

PENGENALAN PYTHON
PRAKTIKUM PEMBELAJARAN MESIN



Disusun Oleh:

Nashirudin Baqiy

24060119130045

Lab A2

DEPARTEMEN ILMU KOMPUTER / INFORMATIKA

FAKULTAS SAINS DAN MATEMATIKA

UNIVERSITAS DIPONEGORO

2021

BAB I PENDAHULUAN

1.1. Rumusan Masalah

- 1.1.1. Cari dataset dataset lain dari repository UCI Machine Learning!
- 1.1.2. Lakukan keenam langkah untuk menampilkan summary dataset!

1.2. Tujuan

- 1.2.1. Mencari dan menemukan dataset di repository UCI machine learning
- 1.2.2. Melakukan enam langkah preprocessing data hingga menampilkan summary dataset.

1.3. Dasar Teori

Pembelajaran mesin (ML) adalah studi tentang algoritma komputer yang dapat ditingkatkan secara otomatis melalui pengalaman dan penggunaan data. Itu dilihat sebagai bagian dari kecerdasan buatan. Algoritma pembelajaran mesin membangun model berdasarkan data sampel, yang dikenal sebagai "data pelatihan", untuk membuat prediksi atau keputusan tanpa diprogram secara eksplisit untuk melakukannya. Algoritma pembelajaran mesin digunakan dalam berbagai macam aplikasi, seperti dalam kedokteran, penyaringan email, pengenalan suara, dan visi komputer, di mana sulit atau tidak mungkin untuk mengembangkan algoritma konvensional untuk melakukan tugas-tugas yang diperlukan.

Bagian dari pembelajaran mesin terkait erat dengan statistik komputasi, yang berfokus pada pembuatan prediksi menggunakan komputer; tetapi tidak semua pembelajaran mesin adalah pembelajaran statistik. Studi optimasi matematika memberikan metode, teori, dan domain aplikasi ke bidang pembelajaran mesin. Data mining adalah bidang studi terkait, dengan fokus pada analisis data eksplorasi melalui pembelajaran tanpa pengawasan. Beberapa implementasi pembelajaran mesin menggunakan data dan jaringan saraf dengan cara yang meniru kerja otak biologis. Dalam penerapannya di seluruh masalah bisnis, machine learning juga disebut sebagai analitik prediktif.

BAB II PEMBAHASAN

2.1. Deskripsi Dataset

Pada praktikum kali ini, dataset yang akan digunakan adalah dataset dari UCI yang berisi data mengenai prediksi pasar saham dengan 9357 baris dan 15 kolom. File dataset yang digunakan bernama AirQualityUCI.csv

```
      Date      Time CO(GT) PT08.S1(CO) ... PT08.S5(O3)      T      RH      AH
0      3/10/2004 18:00:00      2.6      1360 ...      1268      13.6      48.9      0.7578
1      3/10/2004 19:00:00      2.0      1292 ...      972      13.3      47.7      0.7255
2      3/10/2004 20:00:00      2.2      1402 ...      1074      11.9      54.0      0.7502
3      3/10/2004 21:00:00      2.2      1376 ...      1203      11.0      60.0      0.7867
4      3/10/2004 22:00:00      1.6      1272 ...      1110      11.2      59.6      0.7888
...      ...      ...      ...      ...      ...      ...      ...      ...
9352      4/4/2005 10:00:00      3.1      1314 ...      1729      21.9      29.3      0.7568
9353      4/4/2005 11:00:00      2.4      1163 ...      1269      24.3      23.7      0.7119
9354      4/4/2005 12:00:00      2.4      1142 ...      1092      26.9      18.3      0.6406
9355      4/4/2005 13:00:00      2.1      1003 ...      770      28.3      13.5      0.5139
9356      4/4/2005 14:00:00      2.2      1071 ...      816      28.5      13.1      0.5028
```

[9357 rows x 15 columns]

```
      Date      Time CO(GT) PT08.S1(CO) ... PT08.S5(O3)      T      RH      AH
0      False      False      False      False ...      False      False      False      False
1      False      False      False      False ...      False      False      False      False
2      False      False      False      False ...      False      False      False      False
3      False      False      False      False ...      False      False      False      False
4      False      False      False      False ...      False      False      False      False
...      ...      ...      ...      ...      ...      ...      ...      ...
9352      False      False      False      False ...      False      False      False      False
9353      False      False      False      False ...      False      False      False      False
9354      False      False      False      False ...      False      False      False      False
9355      False      False      False      False ...      False      False      False      False
9356      False      False      False      False ...      False      False      False      False
```

[9357 rows x 15 columns]

NaN value: 0

```
Date      0
Time      0
CO(GT)    0
PT08.S1(CO) 0
NMHC(GT)  0
C6H6(GT)  0
PT08.S2(NMHC) 0
NOx(GT)   0
PT08.S3(NOx) 0
NO2(GT)   0
PT08.S4(NO2) 0
PT08.S5(O3) 0
T         0
RH        0
AH        0
dtype: int64
```

Jika dilihat lebih teliti lagi, maka dataset ini mengandung banyak sekali missing value atau NaN value 0 nilai data jadi tidak menghilangkan beberapa fiturnya.

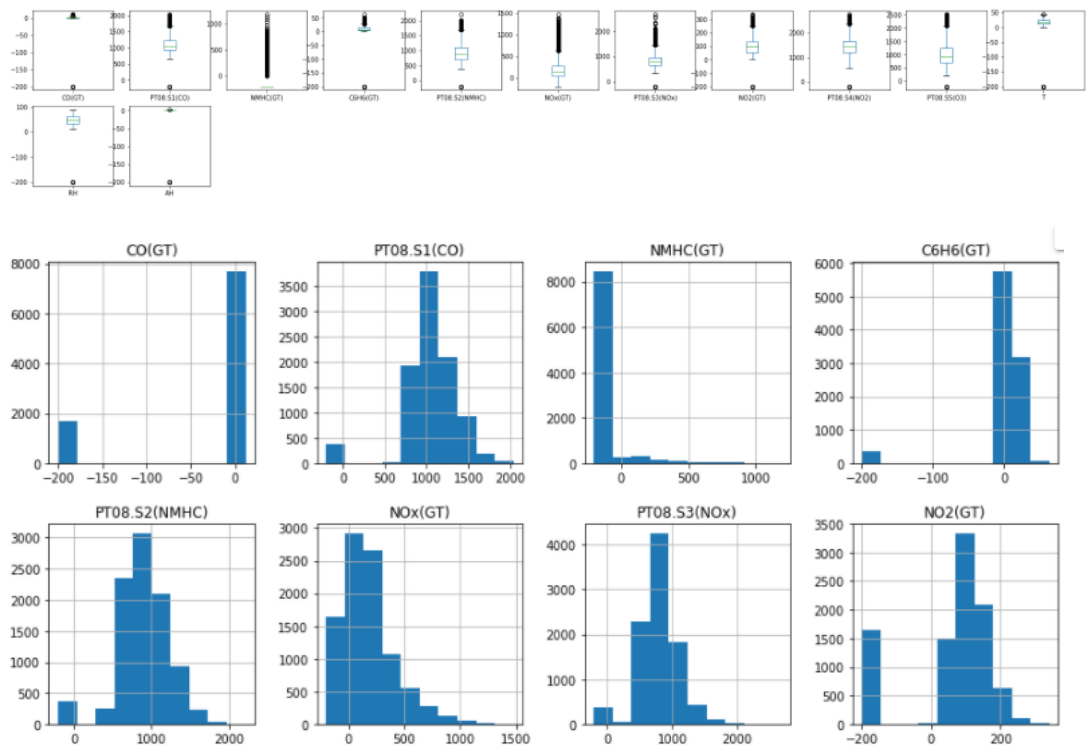
2.2. Statistik

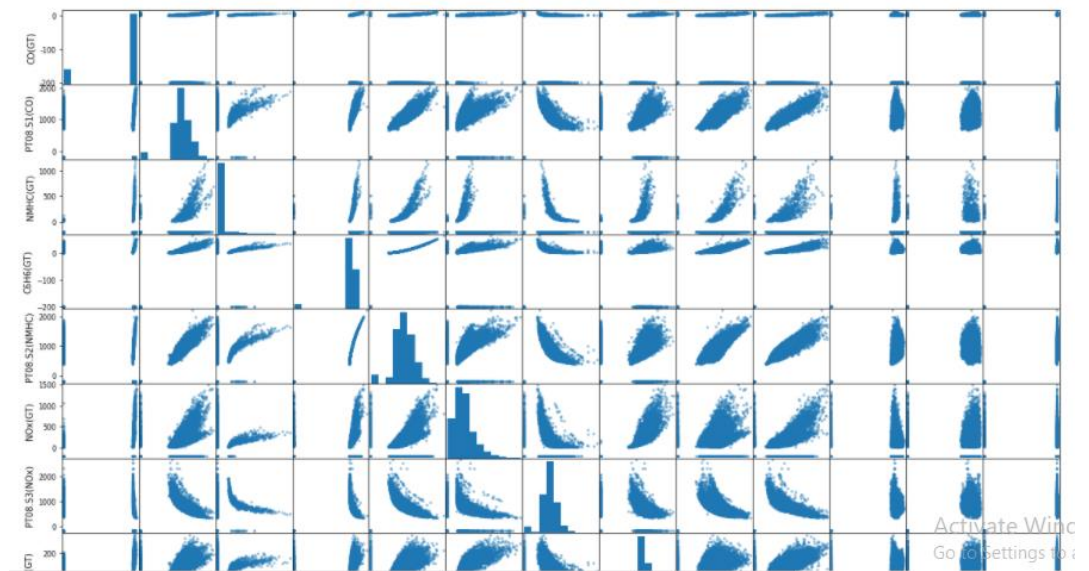
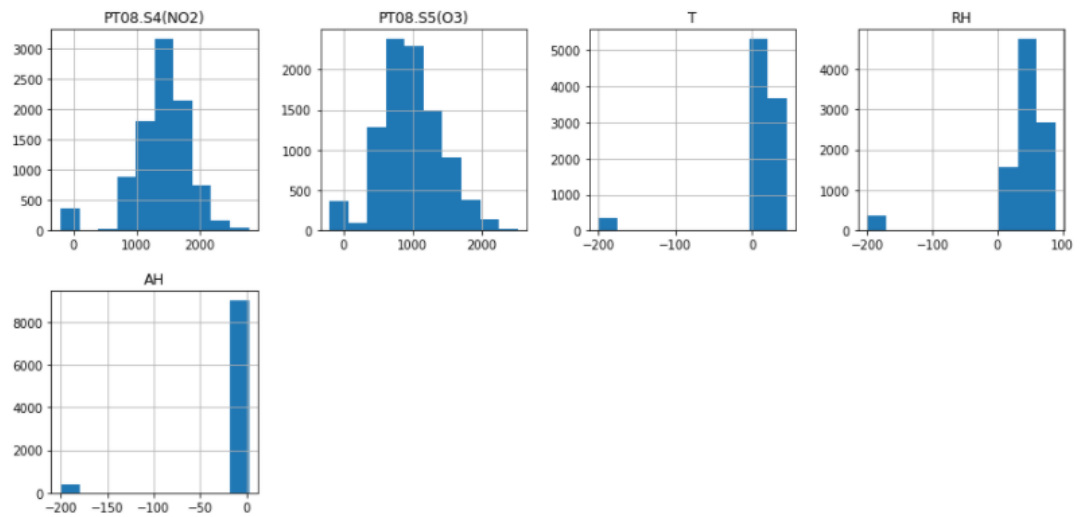
Berikut karakteristik secara statistik dari dataset ini

	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	NOx(GT)	PT08.S3(NOx)	NO2(GT)	PT08.S4(NO2)	PT08.S5(O3)
count	9357.000000	9357.000000	9357.000000	9357.000000	9357.000000	9357.000000	9357.000000	9357.000000	9357.000000	9357.000000
mean	-34.207524	1048.990061	-159.090093	1.865683	894.595276	168.616971	794.990168	58.148873	1391.479641	975.072032
std	77.657170	329.832710	139.789093	41.380206	342.333252	257.433866	321.993552	126.940455	467.210125	456.938184
min	-200.000000	-200.000000	-200.000000	-200.000000	-200.000000	-200.000000	-200.000000	-200.000000	-200.000000	-200.000000
25%	0.600000	921.000000	-200.000000	4.000000	711.000000	50.000000	637.000000	53.000000	1185.000000	700.000000
50%	1.500000	1053.000000	-200.000000	7.900000	895.000000	141.000000	794.000000	96.000000	1446.000000	942.000000
75%	2.600000	1221.000000	-200.000000	13.600000	1105.000000	284.000000	960.000000	133.000000	1662.000000	1255.000000
max	11.900000	2040.000000	1189.000000	63.700000	2214.000000	1479.000000	2683.000000	340.000000	2775.000000	2523.000000

2.3. Visualisasi

Tahap terakhir dari preprocessing kali ini adalah melakukan plotting grafik dengan univarian dan multivarian





Activate Windows
Go to Settings to activate Windows.

BAB III KESIMPULAN

3.1. Kesimpulan

Pada praktikum pertemuan pertama di mata kuliah pembelajaran mesin kali ini, mempelajari mengenai dasar Python untuk machine learning khususnya untuk preprocessing dataset. Dalam belajar mengenai machine learning di praktikum kali ini, dipergunakan beberapa library Python guna memudahkan dalam melakukan preprocessing, seperti Pandas dan Numpy untuk mengolah array dan matriks dan Matplotlib sebagai library untuk membuat grafik dari data yang ada. Dataset yang digunakan adalah dataset mengenai kualitas udara prediction (*AirQuality*) yang didapat dari situs web UCI

DAFTAR PUSTAKA

S. De Vito, E. Massera, M. Piga, L. Martinotto, G. Di Francia, On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario, *Sensors and Actuators B: Chemical*, Volume 129, Issue 2, 22 February 2008, Pages 750-757, ISSN 0925-4005

Saverio De Vito, Marco Piga, Luca Martinotto, Girolamo Di Francia, CO, NO₂ and NO_x urban pollution monitoring with on-field calibrated electronic nose by automatic bayesian regularization, *Sensors and Actuators B: Chemical*, Volume 143, Issue 1, 4 December 2009, Pages 182-191, ISSN 0925-4005,

S. De Vito, G. Fattoruso, M. Pardo, F. Tortorella and G. Di Francia, 'Semi-Supervised Learning Techniques in Artificial Olfaction: A Novel Approach to Classification Problems and Drift Counteraction,' in *IEEE Sensors Journal*, vol. 12, no. 11, pp. 3215-3224, Nov. 2012. doi: 10.1109/JSEN.2012.2192425