

**MODUL 4**  
**(MINGGU KEEMPAT)**

## Praktikum 7



## DATA MINING

Data mining adalah proses penggalian atau ekstraksi informasi yang berguna dari data mentah yang besar dan kompleks. Tujuan utama dari data mining adalah untuk menemukan pola tersembunyi dan wawasan yang dapat membantu pengambilan keputusan yang lebih baik.

Proses data mining melibatkan beberapa tahap, termasuk pengumpulan data, pemrosesan data, transformasi data, pemodelan, dan evaluasi. Pada tahap pengumpulan data, data yang relevan dikumpulkan dari berbagai sumber. Kemudian, data diolah dan dibersihkan untuk menghilangkan kesalahan dan outlier. Selanjutnya, data diubah menjadi format yang dapat diproses oleh algoritma data mining.

Selanjutnya, model data mining dibangun dengan menggunakan teknik seperti clustering, klasifikasi, regresi, asosiasi, atau segmentasi. Setelah model dibangun, model dievaluasi dan diuji dengan menggunakan data yang tidak digunakan selama proses pembangunan model. Hasil evaluasi ini digunakan untuk mengukur keefektifan model dalam memprediksi atau mengidentifikasi pola. Data mining digunakan di banyak bidang, seperti bisnis, keuangan, kesehatan, dan ilmu pengetahuan. Contohnya, di bidang bisnis, data mining dapat membantu dalam identifikasi tren pasar, segmentasi pelanggan, dan analisis risiko keuangan. Sedangkan di bidang kesehatan, data mining dapat membantu dalam analisis data medis untuk mendukung diagnosis dan perawatan pasien.

Dalam ringkasannya, Data mining adalah teknik penggalian informasi yang menggunakan algoritma komputer untuk menemukan pola dan wawasan dari data yang besar dan kompleks. Hal ini dapat membantu dalam mengidentifikasi tren, pola, dan hubungan dalam data yang dapat digunakan untuk pengambilan keputusan yang lebih baik.

Berikut adalah beberapa metode data mining yang umum digunakan:

1. Klasifikasi: metode ini digunakan untuk memprediksi kelas atau label dari suatu objek berdasarkan kumpulan data yang sudah diketahui. Contoh penggunaan klasifikasi adalah untuk memprediksi apakah seorang pelanggan akan membeli produk tertentu atau tidak.

2. Regresi: metode ini digunakan untuk memprediksi nilai kontinu dari suatu variabel berdasarkan variabel lainnya. Contohnya adalah untuk memprediksi harga rumah berdasarkan ukuran, lokasi, dan fitur lainnya.
3. Asosiasi: metode ini digunakan untuk menemukan hubungan antara item dalam dataset. Contoh penggunaan asosiasi adalah untuk menemukan produk yang sering dibeli bersamaan oleh pelanggan.
4. Segmentasi: metode ini digunakan untuk membagi dataset menjadi kelompok yang lebih kecil berdasarkan karakteristik atau atribut yang sama. Contoh penggunaan segmentasi adalah untuk membagi pelanggan ke dalam kelompok berdasarkan preferensi dan perilaku mereka.
5. Klastering: metode ini digunakan untuk membagi dataset menjadi kelompok yang lebih kecil berdasarkan kesamaan karakteristik atau atribut. Contohnya adalah untuk membagi produk ke dalam kelompok berdasarkan fitur atau kriteria tertentu.
6. Pengelompokan: metode ini digunakan untuk mengelompokkan data menjadi kelas atau kelompok tertentu berdasarkan atribut yang telah ditentukan sebelumnya. Contohnya adalah untuk mengelompokkan pelanggan ke dalam kategori seperti pelanggan loyal, pelanggan berpotensi, dan sebagainya.
7. Analisis deret waktu: metode ini digunakan untuk menganalisis data yang terjadi seiring waktu, seperti tren penjualan atau perilaku pelanggan.
8. Analisis teks: metode ini digunakan untuk mengekstraksi informasi dari data teks, seperti ulasan produk atau dokumen bisnis.

Setiap metode data mining memiliki kegunaan yang berbeda dan dapat membantu mengidentifikasi pola dan wawasan yang berbeda dalam data.

**Judul : metode regresi linier sederhana**

**Deksripsi : prediksi harga rumah dengan regresi linier**

**Dataset : <https://tifupb.id/hargarumah>**

**Libray linier regesri : sklearn**

**Instalasi : pip install -U scikit-learn scipy matplotlib**

Analisis Regresi Linear adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat/dependen (Y) dengan satu atau lebih variabel bebas/independen (X). Berdasarkan banyaknya variabel bebas yang ada dalam model, Regresi Linear dibagi menjadi 2 jenis yaitu : Regresi Linear sederhana dan Regresi Linear berganda. Apabila banyaknya variabel bebas (X) hanya ada satu, maka disebut sebagai Regresi Linear sederhana. Sedangkan apabila terdapat lebih dari 1 variabel bebas (X) maka disebut sebagai Regresi Linear berganda.

1. Import library yang dibutuhkan

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn import linear_model
```

- Ambil dataset rumah dengan pandas kemudian tampilkan datanya

```
df = pd.read_csv("datarumah.csv")
print(df)
```

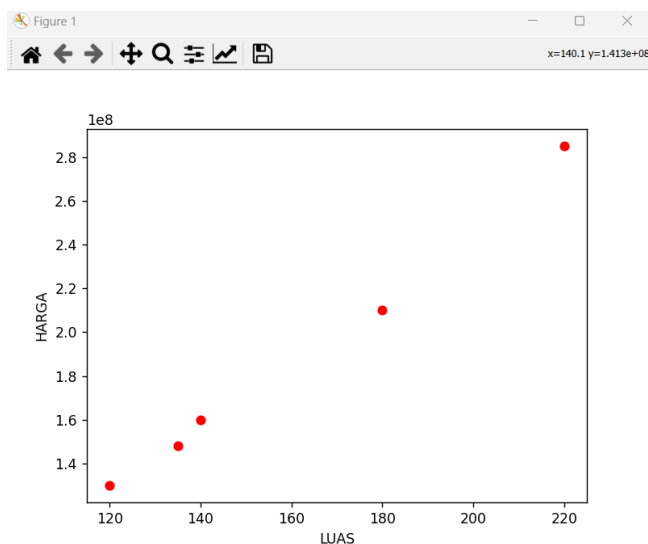
Output :

```
PS D:\praktikum data mining\python> python modul4.py
      luas      harga
0    120  130000000
1    135  148000000
2    140  160000000
3    180  210000000
4    220  285000000
```

- Visualisasi data menggunakan scatter plot

```
plt.xlabel("LUAS");
plt.ylabel("HARGA")
plt.scatter(df.luas, df.harga, color='red')
plt.show()
```

Output



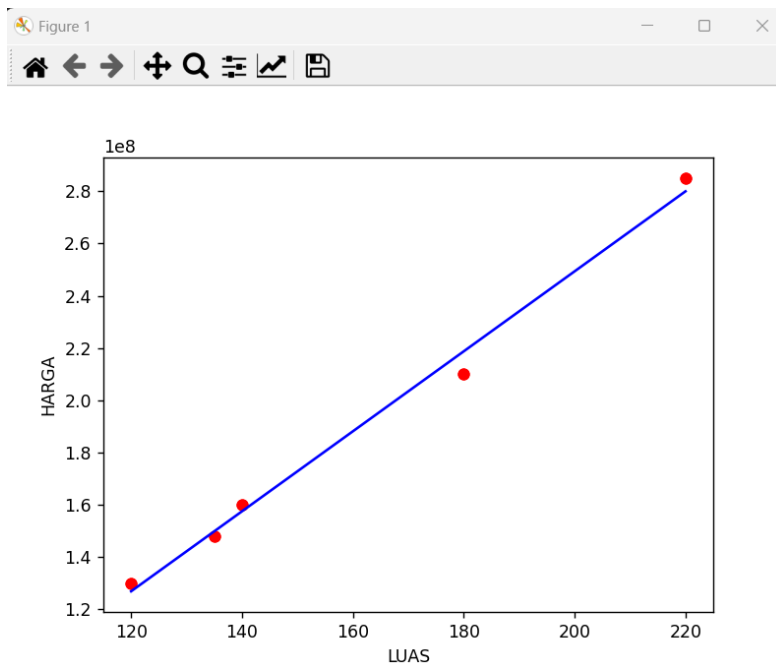
- Membuat model agar mempelajari data dengan sendirinya

```
model_regresi_linier = linear_model.LinearRegression()  
model_regresi_linier.fit(df[['luas']], df.harga)
```

5. Visualisasi data sebenarnya dengan garis prediksi

```
plt.xlabel("LUAS")  
plt.ylabel("HARGA")  
plt.scatter(df.luas, df.harga, color='red')  
plt.plot(df.luas, model_regresi_linier.predict(df[['luas']]), color='blue')  
plt.show()
```

Output



6. Prediksi harga untuk rumah seluas 140 M2

```
hasil = int(model_regresi_linier.predict([[140]]))  
print(hasil)
```

## PRAKTIKUM 8

Judul : library Streamlit



Streamlit adalah sebuah library open-source Python yang digunakan untuk membuat aplikasi web interaktif untuk data science dan machine learning dengan mudah dan cepat. Streamlit memungkinkan pengguna untuk membuat aplikasi web interaktif dengan kode Python sederhana, sehingga dapat dengan mudah memvisualisasikan data, membuat model machine learning, dan mengeksplorasi dataset secara interaktif.

Beberapa fitur dan keuntungan dari Streamlit antara lain:

1. Mudah digunakan: Streamlit dirancang untuk mempermudah pembuatan aplikasi web interaktif. Dengan Streamlit, pengguna dapat membuat aplikasi web dengan kode Python yang sederhana dan mudah dipahami.
2. Terintegrasi dengan mudah: Streamlit dapat diintegrasikan dengan berbagai library Python yang populer untuk data science dan machine learning seperti NumPy, Pandas, Matplotlib, dan Scikit-Learn.
3. Menampilkan data secara interaktif: Streamlit menyediakan berbagai jenis widget interaktif yang dapat digunakan untuk menampilkan data dengan cara yang lebih mudah dipahami seperti plot, tabel, dan grafik.
4. Deploy dengan mudah: Streamlit menyediakan fitur untuk deploy aplikasi web secara mudah dengan layanan hosting seperti Heroku, Google Cloud Platform, dan Amazon Web Services.

Dalam pengembangan aplikasi web interaktif dengan Streamlit, pengguna dapat memanfaatkan fitur-fitur seperti:

- Menggunakan widget seperti slider, checkbox, dan radio button untuk memudahkan interaksi pengguna dengan aplikasi.
- Menggunakan library visualisasi data seperti Matplotlib dan Seaborn untuk memvisualisasikan data dengan berbagai jenis plot.
- Menggunakan library machine learning seperti Scikit-Learn untuk membangun model machine learning dan memprediksi hasil berdasarkan input yang diberikan oleh pengguna.
- Menambahkan layout dan komponen HTML dan CSS untuk membuat tampilan yang lebih menarik dan profesional.

Streamlit dapat digunakan oleh siapa saja yang ingin membuat aplikasi web interaktif dengan mudah dan cepat, baik bagi pengguna yang belum memiliki pengalaman dalam pengembangan web maupun bagi pengguna yang sudah berpengalaman. Dalam penggunaannya, pengguna dapat memanfaatkan dokumentasi resmi Streamlit dan komunitas pengguna yang aktif untuk mendapatkan bantuan dan dukungan.

Dokumentasi selengkapnya bisa dilihat disini : <https://docs.streamlit.io/library/get-started>

Instalasi

### ***Pip Install streamlit***

Cara menjalankan

### ***streamlit run namafile.py***

contoh : membuat aplikasi menghitung segitiga

1. Panggil library yang dibutuhkan

```
import streamlit as st
st.title('Hitung Luas Persegi Panjang')
panjang = st.number_input ("Masukan Nilai Panjang", 0)
lebar = st.number_input ("Masukan Nilai Lebar", 0)
hitung = st.button ("Hitung Luas")

if hitung :
    luas = panjang * lebar
    st.write ("Luas Persegi Panjang Adalah = ",luas)
    st.success("Luas Persegi Panjang Adalah = {luas}")
```

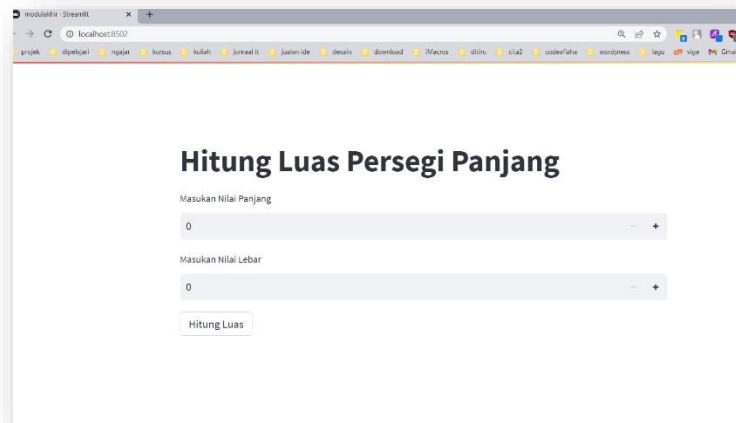
2. Jalankan dengan perintah berikut

```
PS D:\praktikum data mining\python> streamlit run .\modulakhir.py

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8502
Network URL: http://10.10.82.213:8502
```

3. Output bisa dibuka di url diatas



## Deployment prediksi harga rumah

Tulis script berikut

```
import streamlit as st
from sklearn import linear_model
import pandas as pd
import numpy as np

df = pd.read_csv("datarumah.csv")

model_regresi_linier = linear_model.LinearRegression()
model_regresi_linier.fit(df[['luas']], df.harga)

st.title('Prediksi Harga rumah')
luas = st.number_input("Masukan Luas Rumah", 0)

if st.button("cek harga") :
    hasil = int(model_regresi_linier.predict([[luas]]))
    st.success(f"Prediksi Harga Rumah = RP.{hasil}")
```

Output



