt['SALES'])
6 plt.title('Sales per Country')
7 plt.xlabel('Country')
8 plt.ylabel('Sales')
9 plt.show()
```

7. Perhatikan hasil output. Ternyata data penjualan per negara akan terlihat sangat tidak rapi dan sulit dibaca jika menggunakan *bar chart*, karena itu cobalah menggunakan *pie chart* untuk menampilkan persentase penjualan dari setiap negara.

```
1 #menampilkan data dalam bentuk pie chart
2
3 plt.pie(SalesperCnt['SALES'], labels=SalesperCnt['COUNTRY'], autopct='%1.1f%%', startangle=90)
4 plt.show()
```

8. Lakukan perubahan pada data dan tambahkan atribut yang membantu *chart* bisa ditampilkan lebih informatif dan lebih mudah untuk dibaca oleh user. Misalnya dengan menampilkan hanya 5 negara dengan penjualan tertinggi.

```
1 #menampilkan negara dengan top 5 sales
2
3 top5sales = SalesperCnt.nlargest(5,['SALES'])
4
5 plt.pie(top5sales['SALES'], labels=top5sales['COUNTRY'], autopct='%1.1f%%', startangle=90, explode=(0.1, 0, 0, 0, 0))
6 plt.title('Top 5 Sales')
7 plt.show()
```

9. *Chart* lain yang dapat dibuat menggunakan Matplotlib adalah *line chart*. *Line chart* dapat digunakan untuk melihat pergerakan data pada periode tertentu. Sebagai contoh, data yang akan digunakan adalah data jumlah penjualan setiap bulan.

```
1 #mendapatkan data jumlah sales per bulan
2
3 dataSales['MonthYear'] = pd.to_datetime(dataSales['YEAR_ID'].astype('string')+dataSales['MONTH_ID'].astype('string'),
4                                     format='%Y%m')
5
6 SalesMY = dataSales[['MonthYear', 'SALES']].groupby(['MonthYear'])['SALES'].sum().reset_index()
7 SalesMY
```

```

1 #menampilkan tingkat sales per bulan menggunakan line chart
2
3 plt.plot(SalesMY['MonthYear'],SalesMY['SALES'])
4 plt.title('Sales by Month Year')
5 plt.ylabel('Sales')
6 plt.xlabel('Month-Year')

```

10. Dengan Matplotlib kita dapat membuat sebuah figure, atau sebuah kerangka untuk meletakkan *chart* pada posisi tertentu. Berikut ini cara untuk menambahkan subplot, yaitu plot pada figure dengan ukuran 2x2 (seperti tabel). Untuk *chart* pertama dan kedua akan menggunakan row 1 kolom 1 dan row 1 kolom 2, sedangkan untuk *chart* ketiga akan menggunakan semua space dalam row 2 kolom 1 dan 2.

```

1 #menampilkan beberapa chart dan mengatur posisi tampilan chart
2
3 fig = plt.figure(figsize=(15, 15))
4
5 plt.subplots_adjust(wspace= 0.25, hspace= 0.25)
6
7 #bar chart
8 ax1 = fig.add_subplot(2,2,1)
9 ax1.bar(SalesperYr['YEAR_ID'],SalesperYr['SALES'],color=('blue','red','yellow'))
10 ax1.set_title('Sales per Year')
11 ax1.set_ylabel('Sales')
12 ax1.legend(labels=SalesperYr['YEAR_ID'])
13
14 #pie chart
15 ax2 = fig.add_subplot(2,2,2)
16 ax2.pie(top5sales['SALES'], labels=top5sales['COUNTRY'], autopct='%1.1f%%', startangle=90, explode= (0.1, 0, 0, 0, 0))
17 ax2.set_title('Top 5 Sales')
18
19 #line chart
20 ax3 = fig.add_subplot(2,2,(3,4))
21 ax3.plot(SalesMY['MonthYear'],SalesMY['SALES'])
22 ax3.set_ylabel('Sales')
23 ax3.set_xlabel('Month-Year')
24 ax3.set_title('Sales by Month')
25
26 fig.show()
27

```

Challenge

11. Tampilkanlah menggunakan *chart* **data penjualan per produk**. Kreasikan bagaimana data ditampilkan dan gunakan *chart* yang menurut Anda paling sesuai dan mudah dipahami oleh pembaca sehingga informasi dapat tersampaikan dengan benar dan menarik.

PENGUMPULAN

1. File yang dikumpulkan terdiri dari:
 - a. File project (.ipynb)
 - b. File PDF berisi screenshot output dan jawaban (jika ada pertanyaan)
2. File di-compress (.zip) dan diberi nama **KODEMATAKULIAH_KELAS_NIM_NAMA_WEEK KE-XX.zip** (contoh: IS5411_A_13110310017_Monika Evelin Johan_Week-01.zip).

