Module Python for Data Science

Prerequirements : Pernah belajar Python Dasar,punya minat tinggi thd data sains

Duration per meet : 2 Hours

Curriculum Based : DataCamp and DQLAB

Learning Strategy : Interactive Learning with Simple Explanation

Output Peserta yg diharapkan :

* dapat menggunakan google collab/kaggle untuk data analisis pada suatu dataset sederhana
* dapat melakukan EDA (handle missing value,outlier,duplicated,noise,drop unused column,cek korelasi per kolom)
* dapat menggunakan pandas,numpy,matplotlib pada kasus dataset sederhana
* dapat menggunakan seaborn /plotly untuk visualisasi data
* memahami konsep machine learning (unsupervised vs supervised ,makna train test split,evaluation metric)
* dapat menemukan insight dari sebuah dataset sederhana menggunakan metode clustering
* dapat menemukan insight dari sebuah dataset sederhana menggunakan metode regresi linear
* dapat menemukan insight dari sebuah dataset sedehana menggunakan metode klasifikasi

**13 July (Selasa)**

Data Science for Everyone

- perkenalan mentor,alasan kita wajib belajar data science

- Perbedaan data science ,data engineering,machine learning ,businesss intelligent,dan data mining

- info menarik seputar data sains ,metode pembelajaran yang akan dilakukan

- pengenalan tentang tools data science (google collab,kaggle,jupyter notebook)

- Teaser for next meet ( Apa yg perlu dipelajari next meet)

**14July (Rabu) 2 hours**

Remind Python

**-** pengertian variabel,jenis tipe data(numerik : int,float64,kategori: bool,object)

- Pengenalan list,dictionary,set,tuple,looping,function, dan lambda function

- kasih file cheatseet python data science toolbox

- peng

- Teaser for next meet

**15 July (Kamis) 2 hours**

Numpy (frequently used syntax)

- Why we use Numpy (kelebihan)

- 2D Numpy Arrays meaning

- pengenalan nd-array

- Contoh penggunaan numpy ( menentukan median,mean)

- Teaser for next meet

**16 July (Jum’at)** 2 hours (interaktif and easy explanation)

Data Manipilation with Pandas 1 (Frequently Used Syntax)

1. Why we use Pandas
2. Pengenalan dataframes dan series

index = data yg di observasi

kolom = variabel dalam observasi

1. Import file csv ,Penggunaan head(),tail(),sample()

**19 July (Senin)** 2 hours

1. Penggunaan simbol kurung siku “[]” pada pandas untuk select kolom dan subsetting
2. Memilih dan memisah data dengan sintaks iloc/loc

**20 July (Selasa)**

Data Manipulation with Pandas 2 (Frequently Used Syntax)

1. Add,Replace,Copy ,Rename column in pandas
2. Menampilkan hanya 2 buah kolom pada dataset di pandas
3. Membuat data frame dari suatu series/kolom
4. Variabel (predictor ,target ) meaning
5. Sort Values untuk filter data
6. Penggunaan GroupBy pada dataframes
7. Menggabungkan 2 buah dataframes dengan pd.concat()

**21 july (rabu)**

Basic Exploratory data analysis(Teory)

1. pengertian duplicates values

2. memahami ciri kolom yg sebaiknya di hapus pada dataframe :

- tidak mempunyai korelasi yg kuat terhadap variabel pengamatan utama

- datanya sama semua

- datanya kurang bervariasi

- data yg berupa index

- data yg missing valuenya lebih dari 50%

3. pengertian oulier dan noise terus apa perbedaanya

4. pengenalan distribusi normal

**22 july (kamis)**

EDA with pandas

1. sintaks cek duplicates

2. menghapus kolom yang kurang berguna

3. memisahkan outlier dengan IQR

4. Transforming data

5. Normalisasi data (minmax scaler,label encoder,pd.dummies)

**26 july (senin)**

Matplotlib(frequently used syntax)

- Why we use matplotlib

- Sintaks dasar matplotlib

- Basics Line plot (include : legend,title,ticks,x\_lim,plot,show,figsize) meaning for every sintaks

- Basics Scatter plot (include : legend,title,ticks,x\_lim,plot,show,figsize)

- Build Histogram (include : legend,title,ticks,x\_lim,plot,show,figsize)

- Teaser for next meet

**27 july (selasa)**

Visualiasi data with matplotlib,seaborn and plotly

* Teknik pengambilan insight dari suatu diagram
* Scatterplot
* Line plot
* Catplot seaborn

**28 july (rabu)**

Visualiasi data with matplotlib,seaborn and plotly

* Histogram(distribusi data normal/tidak)
* Countplot
* Distplot
* Boxplot dengan seaborn /plotly(cek outlier)

**29 July (kamis)**

Supervised and unsupervised machine learning

* Pengenalan tentang machine learning (easy explanation)

Supervised ada label unsupervised gaada

* Show cheatsheet model machine learning
* Penjelasan mengenai plus minus suatu model terhadap penerapanya dalam data
* Konsep train test split

**2 Agustus (Senin)**

**Clustering**

* Pengenalan clustering sederhana (unsupervised)
* Penggunaan kmeans untuk clustering
* Elbow method dan kmeans++

**3 Agustus (Selasa )**

**Regresi linear sederhana**

* Penggunaan model sklearn linear regression

**4 Agustus (Rabu)**

**Praktek klasifikasi sederhana**

Pemberian tugas akhir yang nanti akan dikumpulkan dan dibahas di akhir kelas

**5 Agustus (Kamis)**

**9 Agustus (Senin)**

**10 Agustus (Selasa)**

**11 Agustus (Rabu)**

**12 Agustus (Kamis)**