Nama Dosen : Teguh Iman Hermanto, M.Kom

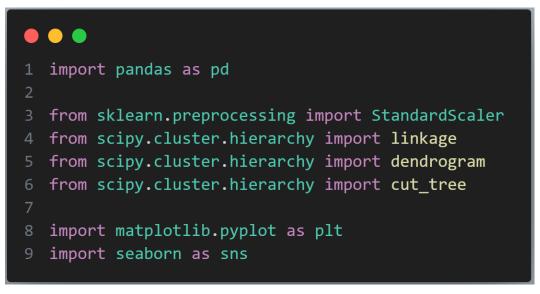
Mata Kuliah : Machine Learning 1
Pembahasan : Hierarchical Clustering

Pokok Pemb : - Membangun Model Hierarchical Clustering

- Simulasi Hierarchical Clustering

- Profiling Hasil CLustering

## 1. Load Library dan dataset pada file Notebook





data.head()  ✓ 0.0s Python													
	ID Nasabah	Usia	Jenis Kelamin	Status Merokok	вмі	Jumlah Klaim	Total Biaya Klaim						
0	1	58	Pria	Ya	30.7	1	4744854						
1	2	26	Pria	Ya	28.2	0	0						
2	3	19	Pria	Ya	22.3	4	5780796						
3	4	53	Pria	Tidak	22.1	4	22669060						
4	5	69	Pria	Ya	30.0	2	11323814						

- 2. Buatkan 5 variasi Exploratory data analysis dan berikan penjelasannya yang terdiri dari
  - a. Pie Plot
  - b. Box plot
  - c. Violin plot
  - d. Scatter plot
  - e. Count plot
- 3. Buat model Hierarchical Clustering

```
# Pra-pemrosesan data
# Pra-pemrosesan data
# Mengonversi kolom kategorikal menjadi numerik
data['Jenis Kelamin'] = data['Jenis Kelamin'].map({'Pria': 0, 'Wanita': 1})
data['Status Merokok'] = data['Status Merokok'].map({'Tidak': 0, 'Ya': 1})
```

```
data.head()
                                                                         Python
          ID
                          Jenis
                                      Status
                                                      Jumlah
                                                                 Total Biaya
                                              ВМІ
              Usia
                       Kelamin
                                    Merokok
                                                        Klaim
                                                                       Klaim
    Nasabah
                                           1 30.7
                58
                             0
                                                                    4744854
0
           2
                26
                             0
                                           1 28.2
                                                            0
2
           3
                19
                             0
                                           1 22.3
                                                            4
                                                                    5780796
3
                             0
                                           0 22.1
           4
                                                                   22669060
                                                            4
4
                69
                             0
                                           1 30.0
                                                            2
                                                                   11323814
```

```
1 # Standardize the data
2 scaler = StandardScaler()
3 X_scaled = scaler.fit_transform(X)
```

```
# Create linkage matrix for dendrogram
linked = linkage(X_scaled, method='ward')
```

```
# Plot dendrogram
plt.figure(figsize=(10, 7))
plt.title("Dendrogram for Hierarchical Clustering")
dendrogram(linked, truncate_mode='lastp', p=30, leaf_rotation=90, leaf_font_size=10)
plt.xlabel("Cluster Size")
plt.ylabel("Distance")
plt.show()
```

```
# membuat cluster label
cluster_labels = cut_tree(linked, n_clusters=5).reshape(-1, )
cluster_labels
```

```
# gabungkan cluster label dengan data
data['Cluster_Labels'] = cluster_labels
```

_	data.head()  ✓ 0.0s  Python													
	ID Nasabah	Usia	Jenis Kelamin	Status Merokok	ВМІ	Jumlah Klaim	Total Biaya Klaim	Cluster_Labels						
0	1	58	0	1	30.7	1	4744854	0						
1	2	26	0	1	28.2	0	0	0						
2	3	19	0	1	22.3	4	5780796	0						
3	4	53	0	0	22.1	4	22669060	1						
4	5	69	0	1	30.0	2	11323814	0						

- 4. Buatkan profiling masing-masing cluster
  - a. Buatkan profiling menggunakan scaterplot, lakukan perbandingan dengan merubah nilai x dan y lalu berikan kesimpulannya

b. Buatkan profiling menggunakan boxplot, lakukan perbandingan dengan merubah nilai y lalu berikan kesimpulannya