

Nama : 1. Ari Winanto 1815025027
2. Agung Budi Mulyono 1815025004

Matkul : Data Mining

Kelas : Data Mining C 2018

Dosen Pengampu: Dr. Anindita Septiarini, M.Cs
Tugas II Histogram

1. Tujuan

1. Untuk dapat menentukan sebaran data yang berbentuk grafik dalam sebuah diagram.

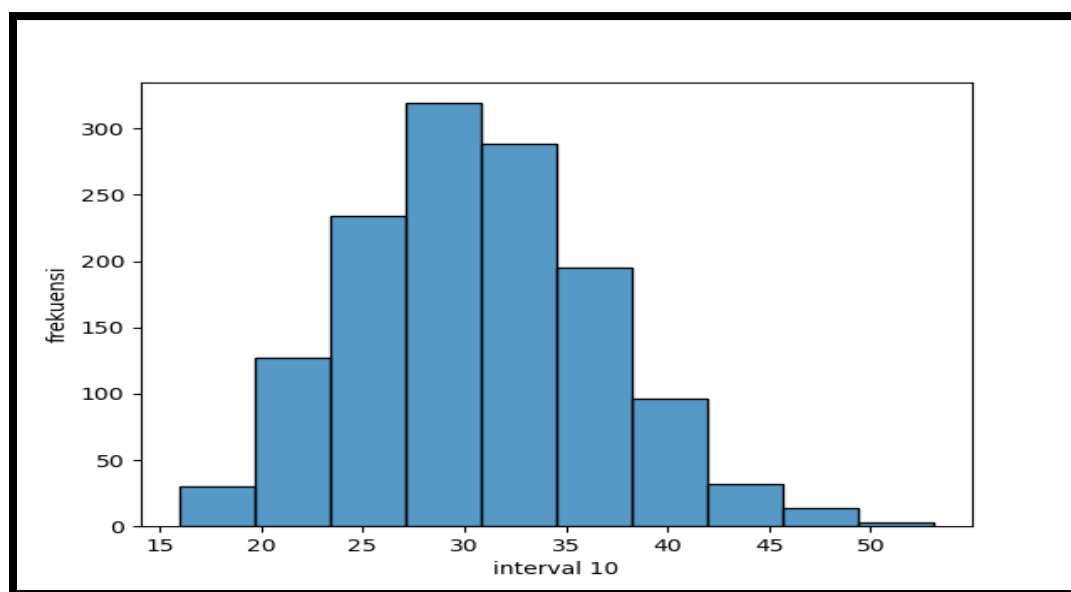
2. Dasar Teori

Histogram adalah tampilan grafis dari tabulasi frekuensi yang digambarkan dengan grafis batangan sebagai manifestasi data binning.

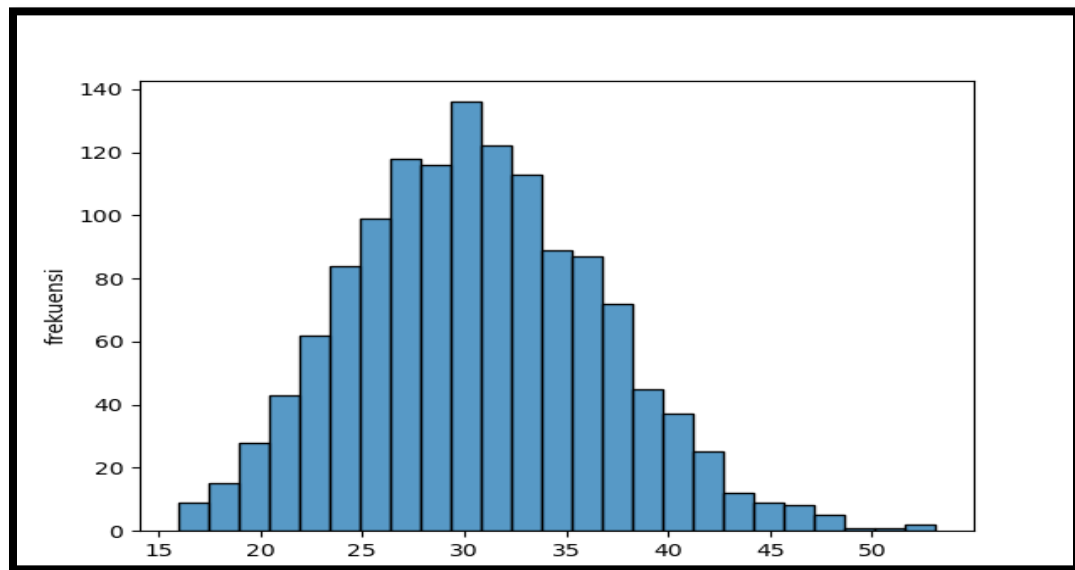
Dalam tugas ini kami menggunakan data yang bersumber dari Kaggle.com dengan keyword insurance.csv, dengan jumlah data sebanyak 1339 data yang digunakan memuat berupa body mass index untuk histogram dan data age (Umur) dan body massa index(BMI) untuk scatter plot.

Hasil dari histogram dan scatter plot dari data di atas menggunakan program python adalah sebagai berikut:

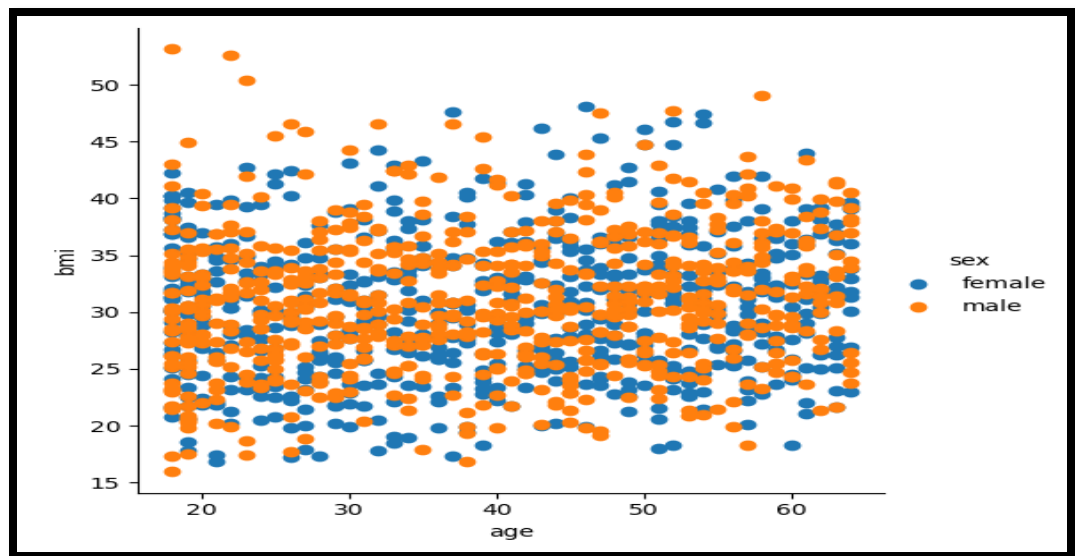
1. Histogram data kelompok dengan Interval



2. Histogram dengan data tunggal



3. Scatter Plot



4. Source code

1. Source code Histogram data kelompok

```
#Pandas library python digunakan untuk membaca file csv
import pandas as pd

#Matplotlib memvisualisasikan data secara 2D atau 3D
import matplotlib.pyplot as plt

#Visualisasi data secara statistik agar tampak lebih menarik
import seaborn as sns

#Membaca file csv dan meyimpannya pada variabel belajar
belajar = pd.read_csv("Dataset/insurance.csv")

#menampilkan semua data
belajar.head()

#memvisualisasikan distribusi data dengan interval 10
sns.histplot(belajar['bmi'],bins=10)
plt.xlabel('interval 10')
plt.ylabel('frekuensi')

#menampilkan hasil visualisasi
plt.show()
```

2. Source code histogram data tunggal

```
#Pandas library python digunakan untuk membaca file csv
import pandas as pd

#Matplotlib memvisualisasikan data secara 2D atau 3D
import matplotlib.pyplot as plt

#Visualisasi data secara statistik agar tampak lebih menarik
import seaborn as sns

#Membaca file csv dan meyimpannya pada variabel belajar
belajar = pd.read_csv("Dataset/insurance.csv")

#menampilkan semua data
belajar.head()

#memvisualisasikan distribusi data tunggal
sns.histplot(belajar['bmi'])
plt.ylabel('frekuensi')

#menampilkan hasil visualisasi
plt.show()
```

3. Source code scatter plot

```
#Pandas library python digunakan untuk membaca file csv
import pandas as pd

import warnings ##untuk mengabaikan perintah warning pada seaborn
warnings.filterwarnings("ignore")

#Matplotlib memvisualisasikan data secara 2D atau 3D
import matplotlib.pyplot as plt

#Visualisasi data secara statistik agar tampak lebih menarik
import seaborn as sns

#Membaca file csv dan menyimpannya pada variabel belajar
belajar = pd.read_csv("Dataset/insurance.csv")

#menampilkan semua data
belajar.head()

#memvisualisasikan distribusi data dengan x (age) dan y (bmi)
sns.FacetGrid(belajar, hue="sex", size=5) .map(plt.scatter, "age",
"bmi") .add_legend()

#menampilkan hasil visualisasi
plt.show()
```

4. Dataset

34 female	31.92	1 yes	northeast	37701.88
37 male	28.025	2 no	northwest	6203.902
59 female	27.72	3 no	southeast	14001.13
63 female	23.085	0 no	northeast	14451.84
55 female	32.775	2 no	northwest	12268.63
23 male	17.385	1 no	northwest	2775.192
31 male	36.3	2 yes	southwest	38711
22 male	35.6	0 yes	southwest	35585.58
18 female	26.315	0 no	northeast	2198.19
19 female	28.6	5 no	southwest	4687.797