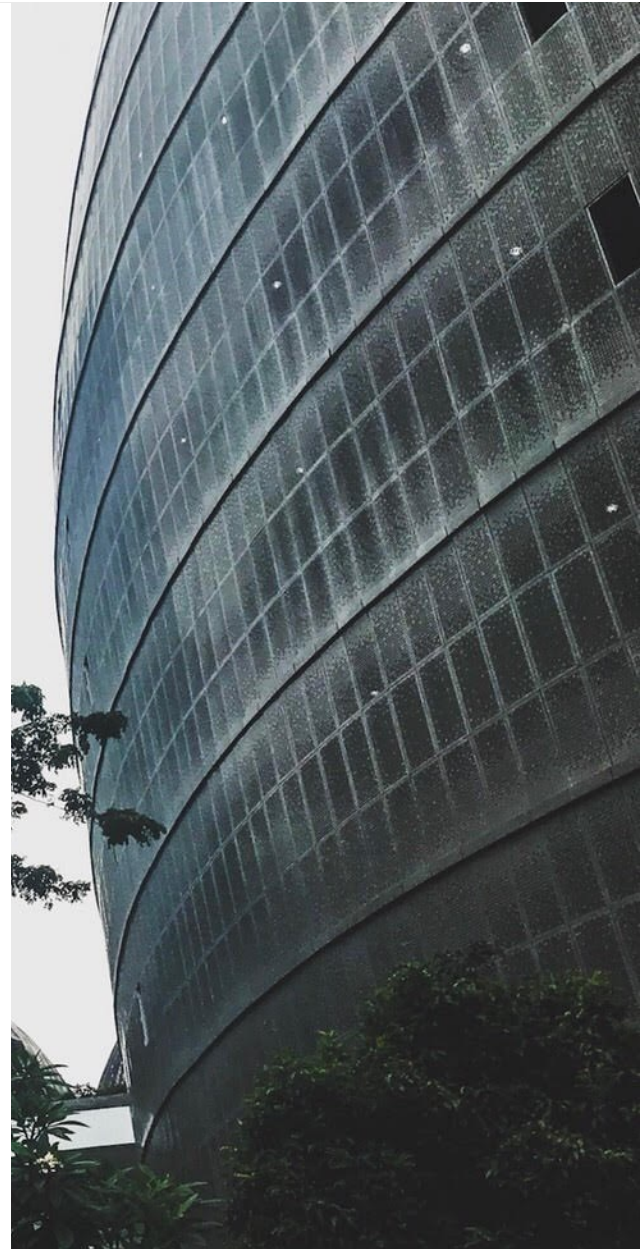


MODUL PRAKTIKUM

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 1

INTRODUCTION OF DATA SCIENCE AND DATA MINING



DESKRIPSI TEMA

1. Python Fundamental for Data Science
2. Introduction to Data Scientist with Case Study: Retail

CAPAIAN PEMBELAJARAN MINGGUAN (SUB-CAPAIAN PEMBELAJARAN)

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

PENUNJANG PRAKTIKUM

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

LANGKAH-LANGKAH PRAKTIKUM

PERSIAPAN INSTALASI

Unduh anaconda pada tautan berikut ini:

1. Tautan ofisial
 - a. anaconda

<https://www.anaconda.com/distribution/#download-section>

PS:

Pada saat penulisan modul ini, versi anaconda yang digunakan adalah versi anaconda3-2020.07

INSTALASI ANACONDA

1. Jalankan installer anaconda - Anaconda3-2020.07-Windows-x86_64.exe.
2. Setujui License Agreement-nya
3. Untuk Install for, gunakan yang direkomendasikan saja (Just Me)
4. PS: Pada komputer yang digunakan di kelas, menggunakan instalasi All Users
5. Pilih tempat instalasi sesuai yang diinginkan (WARNING: membutuhkan space sekitar 3GB dalam instalasi anaconda)
6. Pada Advanced Options, uncheck opsi Add Anaconda to my PATH environment variable, lalu tekan Install
7. Program anaconda siap dijalankan.

Cara 1:

1. Pada start menu Windows, carilah program **Anconda Prompt (Anaconda)**.
2. Pada saat **Anaconda Prompt** terbuka, Anda dapat melihat drive dan folder yang sedang aktif. Ubah posisi ke drive tempat Anaconda terinstal, lalu pindah ke folder **Anaconda\Scripts**.

Contoh:

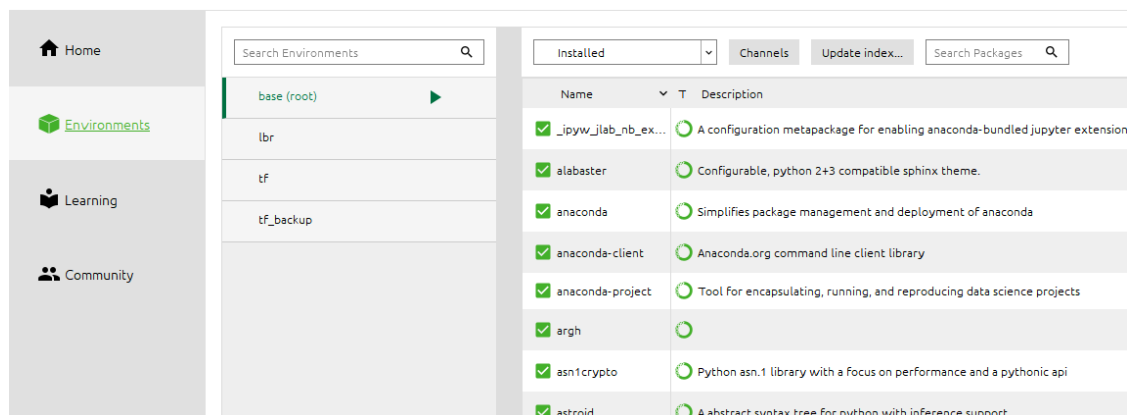
```
Anaconda Prompt (Anaconda)

(base) C:\Users\Admin> D:

(base) D:\>cd Anaconda/Scripts

(base) D:\Anaconda\Scripts>
```

3. Buatlah sebuah *environment* baru untuk melakukann instalasi *library TensorFlow*. Gunakan *command* berikut: **conda create -n <nama_env> tensorflow**
4. Prompt akan melakukan instalasi *library tensorflow* pada *environment* baru tersebut. Input “y” pada saat program menanyakan **Proceed? (y/n)**
5. Setelah proses instalasi selesai, lakukan aktivasi environment yang sudah dibuat dengan *command* berikut: **conda activate <nama_env>**
6. Anda dapat melihat daftar *environment* yang ada pada Anaconda Prompt (tab Environment) dan Environment mana yang sedang aktif.



7. Untuk menginstal *library* lain yang dibutuhkan, dapat menggunakan *command* berikut: **conda install --name <nama_env> <nama package>**

Cara 2:

Instal melalui Anaconda Navigator.

Environment → Not installed → cari package “Tensorflow”.

Jupyter Notebook

8. Pada start menu Windows, bukalah Anaconda Navigator.
9. Pilih “install” pada Jupyter Notebook. Tunggu hingga proses instalasi selesai, sampai tulisan “install” menjadi “Launch”.
10. Pastikan Environments sudah yang terinstall Tensorflow
11. Launch Jupyter Notebook yang sudah diinstal. Pada web browser akan muncul tampilan Jupyter Notebook.

Mengenal Basic Python

12. Men-define variable dengan tipe String dan Numeric.

```
1 #define variable
2 text = 'test'
3 angka = 5
4 angka2 = 5.5
```

13. Print value yang sudah ditentukan di awal.

```
1 #print variable value
2 print(text)
3 print(angka+angka2)
```

14. Membuat percabangan IF..ELSE..

```
1 #Percabangan if else
2 if angka+angka2 <= 10:
3     print(10)
4 else:
5     print(angka+angka2)
```

15. Membuat function di Python.

```
1 #Membuat function
2 def penjumlahan(x,y):
3     hasil = x+y
4     return(hasil)
5
6 test = penjumlahan(angka,angka2)
7 print(test)
```

16. Menerima input user.

```
1 #Menerima input dari user dengan tipe data int
2 angka = int(input('angka baru :'))
3 angka2 = int(input('angka 2 baru : '))
4 test = penjumlahan(angka,angka2)
5 print(test)
```

17. Melakukan konversi tipe data.

```
1 #konversi number to string
2 angka_1 = 12
3 angka_2 = 20
4 print (angka_1 + angka_2)
5 print(str(angka_1) + str(angka_2))
```

18. Melakukan looping.

```
1 #melakukan looping pada basic python
2 test_loop = 0
3
4 while(test_loop < 10):
5     print(test_loop)
6     test_loop += 1
7
```

Mengenal Basic Numpy

NumPy (Numerical Python) adalah library Python yang fokus pada scientific computing. NumPy memiliki kemampuan untuk membentuk objek N-dimensional array, yang mirip dengan list pada Python. Print value yang sudah ditentukan di awal.

19. Import Numpy

```
1 import numpy as np
```

20. Membuat array Numpy sederhana

```
1 #membuat array numpy sederhana
2 a = np.array([[1,2,3],
3              [4,5,6]])
```

21. Array re-dimensioning

```
1 #array re-dimensioning
2 a = np.array([x for x in range(27)])
3
4 npshape = a.reshape((3,3,3))
5
6 print(npshape)
```

22. Stacking Numpy array

```
1 #stacking numpy array
2 a1 = np.array([[1,2,3],
3               [4,5,6]])
4
5 a2 = np.array([[7,8,9],
6               [10,11,12]])
7
8 npstack = np.hstack((a1, a2))
9
10 print(npstack)
```

23. Membuat array numpy dari angka 0 – 100 dengan jarak 2 angka

```
1 #membuat array numpy dari angka 0 - 100 dengan jarak 2 angka
2 list_of_numbers = [x for x in range(0, 101, 2)]
3
4 listnp = np.array(list_of_numbers)
5
6 print(listnp)
```

```
1 #cara alternatif
2 npalt = np.arange(0, 101, 2)
3
4 print(npalt)
```

24. Membuat array numpy dengan random value

```
1 #membuat array numpy dengan random value
2 np.random.seed(123) # setting the seed
3
4 nprand = np.random.randint(0, 10, size = (5,5))
5
6 print(nprand)
```

Mengenal Pandas

Pandas (Python for Data Analysis) adalah library Python yang fokus untuk proses analisis data seperti manipulasi data, persiapan data, dan pembersihan data. Pandas menyediakan struktur data dan fungsi high-level untuk membuat pekerjaan dengan data terstruktur/tabular lebih cepat, mudah, dan ekspresif.

Dalam pandas terdapat dua objek yang sering digunakan, yaitu DataFrame dan Series. DataFrame adalah objek yang memiliki struktur data tabular, berorientasi pada kolom dengan label baris dan kolom. Sedangkan Series adalah objek array 1-dimensi yang memiliki label. Print value yang sudah ditentukan di awal.

25. Import library Pandas

```
1 import pandas as pd
```

26. Membaca data file yang tersedia

```
1 #membaca data dari file yang tersedia
2 data = pd.read_csv('MBA_Python.csv', header=0)
3
4 data.info()
```

Study Kasus Market Basket Analysis

Persiapan data untuk market basket analysis.

27. Dalam mempersiapkan data yang akan digunakan terutama dalam machine learning, data harus di proses agar tidak ada value yang NaN atau Null. Salah satu cara untuk menghilangkan missing value tersebut, dapat menggunakan function dropna() dari pandas. Print value yang sudah ditentukan di awal.

```
1 data.dropna()
```

28. Dalam membuat market basket analysis, pastikan data yang digunakan bernilai positif untuk memaksimalkan analisa hubungan antar barang dalam basket. Salah satunya dengan menggunakan function min(). Kemudian hapus data yang bernilai negatif.dengan menambahkan filter pada dataset.

```
1 data['Quantity'].min()
2
3 dataPositif = data[data['Quantity']>=0]
4 dataPositif['Quantity'].min()
5
```

29. Untuk membuat basket dari data yang ada, dapat menggunakan group by. Beberapa function untuk mendukung group by adalah unstack() untuk menghilangkan data yang menjadi pivot akibat group by, sehingga tampilan data tetap dalam bentuk tabular. reset_index() menghapus index dari data. fillna() mengisi data yang valuenya NaN dengan value lain yang sudah ditentukan. pada praktikum ini contohnya angka 0. set_index() menambahkan index baru ke dalam table.

```
basket = (dataPositif.groupby(['Invoice','Description'])['Quantity'].sum().unstack()
          .reset_index().fillna(0).set_index(['Invoice']))
basket
```

30. Dalam menganalisa market basket, data yang harus diperhatikan adalah data barang apa saja yang dibeli oleh customer dalam satu basket. Banyaknya jumlah barang yang dibeli tidak memiliki hubungan yang signifikan dengan analisa. karena itu untuk memudahkan proses analisa dari basket penjualan, data dapat disederhanakan menjadi 0 dan 1. dimana 1 menandakan barang tersebut ada dalam basket dan 0 berarti barang tersebut tidak ada.

```
def encode(qty) :
    if qty >= 1:
        return 1
    if qty <= 0:
        return 0

basketSimplified = basket.applymap(encode)
basketSimplified
```

31. Selain jumlah data, basket yang hanya memiliki 1 data juga tidak dapat digunakan dalam market basket analisis, untuk itu data perlu difilter kembali untuk basket - basket yang hanya memiliki 1 item.

```
1 finalBasket = basketSimplified[(basketSimplified > 0).sum(axis=1) >= 2]
```

32. Untuk membuat market basket analysis, gunakan library apriori

```
1 from mlxtend.frequent_patterns import apriori
2
3 frequent_item = apriori(finalBasket, min_support=0.03, use_colnames=True)
4 .sort_values('support', ascending=False).reset_index(drop=True)
5
6 frequent_item
```

```
1 from mlxtend.frequent_patterns import association_rules
2
3 association_rules(frequent_item, metric='lift', min_threshold=1)
4 .sort_values('lift', ascending=False).reset_index(drop=True)
```

REFERENSI

Bengfort B., Bibro R., & Ojeda T. (2018). Applied text analysis with Python. O'Reilly Media.
 Deitel, P., & Deitel, H. (2020). Intro to Python for Computer Science and Data Science. Pearson Education.