

PENERAPAN ALGORITMA K-MEANS UNTUK SEGMENTASI DAERAH DI JAWA TIMUR BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT**M. Kandias Happy Maulana, Nur Cahyo Wibowo, Abdul Rezha Efrat Najaf**

Program Studi Sistem Informasi, Fakultas Ilmu Komputer

Universitas Pembangunan Nasional “Veteran” Jawa Timur

muhammadkandias.hm@gmail.com, nurcahyo.si@upnjatim.ac.id,Rezha.efrat.sifo@upnjatim.ac.id**Abstrak (Indonesia)**

Kesejahteraan masyarakat merupakan indikator penting dalam menilai kualitas hidup suatu daerah. Penelitian ini bertujuan untuk mengelompokkan daerah-daerah di Provinsi Jawa Timur berdasarkan indikator kesejahteraan menggunakan algoritma K-Means. Teknik Principal Component Analysis (PCA) digunakan untuk mereduksi dimensi variabel kesejahteraan sebelum dilakukan klusterisasi. Penelitian ini mengikuti tahapan dalam proses data mining dengan model CRISP-DM yang terdiri dari enam tahapan di antaranya business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Dalam tahap modeling, algoritma K-Means diterapkan untuk menemukan karakteristik atau informasi dalam mengelompokkan daerah dengan tingkat kesejahteraan di Provinsi Jawa Timur pada tahun 2020 hingga 2022. Visualisasi hasil cluster dilakukan dalam bentuk peta interaktif menggunakan library Geopandas dan Folium pada Python. Hasil penelitian menunjukkan bahwa pada tahun 2020 dan 2021 terbentuk 2 cluster dimana cluster 2 adalah cluster sejahtera dan cluster 1 tidak sejahtera, sementara pada tahun 2022 terbentuk 6 cluster dengan karakteristik masing-masing dimana

Sejarah Artikel*Submitted: 14 July 2024**Accepted: 23 July 2024**Published: 24 July 2024***Kata Kunci**Klusterisasi, Kesejahteraan, *Principal Component Analysis*, K-Means, Segmentasi**PENDAHULUAN**

Undang-undang No. 11 Tahun 2009, tentang Kesejahteraan Masyarakat, kesejahteraan masyarakat diartikan kondisi telah terpenuhinya kebutuhan material, spiritual, dan sosial warga negara agar dapat hidup layak dan mampu mengembangkan diri, sehingga dapat melaksanakan fungsi sosialnya. Berdasarkan undang-undang tersebut dapat diketahui bahwa tingkat kesejahteraan dapat diukur dan dinilai berdasarkan atas kemampuan dari seorang individu atau kelompok di dalam usahanya untuk memenuhi kebutuhan baik material maupun spiritualnya (Kadeni dan Srijani, 2020).

Kesejahteraan dapat diukur dari kesehatan, keadaan ekonomi, kebahagiaan dan kualitas hidup rakyat, Pandangan masyarakat umum, dalam keluarga yang sejahtera maka mampu menyekolahkan anggota keluarganya hingga setinggi mungkin. Sama halnya jika semakin tinggi tingkat pendidikan seseorang maka akan membawa keluarganya semakin sejahtera karena mendapatkan timbal balik seperti pekerjaan yang mapan dan pendapatan yang mencukupi (Mulia dan Saputra, 2020). Kemiskinan menjadi satu diantara faktor yang memengaruhi kesejahteraan masyarakat di Jawa Timur. Kesejahteraan masyarakat dapat ditinjau melalui pertumbuhan ekonomi yang dipengaruhi oleh beberapa kondisi, yaitu jumlah penduduk miskin, tingkat pengangguran terbuka, Indeks Pembangunan Manusia (IPM), rata-rata pengeluaran per kapita untuk makanan sebulan, persentase rumah tangga yang memiliki akses terhadap sumber air minum layak, dan persentase rumah tangga yang memiliki akses terhadap layanan sanitasi layak (Farizi dkk, 2023).

Hasil visualisasi dari penelitian ini akan di tampilkan dalam bentuk peta dan dibedakan berdasarkan karakteristik *cluster* atau kesejahteraan Kota/Kabupaten yang ada pada Provinsi Jawa Timur agar dapat mudah dipahami oleh masyarakat awam. Dari hasil penelitian ini tidak

hanya masyarakat namun juga terutama pemerintah diharapkan dapat menjadikan hasil segmentasi sebagai informasi dimana saja daerah yang masih tergolong dalam tingkat kesejahteraan rendah, sehingga pemerintah dapat melakukan tindakan dalam meningkatkan kesejahteraan serta dapat memprioritaskan daerah yang tergolong rendah dalam tingkat kesejahteraan.

Penelitian ini data jumlah penduduk miskin, Tingkat pengangguran terbuka (TPT), Indeks Pembangunan Manusia, rata-rata pengeluaran perkapita untuk makanan sebulan, persentase rumah tangga yang memiliki akses terhadap sumber air minum layak, persentase rumah tangga yang memiliki akses terhadap sanitasi layak didapatkan dari situs resmi Badan Pusat Statistik Provinsi Jawa Timur akan diklasterisasi berdasarkan kriteria kesejahteraan pada masing-masing daerah yang ada di Provinsi Jawa Timur menggunakan algoritma *K-Means*. Data dari penelitian ini kemudian akan diklasterisasi berdasarkan indikator kesejahteraan pada setiap kota-kabupaten Provinsi Jawa Timur menggunakan metode *K-Means*. Algoritma *K-Means* dipilih, diantaranya karena algoritma *K-Means* ini tidak terpengaruh oleh urutan objek (Samosir, Amin, dan Harahap, 2021). Karena metode *K-Means* mempunyai hasil pembagian yang lebih akurat (Yaumi, Zulfiqar, dan Nugroho, 2022). Adapun alasan lainnya mengapa penulis menggunakan metode *K-Means* ini dikarenakan metode *K-Means* merupakan metode yang cukup populer, mempunyai keakuratan data yang cukup baik (Pahlevi dkk, 2022).

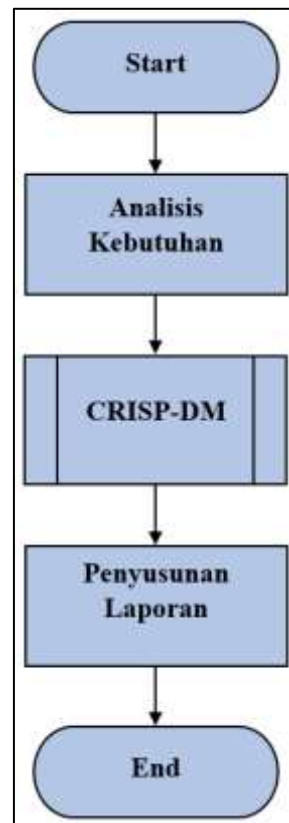
Penggunaan teknik *Principal Component Analysis* (PCA) dapat mereduksi dimensi variabel kesejahteraan sebelum dilakukan clustering dengan *K-Means*. PCA berguna dalam meningkatkan efisiensi model dan menghindari dominasi atribut dengan domain yang lebih besar (Nurohmah, Mayasari, dan Sari, 2023). Penerapan PCA dalam penelitian ini bertujuan untuk mereduksi dimensi variabel kesejahteraan di Jawa Timur.

Data yang telah diolah selanjutnya akan divisualisasikan dengan *library* Geopandas dan Folium. *Library* Geopandas merupakan *library* Python yang digunakan untuk mengolah dan memanipulasi data geospasial, seperti shapefile (shp), GeoJson dan lainnya. Data geospasial sendiri merupakan data geografis yang memiliki titik koordinat dari sebuah objek tertentu. Geopandas juga merupakan plotting package yang berintegrasi dengan Matplotlib yang berfungsi membaca file dalam bentuk format GIS (Wicaksono, dan Susetyo, 2023).

Folium merupakan struktur yang berbentuk daun dimana folium merupakan sebuah library pada python yang biasanya berfungsi untuk menampilkan peta menggunakan OpenStreetMap (OSM) yang mana merupakan peta yang digunakan pada library folium karena sifatnya gratis digunakan atau open source (Sukmayanti, Asmarajaya, dan Sanjaya, 2023). Folium adalah pustaka Python yang membantu dalam membuat beberapa jenis peta leaflet untuk menghasilkan visualisasi peta lokasi. Argumen lokasi memungkinkan untuk memusatkan peta di lokasi tertentu, kemudian juga dapat memberikan tingkat zoom awal ke lokasi tersebut untuk memperbesar peta ke tengah (Subiksa, Ariawan, dan Peling, 2023).

METODOLOGI PENELITIAN

Pada bab metodologi penelitian ini membahas tahapan-tahapan yang digunakan dalam penelitian ini agar terstruktur dengan baik. Berikut adalah tahapan-tahapan metodologi yang dilakukan dalam penelitian ini.



Gambar 3. 1 Alur Metode Penelitian

Analisis Kebutuhan

Pada tahap analisis kebutuhan ini mencakup analisis kebutuhan data serta kebutuhan *software* dan *hardware* agar tujuan dari penelitian ini dapat tercapai.

Kebutuhan Data

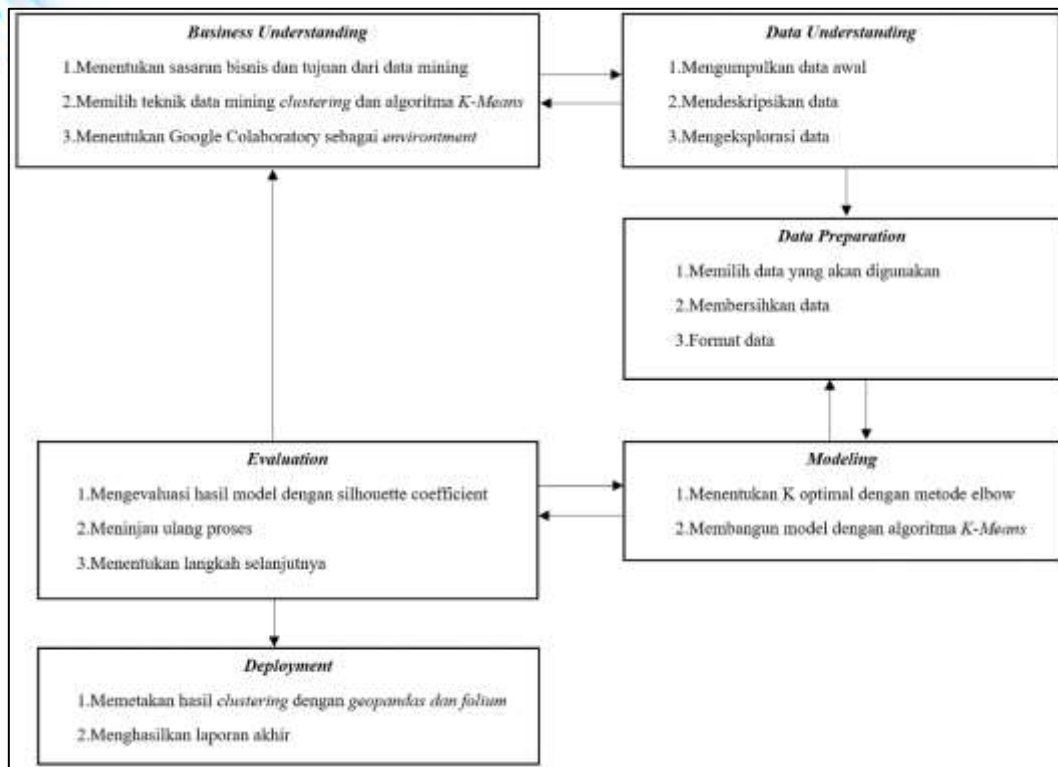
Penelitian ini menggunakan data jumlah penduduk miskin, Tingkat pengangguran terbuka (TPT), Indeks Pembangunan Manusia, rata-rata pengeluaran perkapita untuk makanan sebulan, persentase rumah tangga yang memiliki akses terhadap sumber air minum layak, dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak.

Kebutuhan Software dan Hardware

Dalam melakukan penelitian ini digunakan *software* dan *hardware* untuk merancang dan menjalankan sistem yang telah dibuat. Hardware yang digunakan yaitu laptop Asus VivoBook A412FL dengan spesifikasi RAM 8GB serta penyimpanan internal SSD sebesar 512GB. Software yang digunakan yaitu web browser Chrome, Microsoft Excel untuk melihat data, Microsoft Excel untuk preparasi data dan Google Colab untuk menjalankan bahasa pemrograman Python.

Cross Industry Standard Process for Data Mining (CRISP-DM)

Pada penelitian ini yaitu mengikuti tahapan yang terdapat dalam proses data mining dengan menggunakan model CRISP-DM yang dibangun dari 6 tahapan yang direpresentasikan melalui rancangan penelitian pada Gambar 3.2.



Gambar 3. 2 Model Analitik CRISP-DM

HASIL DAN PEMBAHASAN

Bab ini memuat hasil dan pembahasan dari penelitian yang telah dilakukan sebelumnya yaitu terkait visualisasi hasil clustering data laporan kriminalitas di kota Surabaya menggunakan algoritma *K-Means*.

Implementasi Kebutuhan

Dalam melaksanakan penelitian *clustering* data indikator kesejahteraan masyarakat di Provinsi Jawa Timur menggunakan algoritma *K-Means* diperlukan data yang akan diolah serta spesifikasi perangkat lunak dan perangkat keras.

Kebutuhan Data

Data yang diperlukan dalam penelitian ini adalah data yang di unduh dari Badan Pusat Statistik (BPS) Provinsi Jawa Timur berupa data dari 6 indikator kesejahteraan masyarakat seperti pada pada Tabel 4.1, dan pada Gambar 4.1 adalah salah satu contoh tampilan website BPS, untuk mengunduh data pada situs BPS, terdapat *button* hijau *xlsx* untuk mengunduh data nya.

Tabel 4. 1 Data Indikator Kesejahteraan Masyarakat

Definisi Variabel	Satuan
Jumlah Penduduk Miskin	Ribuan
Tingkat Pengangguran Terbuka (TPT)	Persentase (%)
Indeks Pembangunan Manusia (IPM)	Desimal
Rata-rata Pengeluaran Perkapita	Rupiah
Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak	Persentase (%)
Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak	Persentase (%)



Gambar 4. 1 Website BPS

Dengan menyertakan lokasi berupa koordinat di setiap daerah Kota dan Kabupaten di Provinsi Jawa Timur data yang di ambil adalah data dari tahun 2020, 2021, hingga 2022.

Import Data

Data yang diimport merupakan file CSV yang di dalam nya terdapat daerah atau Kota dan kabupaten dalam Provinsi Jawa Timur. Dalam penelitian ini data yang tersedia antara lain Kabupaten-Kota, Jumlah penduduk miskin, Tingkat pengangguran terbuka (TPT), Indeks Pembangunan Manusia, rata-rata pengeluaran perkapita untuk makanan sebulan, persentase rumah tangga yang memiliki akses terhadap sumber air minum layak, dan persentase rumah tangga yang memiliki akses terhadap sanitasi layak, latitude, longitude. Berikut pada Gambar 4.2 dibawah ini adalah data yang akan diolah.

Kabupaten_Kota	Jumlah Penduduk Miskin (ribu) 2020	Tingkat Pengangguran Terbuka (persentase %) 2020	Indeks Pembangunan Manusia (desimal) 2020	Rata-rata Pengeluaran Perkapita (rupiah) 2020	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2020	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2020
Pabelan	84.00	2.29	83.26	78000	94.43	89.90
Pasarengga	89.74	8.46	71.57	107000	91.75	84.07
Trenggales	87.00	4.11	77.10	98724	88.84	79.01
Tubungagung	76.40	4.01	73.15	102800	88.73	89.40
Blitar	104.96	7.00	71.00	98470	90.35	85.39

Gambar 4. 2 Contoh Data yang Akan Diolah

Klasterisasi

Untuk perlakuan pertama dengan bentuk data yang memiliki satuan berbeda-beda, maka dilakukan standarisasi menggunakan *StandardScaler* untuk melakukan normalisasi fitur, kemudian juga dilakukan atau mengimplikasikan normalisasi pada data (df) serta mengembalikan data yang telah di skalakan dengan *sourcecode* pada Gambar 4.3 dibawah ini untuk standarisasi data:

```
[ ] scaler = StandardScaler() # Membuat instance dari StandardScaler untuk melakukan normalisasi fitur
scaled_features = scaler.fit_transform(df) # Mengaplikasikan normalisasi pada data df dan mengembalikan data yang telah diskalakan
```

Gambar 4. 3 Sourcecode untuk standarisasi data df

Data Preparation

Pada tahap persiapan ini memuat beberapa hal seperti import library pada Gambar 4.4, dan import dataset pada Gambar 4.5. Library yang diimport berupa library untuk analisis data, visualisasi, dan algoritma *K-Means*.

```
[ ] import numpy as np # linear algebra
# Mengimport library numpy untuk operasi aljabar linear dan manipulasi array.
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
# Mengimport library pandas untuk pemrosesan data dan operasi I/O file CSV.
import os
# Mengimport modul os untuk berinteraksi dengan sistem operasi, seperti manajemen file dan direktori.
import geopandas as gpd
# Mengimport library geopandas untuk pemrosesan data geospasial dan operasi geografis.
import matplotlib.pyplot as plt
# Mengimport modul pyplot dari matplotlib untuk membuat visualisasi data dalam bentuk grafik.
import seaborn as sns
# Mengimport library seaborn untuk membuat visualisasi data statistik yang lebih estetik dan informatif.
from sklearn.preprocessing import StandardScaler
# Mengimport StandardScaler dari scikit-learn untuk standarisasi fitur data sebelum analisis lebih lanjut.
from sklearn.decomposition import PCA
# Mengimport PCA (Principal Component Analysis) dari scikit-learn untuk mengurangi dimensi data.
from sklearn.cluster import KMeans
# Mengimport algoritma KMeans dari scikit-learn untuk melakukan klusterisasi pada data.
from sklearn.metrics import silhouette_score
# Mengimport silhouette_score dari scikit-learn untuk mengevaluasi kualitas kluster yang dihasilkan.
from matplotlib.patches import Patch
# Mengimport Patch dari matplotlib untuk membuat elemen legenda khusus pada plot.
sns.set(context='notebook', palette='spectral', style = 'darkgrid', font_scale = 1.5, color_codes=True)
# Mengatur konfigurasi tampilan seaborn: konteks notebook, palet warna Spectral, gaya grid gelap, skala font 1.5, dan penggunaan kode warna.
```

Gambar 4. 4 Load Library

```
[ ] # Mengimport modul untuk mengakses Google Drive
from google.colab import drive

# Memasang Google Drive ke dalam sistem file Colab
drive.mount('/content/drive')

# Menampilkan daftar file dan direktori di dalam folder "test-skripsi" yang berada di Google Drive
!ls "/content/drive/My Drive/test-skripsi"

# Mengubah direktori kerja saat ini ke folder "test-skripsi" di Google Drive
%cd /content/drive/My Drive/test-skripsi
```

Gambar 4. 5 Import Dataset

```
[ ] # Fit PCA ke data tanpa mengurangi dimensi dan hitung rasio variansi yang dijelaskan
pca_full = PCA() # Membuat instance PCA
pca_full.fit(scaled_features) # Melatih model PCA dengan fitur yang sudah distandarisasi

# Hitung rasio variansi yang dijelaskan kumulatif
explained_variance_ratio = pca_full.explained_variance_ratio_ # Mengambil rasio variansi yang dijelaskan oleh setiap komponen
cumulative_explained_variance = explained_variance_ratio.cumsum() # Menghitung kumulatif dari rasio variansi yang dijelaskan

# Plot rasio variansi yang dijelaskan kumulatif untuk menentukan jumlah komponen yang optimal
plt.figure(figsize=(10, 8)) # Membuat figure dengan ukuran tertentu
plt.plot(range(1, len(cumulative_explained_variance) + 1), cumulative_explained_variance, marker='o', linestyle='--') # Membuat plot garis dengan marker
plt.title('Cumulative Explained Variance by PCA Components') # Menambahkan judul pada plot
plt.xlabel('Number of PCA Components') # Menambahkan label pada x
plt.ylabel('Cumulative Explained Variance') # Menambahkan label pada y
plt.grid(True) # Menambahkan grid pada plot
plt.axhline(0.95, color='r', linestyle='-') # Menambahkan garis horizontal pada level 95% variansi sebagai referensi
plt.text(0.5, 0.95, '95% cut-off threshold', color='red', fontsize=14) # Menambahkan teks untuk garis cut-off 95%

# Menentukan jumlah komponen yang menjelaskan setidaknya 95% dari variansi
optimal_num_components = len(cumulative_explained_variance[cumulative_explained_variance >= 0.95]) + 1 # Menghitung jumlah komponen yang memenuhi syarat 95%

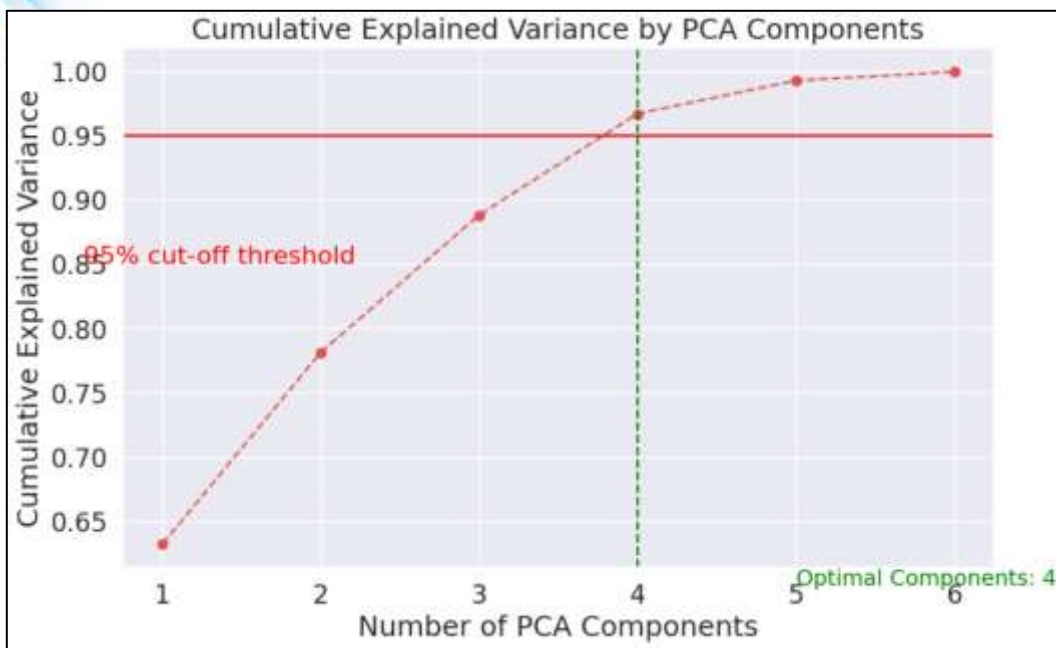
# Menandai jumlah komponen optimal pada plot
plt.axvline(x=optimal_num_components, color='g', linestyle='--') # Menambahkan garis vertikal pada jumlah komponen optimal
plt.text(optimal_num_components + 1, 0.6, f'Optimal Components: {optimal_num_components}', color='green', fontsize=14) # Menambahkan teks untuk jumlah komponen optimal

plt.show() # Menampilkan plot

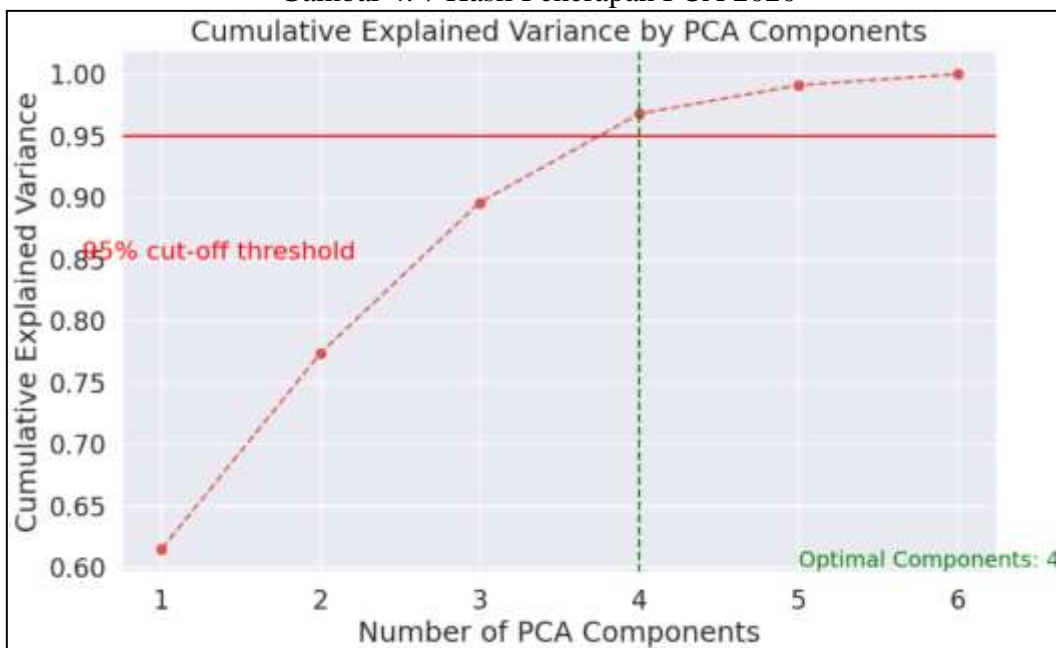
# Mengetahui jumlah komponen yang optimal
optimal_num_components # Output: jumlah komponen optimal
```

Gambar 4. 6 Sourcecode untuk menerapkan PCA

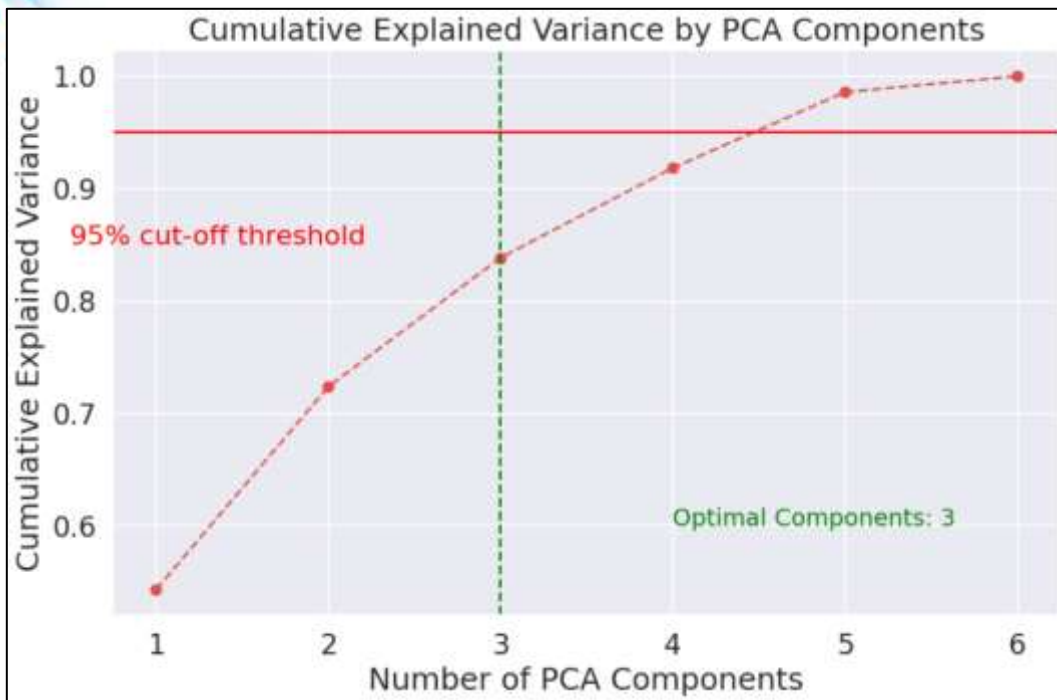
Pada Gambar 4.6 dilakukan penerapan PCA, ‘pca_full = PCA()’ untuk membuat instance PCA, kemudian dilakukan pengskalaan fitur dengan code ‘pca_full.fit(scaled_features)’, dan dilakukan penentuan jumlah komponen yang menunjukkan 95% dari variasi dengan code ‘optimal_num_components = len(cumulative_explained_variance[cumulative_explained_variance >= 0.95]) + 1’. Berikut pada Gambar 4.7, Gambar 4.8, dan Gambar 4.9 adalah hasil dari penerapan PCA yang menunjukkan 4 komponen PCA optimal:



Gambar 4. 7 Hasil Penerapan PCA 2020



Gambar 4. 8 Hasil Penerapan PCA 2021



Gambar 4. 9 Hasil Penerapan PCA 2022

Pada Gambar 4.10 dilakukan penetapan komponen PCA, kemudian pada Gambar 4.11 dilakukan penerapan PCA dengan jumlah komponen optimal yang telah di tentukan sebelumnya, dan mengubah fitur dengan skala yang telah disesuaikan oleh PCA.

```
[ ] n_components = 4 # Menetapkan jumlah komponen utama (principal components) yang akan digunakan, berdasarkan identifikasi dari plot sebelumnya
pca = PCA(n_components=n_components) # Membuat objek PCA dengan jumlah komponen yang ditentukan
```

Gambar 4. 10 Penetapan Komponen PCA

```
[ ] # Menerapkan PCA dengan jumlah komponen optimal
pca = PCA(n_components=n_components)
# Mengubah fitur-fitur dengan skala yang telah disesuaikan menggunakan PCA
pca_result = pca.fit_transform(scaled_features)
```

Gambar 4. 11 Penerapan PCA dengan Jumlah Komponen Optimal

Pada Gambar 4.12 merupakan *source code* untuk ekstrak nilai absolut dari beban PCA serta pada Gambar 4.11, Gambar 4.11, dan Gambar 4.13 adalah hasil ekstrak nya.

```
[58] # Ekstrak nilai absolut dari beban PCA
pca_loadings_analysis = pd.DataFrame(
    np.abs(pca.components_), # Menganbil nilai absolut dari komponen PCA
    columns=df.columns, # Menamai kolom dengan nama kolom pada dataframe asli
    index=[f'PC{i+1}' for i in range(pca.n_components)] # Memberi nama indeks dengan nomor komponen PCA
)

# Identifikasi fitur yang berkontribusi teratas untuk setiap komponen utama
top_features_per_pc = pca_loadings_analysis.apply(lambda s: s.nlargest(5).index.tolist(), axis=1)

# Tampilkan fitur yang berkontribusi teratas untuk setiap komponen utama
top_features_per_pc
```

Gambar 4. 12 Source Code Untuk Ekstrak Beban PCA

```
PC1 [Indeks Pembangunan Manusia (desimal) 2020, Ra...
PC2 [Jumlah Penduduk Miskin (ribu) 2020, Tingkat P...
PC3 [Persentase Rumah Tangga yang Memiliki Akses S...
PC4 [Persentase Rumah Tangga yang Memiliki Akses T...
dtype: object
```

Gambar 4. 13 Hasil Ekstrak Beban PCA 2020

```
PC1 [Indeks Pembangunan Manusia (desimal) 2021, Ra...
PC2 [Jumlah Penduduk Miskin (ribu) 2021, Persentas...
PC3 [Persentase Rumah Tangga yang Memiliki Akses S...
PC4 [Persentase Rumah Tangga yang Memiliki Akses T...
dtype: object
```

Gambar 4. 14 Hasil Ekstrak Beban PCA 2021

```
PC1 [Indeks Pembangunan Manusia (desimal) 2022, Ra...
PC2 [Persentase Rumah Tangga yang Memiliki Akses S...
PC3 [Jumlah Penduduk Miskin (ribu) 2022, Persentas...
dtype: object
```

Gambar 4. 15 Hasil Ekstrak Beban PCA 2022

Pada Gambar 4.16 merupakan *source code* untuk menampilkan hasil ekstrak beban PCA ke dalam bentuk data frame, serta pada Gambar 4.17, Gambar 4.18, dan Gambar 4.19 adalah hasil dalam bentuk data frame.

```
# Membuat dataframe dari data fitur teratas untuk setiap komponen prinsip
top_features_per_pc_df = pd.DataFrame(top_features_per_pc, columns=["result"])
# Menghasilkan dataframe yang berisi data fitur teratas untuk setiap komponen prinsip
top_features_per_pc_df
```

Gambar 4. 16 Source Code untuk Membuat data frame Ekstrak Beban PCA

index	result
PC1	Indeks Pembangunan Manusia (desimal) 2020,Rata-rata Pengeluaran Perkapita (rupiah) 2020,Tingkat Pengangguran Terbuka (persentase %) 2020,Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2020,Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2020
PC2	Jumlah Penduduk Miskin (ribu) 2020,Tingkat Pengangguran Terbuka (persentase %) 2020,Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2020,Rata-rata Pengeluaran Perkapita (rupiah) 2020,Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2020
PC3	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2020,Rata-rata Pengeluaran Perkapita (rupiah) 2020,Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2020,Tingkat Pengangguran Terbuka (persentase %) 2020,Jumlah Penduduk Miskin (ribu) 2020
PC4	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2020,Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2020,Jumlah Penduduk Miskin (ribu) 2020,Tingkat Pengangguran Terbuka (persentase %) 2020,Rata-rata Pengeluaran Perkapita (rupiah) 2020

Gambar 4. 17 Data Frame Ekstrak Beban PCA 2020

index	result
PC1	Indeks Pembangunan Manusia (desimal) 2021,Rata-rata Pengeluaran Perkapita (rupiah) 2021,Tingkat Pengangguran Terbuka (persentase %) 2021,Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2021,Jumlah Penduduk Miskin (ribu) 2021
PC2	Jumlah Penduduk Miskin (ribu) 2021,Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2021,Tingkat Pengangguran Terbuka (persentase %) 2021,Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2021,Rata-rata Pengeluaran Perkapita (rupiah) 2021
PC3	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2021,Jumlah Penduduk Miskin (ribu) 2021,Tingkat Pengangguran Terbuka (persentase %) 2021,Rata-rata Pengeluaran Perkapita (rupiah) 2021,Indeks Pembangunan Manusia (desimal) 2021
PC4	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2021,Jumlah Penduduk Miskin (ribu) 2021,Tingkat Pengangguran Terbuka (persentase %) 2021,Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2021,Rata-rata Pengeluaran Perkapita (rupiah) 2021

Gambar 4. 18 Data Frame Ekstrak Beban PCA 2021

index	result
PC1	Indeks Pembangunan Manusia (desimal) 2022,Rata-rata Pengeluaran Perkapita (rupiah) 2022,Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2022,Jumlah Penduduk Miskin (ribu) 2022,Tingkat Pengangguran Terbuka (persentase %) 2022
PC2	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2022,Tingkat Pengangguran Terbuka (persentase %) 2022,Jumlah Penduduk Miskin (ribu) 2022,Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2022,Indeks Pembangunan Manusia (desimal) 2022
PC3	Jumlah Penduduk Miskin (ribu) 2022,Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2022,Tingkat Pengangguran Terbuka (persentase %) 2022,Rata-rata Pengeluaran Perkapita (rupiah) 2022,Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2022

Gambar 4. 19 Data Frame Ekstrak Beban PCA 2022

Pada Gambar 4.20 merupakan *source code* untuk membuat *heatmap* dari PCA loadings.

```
[132] # Mendapatkan komponen PCA (loadings)
      pca_components = pca.components_

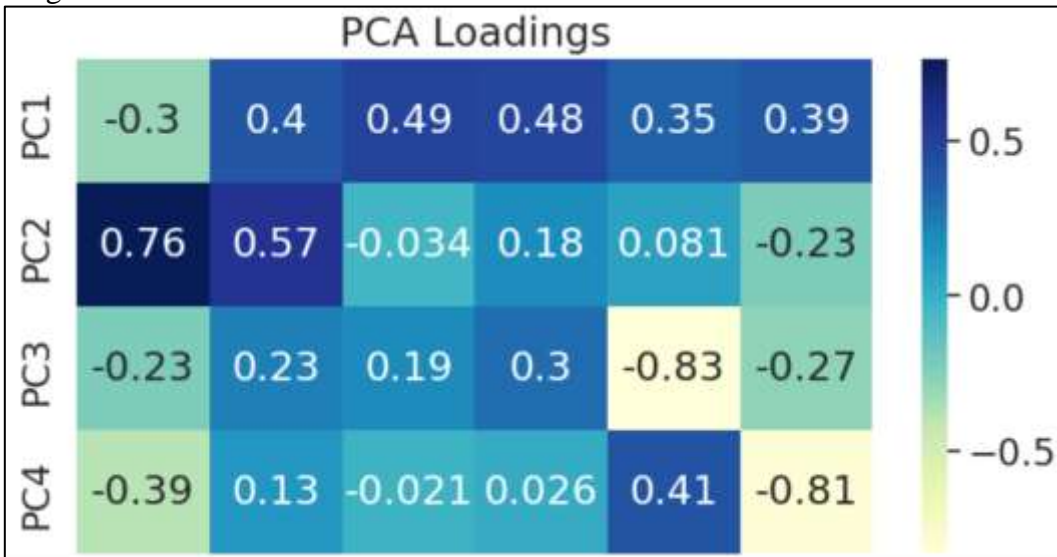
# Membuat DataFrame untuk visualisasi dan analisis yang lebih baik
pca_loadings_df = pd.DataFrame(pca_components, columns=df.columns, index=[f'PC{i+1}' for i in range(n_components)])

# Menampilkan loadings
pca_loadings_df

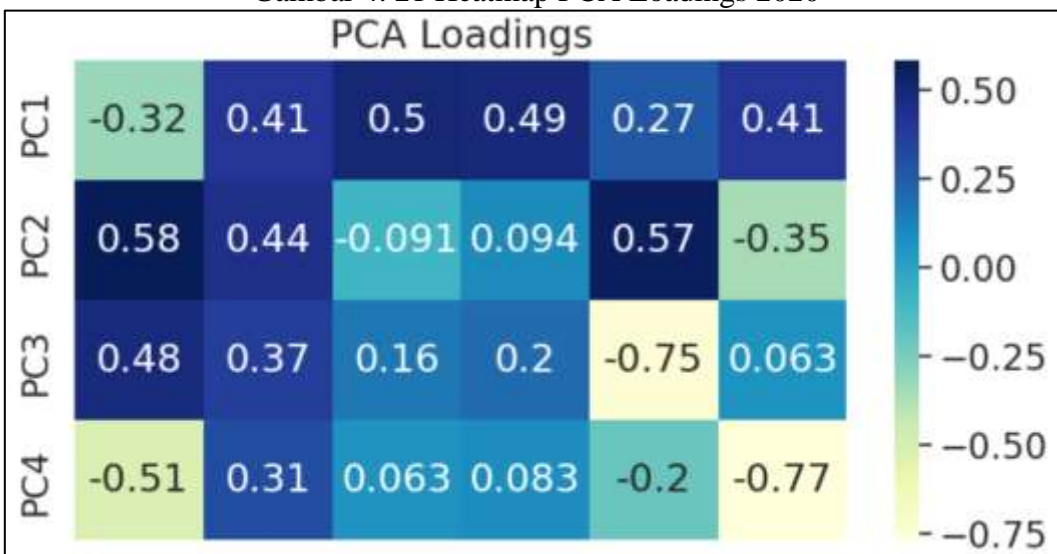
# Heatmap dari loadings
plt.figure(figsize=(8, 4))
sns.heatmap(pca_loadings_df, cmap="YlGnBu", annot=True)
plt.title('PCA Loadings')
plt.show()
```

Gambar 4. 20 Source Code untuk heatmap PCA Loadings

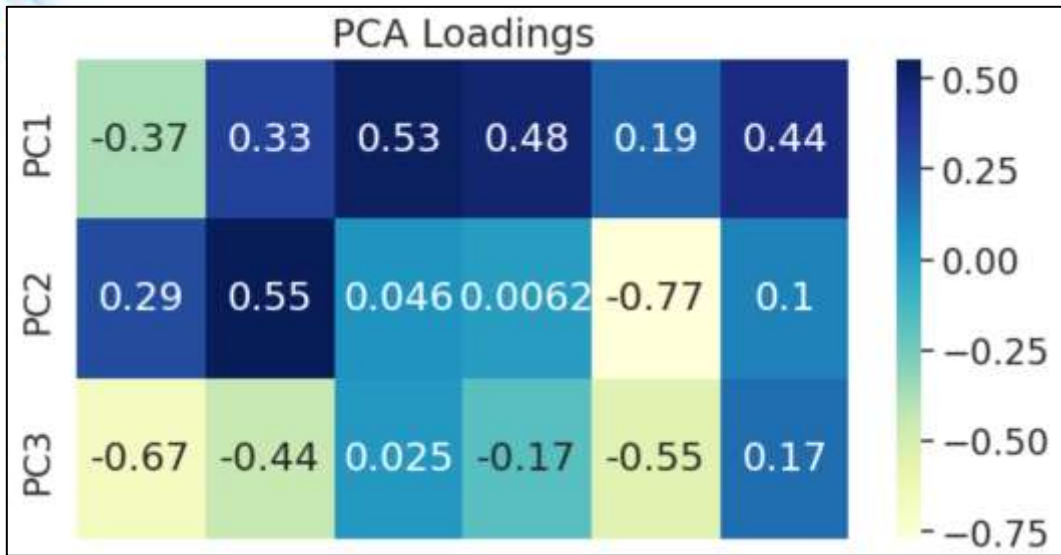
Pada Gambar 4.21, Gambar 4.22, dan Gambar 4.23 di bawah, merupakan hasil atau tampilan *heatmap* PCA loadings. Dapat dilihat dari PCA Loadings 2020, bahwa PC1 dan PC 2 memiliki *value* lebih tinggi dibandingkan PC3 dan PC4, begitu juga pada PCA Loadings 2021 dimana PC1 dan PC 2 memiliki *value* lebih tinggi dibandingkan PC3 dan PC4, kemudian pada PCA Loadings 2022 juga di dapati PC1 dan PC 2 memiliki *value* lebih tinggi dibandingkan PC3.



Gambar 4. 21 Heatmap PCA Loadings 2020



Gambar 4. 22 Heatmap PCA Loadings 2021

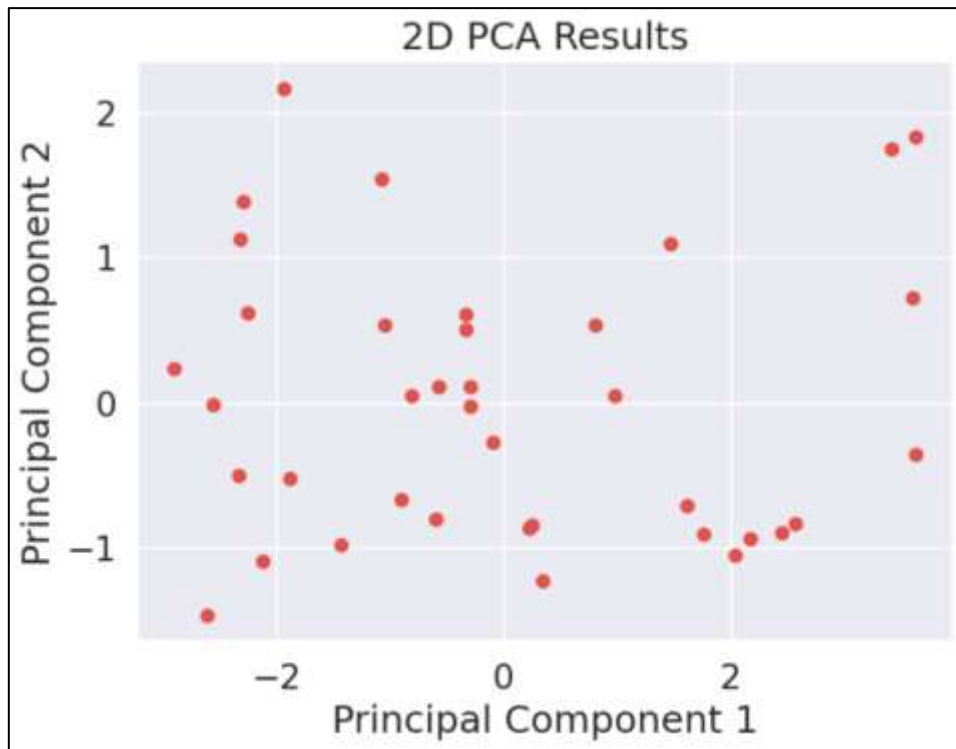


Gambar 4. 23 Heatmap PCA Loadings 2022

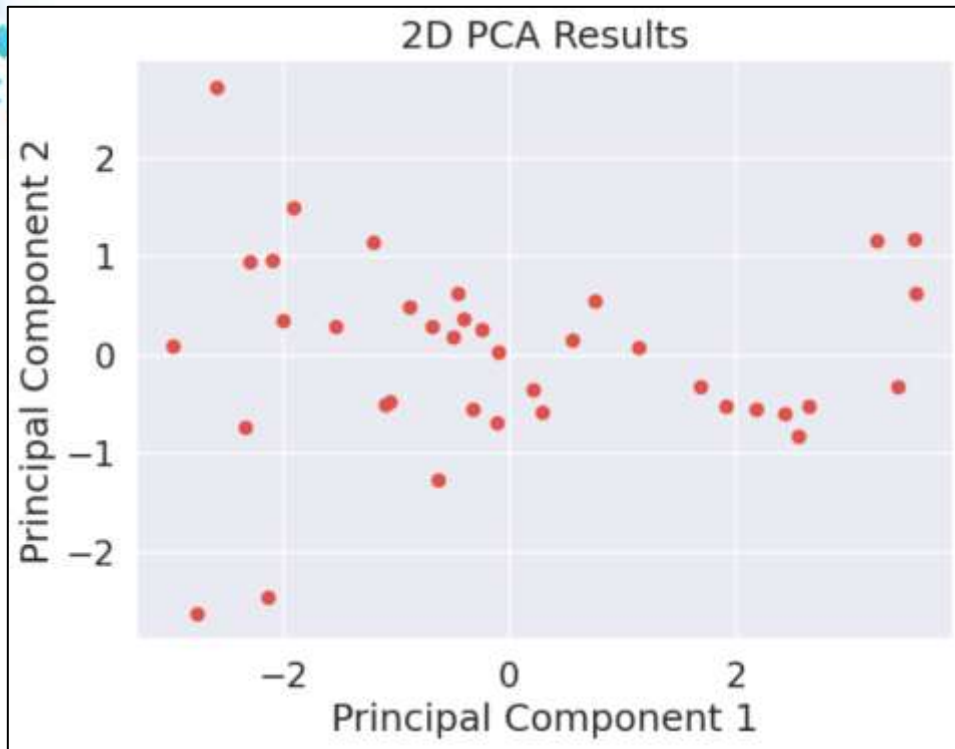
Pada Gambar 4.24 adalah *source code* untuk memvisualisasikan hasil reduksi dimensi oleh PCA ke dalam 2 dimensi. Berikut hasil visualisasi nya pada Gambar 4.25, Gambar 4.26, dan Gambar 4.27.

```
[ ] # Visualisasi hasil reduksi dimensi PCA dalam 2 dimensi
plt.figure(figsize=(7, 5)) # Mengatur ukuran gambar
plt.scatter(pca_result[:, 0], pca_result[:, 1]) # Menyebarkan titik pada plot dengan nilai dari dua komponen PCA
plt.xlabel('Principal Component 1') # Memberi label sumbu x
plt.ylabel('Principal Component 2') # Memberi label sumbu y
plt.title('2D PCA Results') # Memberi judul plot
plt.grid(True) # Menampilkan grid di plot
plt.show() # Menampilkan plot
```

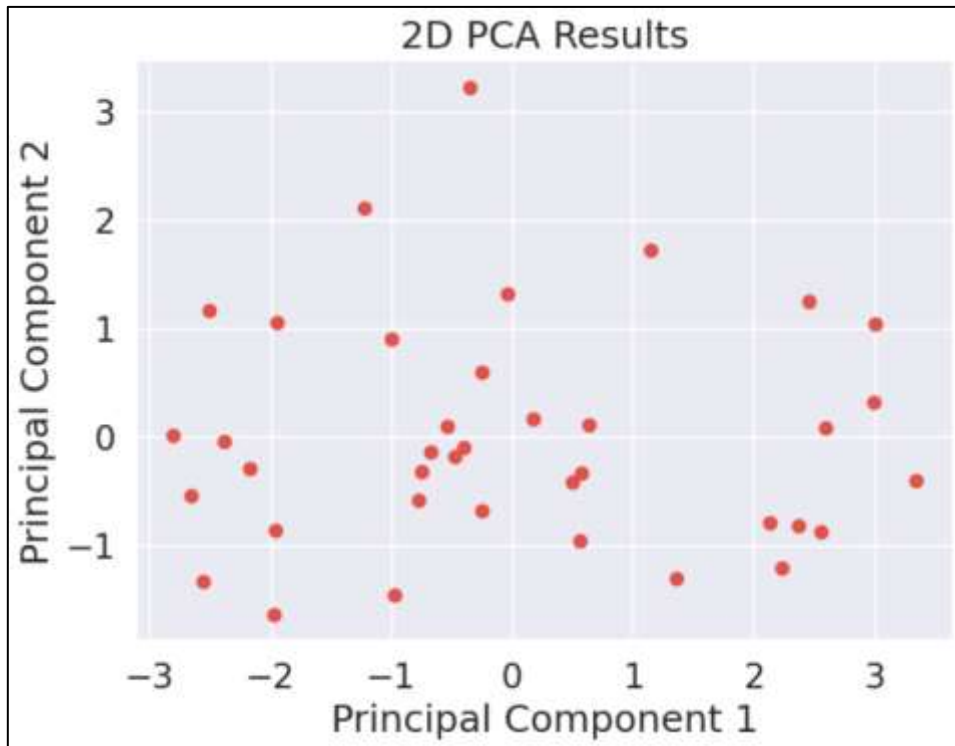
Gambar 4. 24 Source Code untuk Plot 2 Dimensi



Gambar 4. 25 Hasil Visualisasi 2 Dimensi 2020



Gambar 4. 26 Hasil Visualisasi 2 Dimensi 2021



Gambar 4. 27 Hasil Visualisasi 2 Dimensi 2022

1.1.1 Modeling

Untuk menentukan jumlah K paling optimal perlu dideklarasikan variabel apa saja yang akan diambil. *Cluster* yang dilakukan uji coba berjumlah mulai dari 1 sampai dengan 15. Namun, perlu ada nilai pasti yang memastikan bahwa jumlah *cluster* tersebut merupakan jumlah *cluster* yang paling optimal. Pada Gambar di bawah merupakan source code metode

elbow pada Gambar 4.28 dan hasil dari tahapan menentukan jumlah K paling optimal, dimana dari tahapan ini dapat dilihat pada visualisasi metode *Elbow* bahwa nilai K yang paling optimal untuk data tahun 2020 adalah antara 2 sampai dengan 4 pada Gambar 4.29, pada visualisasi metode *Elbow* bahwa nilai K yang paling optimal untuk data tahun 2021 adalah antara 2 sampai dengan 4 pada Gambar 4.30, pada visualisasi metode *Elbow* bahwa nilai K yang paling optimal untuk data tahun 2022 adalah antara 2 sampai dengan 4 pada Gambar 4.31.

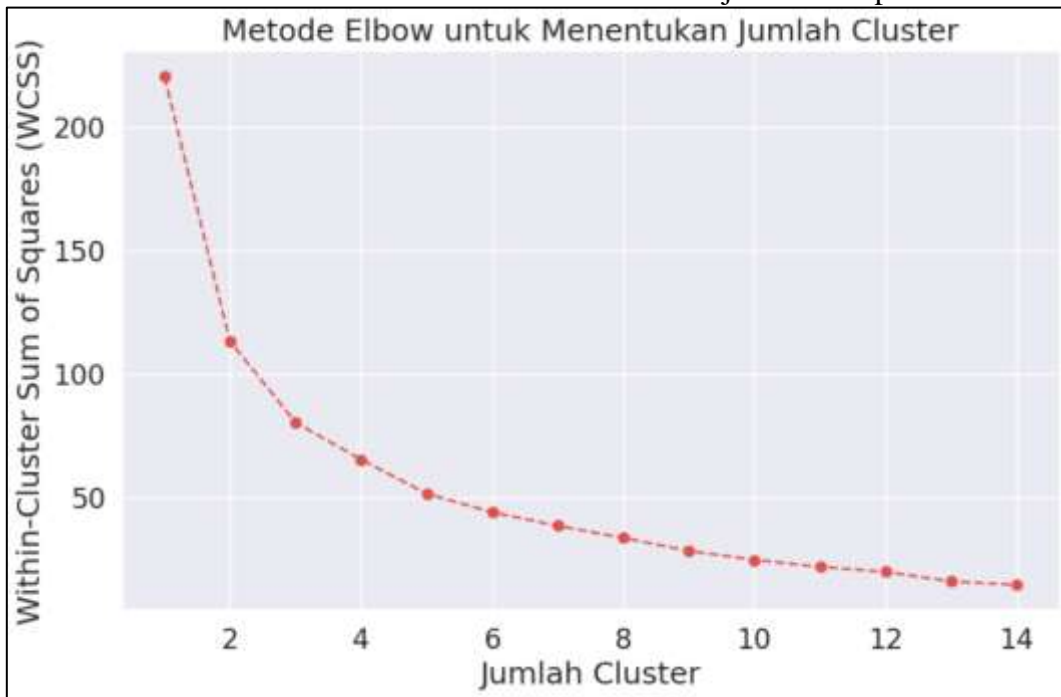
```
[ ] # Menentukan jumlah cluster optimal menggunakan Metode Elbow

wcss = [] # Menginisialisasi daftar kosong untuk menyimpan within-cluster sum of squares (WCSS)

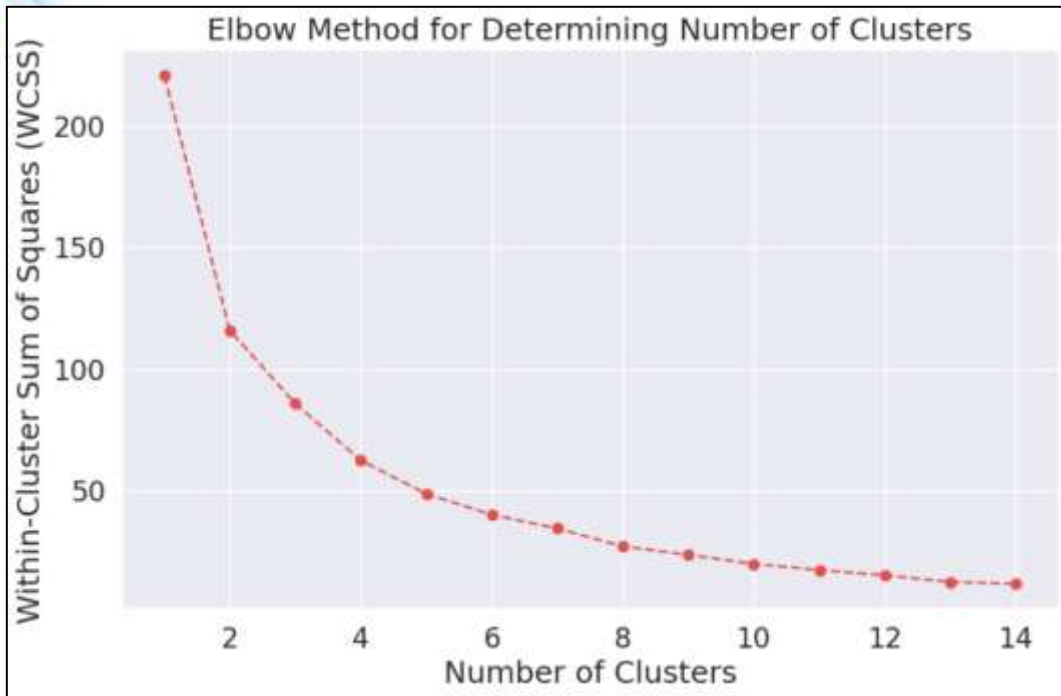
# Melakukan perulangan untuk berbagai jumlah cluster
for i in range(1, 15):
    kmeans = KMeans(n_clusters=i, init='k-means++', n_init=10, random_state=42) # Membuat objek KMeans dengan berbagai jumlah cluster
    kmeans.fit(pca_result) # Train model KMeans dengan data
    wcss.append(kmeans.inertia_) # Menambahkan WCSS dari setiap model ke dalam daftar

# Plot grafik Metode Elbow
plt.figure(figsize=(10, 6))
plt.plot(range(1, 15), wcss, marker='o', linestyle='--') # Plot jumlah cluster - WCSS
plt.title('Metode Elbow untuk Menentukan Jumlah Cluster') # Memberi judul
plt.xlabel('Jumlah Cluster') # Memberi label sumbu x
plt.ylabel('Within-Cluster Sum of Squares (WCSS)') # Memberi label sumbu y
plt.grid(True) # Menambahkan grid untuk visualisasi yang lebih baik
plt.show() # Menampilkan plot
```

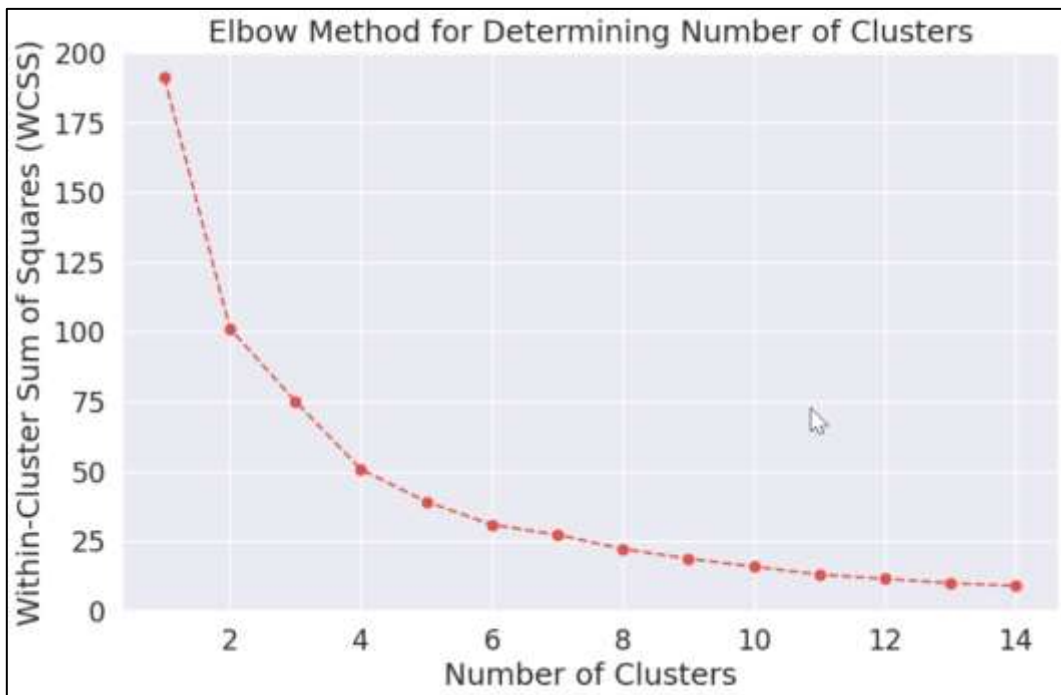
Gambar 4. 28 Source Code untuk Menentukan jumlah K Optimal



Gambar 4. 29 Hasil Visualisasi Metode *Elbow* 2020



Gambar 4. 30 Hasil Visualisasi Metode *Elbow* 2021



Gambar 4. 31 Hasil Visualisasi Metode *Elbow* 2022

1.1.2 Evaluation

Tahapan ini merupakan lanjutan dari tahapan sebelumnya dimana secara kasat mata telah terlihat bahwa ada garisnya paling menekuk sehingga jumlah *cluster* yang paling optimal adalah antara 2 sampai dengan 4. Oleh karena itu pada tahapan ini digunakan *Silhouette Coefficient* untuk mengetahui *cluster* mana yang paling tepat untuk dijadikan jumlah *cluster*. Berikut pada Gambar 4.32 adalah *source code* untuk perhitungan *Silhouette Coefficient* dan pada Gambar 4.33 adalah hasil perhitungan untuk tahun 2020 dan di dapati skor yang paling mendekati 1 adalah 2 *cluster* dengan skor 0,404, Gambar 4.34 adalah hasil perhitungan untuk tahun 2021 dan di dapati skor yang paling mendekati 1 adalah 2 *cluster* dengan skor 0,438, serta Gambar 4.35 adalah hasil perhitungan untuk tahun 2022 dan di dapati skor yang paling mendekati 1 adalah 6 *cluster* dengan skor 0,410.

```
[ ] for i in range(2, 15): # Loop melalui nomor kluster dari 2 sampai 14
kmeans = KMeans(n_clusters=i, init='k-means++', n_init=10, random_state=42) # Inisialisasi KMeans dengan jumlah kluster saat ini
kmeans.fit(pca_result) # Menfitur model KMeans ke data
cluster_labels = kmeans.labels_ # Menetapkan label kluster ke titik data
silhouette_avg = silhouette_score(pca_result, cluster_labels) # Menghitung skor siluet untuk pengelompokan saat ini

print(f'Silhouette Score untuk {i} klaster: {silhouette_avg}') # Mencetak skor siluet untuk pengelompokan saat ini
```

Gambar 4. 32 Source code untuk Perhitungan *Silhouette Coefficient*

```
Silhouette Score untuk 2 klaster: 0.40443095547141844
Silhouette Score untuk 3 klaster: 0.350513224122681
Silhouette Score untuk 4 klaster: 0.3598460672982156
Silhouette Score untuk 5 klaster: 0.3794702666489977
Silhouette Score untuk 6 klaster: 0.37016841046534943
Silhouette Score untuk 7 klaster: 0.2974824428455953
Silhouette Score untuk 8 klaster: 0.3110997397946158
Silhouette Score untuk 9 klaster: 0.3471119941630094
Silhouette Score untuk 10 klaster: 0.34438167155358
Silhouette Score untuk 11 klaster: 0.34648429730209124
Silhouette Score untuk 12 klaster: 0.31294172081656785
Silhouette Score untuk 13 klaster: 0.37914923326234584
Silhouette Score untuk 14 klaster: 0.3526340092474064
```

Gambar 4. 33 Hasil Perhitungan *Silhouette Coefficient* 2020

```

Silhouette Score for 2 clusters: 0.438382536887354
Silhouette Score for 3 clusters: 0.3546035491636868
Silhouette Score for 4 clusters: 0.3903655153870855
Silhouette Score for 5 clusters: 0.4077781212957986
Silhouette Score for 6 clusters: 0.4065873187480721
Silhouette Score for 7 clusters: 0.4205482506882799
Silhouette Score for 8 clusters: 0.40005646640575354
Silhouette Score for 9 clusters: 0.3578725722463061
Silhouette Score for 10 clusters: 0.36010527671322007
Silhouette Score for 11 clusters: 0.3477429906573831
Silhouette Score for 12 clusters: 0.34165830280218207
Silhouette Score for 13 clusters: 0.35795136274561706
Silhouette Score for 14 clusters: 0.3108923452954521

```

Gambar 4. 34 Hasil Perhitungan *Silhouette Coeficient* 2021

```

Silhouette Score for 2 clusters: 0.3831839532369151
Silhouette Score for 3 clusters: 0.3464958304401301
Silhouette Score for 4 clusters: 0.40310988472190173
Silhouette Score for 5 clusters: 0.401700098584294
Silhouette Score for 6 clusters: 0.4100722567435728
Silhouette Score for 7 clusters: 0.37862931913656833
Silhouette Score for 8 clusters: 0.3678330063362678
Silhouette Score for 9 clusters: 0.3671236121734637
Silhouette Score for 10 clusters: 0.3737119231592758
Silhouette Score for 11 clusters: 0.3616509618836378
Silhouette Score for 12 clusters: 0.3558863289797654
Silhouette Score for 13 clusters: 0.35022534897667806
Silhouette Score for 14 clusters: 0.3382434098237913

```

Gambar 4. 35 Hasil Perhitungan *Silhouette Coeficient* 2022

Pada Gambar 4.36 di bawah merupakan *source code* untuk menetapkan jumlah *cluster* yang paling tepat yaitu 2 *cluster* untuk tahun 2020, pada Gambar 4.37 di bawah merupakan *source code* untuk menetapkan jumlah *cluster* yang paling tepat yaitu 2 *cluster* untuk tahun 2021, serta pada Gambar 4.38 di bawah merupakan *source code* untuk menetapkan jumlah *cluster* yang paling tepat yaitu 6 *cluster* untuk tahun 2022 sebagaimana telah di tentukan dengan metode *elbow* dan di validasi ulang dengan *silhouette coefficient*.

```

[ ] # Menetapkan jumlah kluster
# (Jumlah ini dapat ditentukan berdasarkan Metode Elbow dan validasi dengan silhouette coefficient )
n_clusters = 2 # jumlah kluster yang sesuai

# Menerapkan klustering KMeans
kmeans_pca = KMeans(n_clusters=n_clusters, random_state=42)
kmeans_pca.fit(pca_result)

# Label kluster untuk setiap titik data
cluster_labels = kmeans_pca.labels_

```

Gambar 4. 36 *Source Code* untuk penerapan K-Means 2020

```
[ ] # Menetapkan jumlah kluster
# (Jumlah ini dapat ditentukan berdasarkan Metode Elbow dan validasi dengan silhouette coefficient )
n_clusters = 2 # appropriate number of clusters

# Applying KMeans clustering
kmeans_pca = KMeans(n_clusters=n_clusters, random_state=42)
kmeans_pca.fit(pca_result)

# The cluster labels for each data point
cluster_labels = kmeans_pca.labels_
```

Gambar 4. 37 Source Code untuk penerapan K-Means 2021

```
[ ] # Menetapkan jumlah kluster
# (Jumlah ini dapat ditentukan berdasarkan Metode Elbow dan validasi dengan silhouette coefficient )
n_clusters = 6 # appropriate number of clusters

# Applying KMeans clustering
kmeans_pca = KMeans(n_clusters=n_clusters, random_state=42)
kmeans_pca.fit(pca_result)

# The cluster labels for each data point
cluster_labels = kmeans_pca.labels_
```

Gambar 4. 38 Source Code untuk penerapan K-Means 2022

Berikut pada Gambar 4.39 dibawah merupakan *source code* untuk melakukan plotting *cluster*.

```
[38] # Plotting
plt.figure(figsize=(8, 6)) # Mengatur ukuran gambar plot

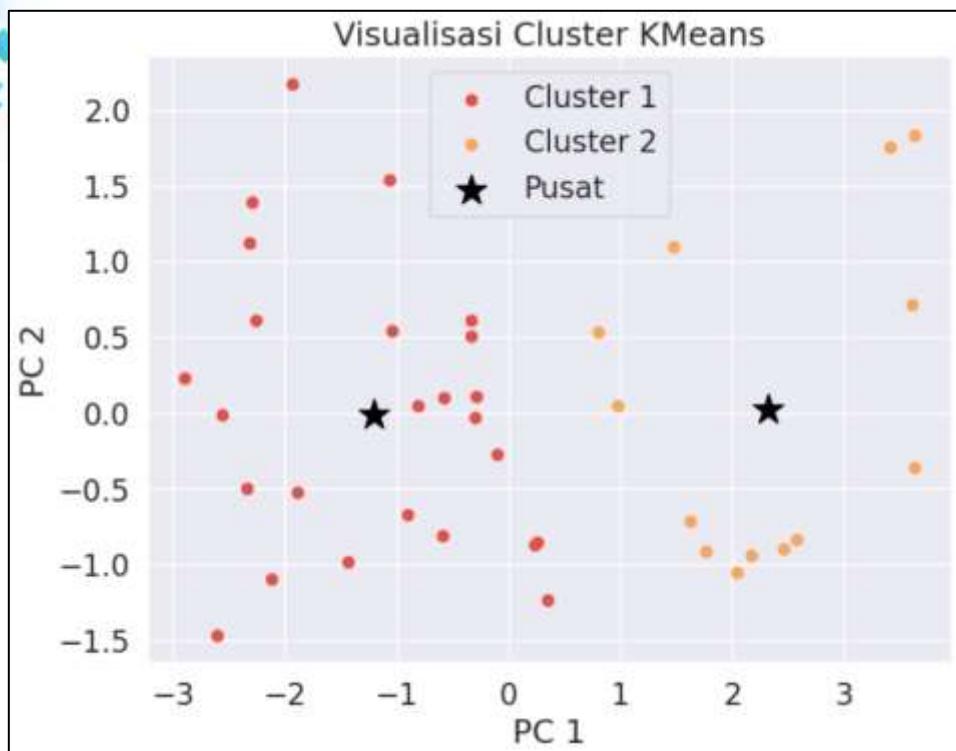
# Plot titik data, diwarnai berdasarkan label kluster mereka
for i in range(n_clusters):
    # Memilih hanya titik data dengan label kluster 'i'
    ds = pca_result[cluster_labels == i]
    # Memplot titik-titik data ini
    plt.scatter(ds[:, 0], ds[:, 1], label=f'Cluster {i+1}')

# Menandai pusat-pusat kluster
centers = kmeans_pca.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], s=300, c='black', label='Pusat', marker='*')

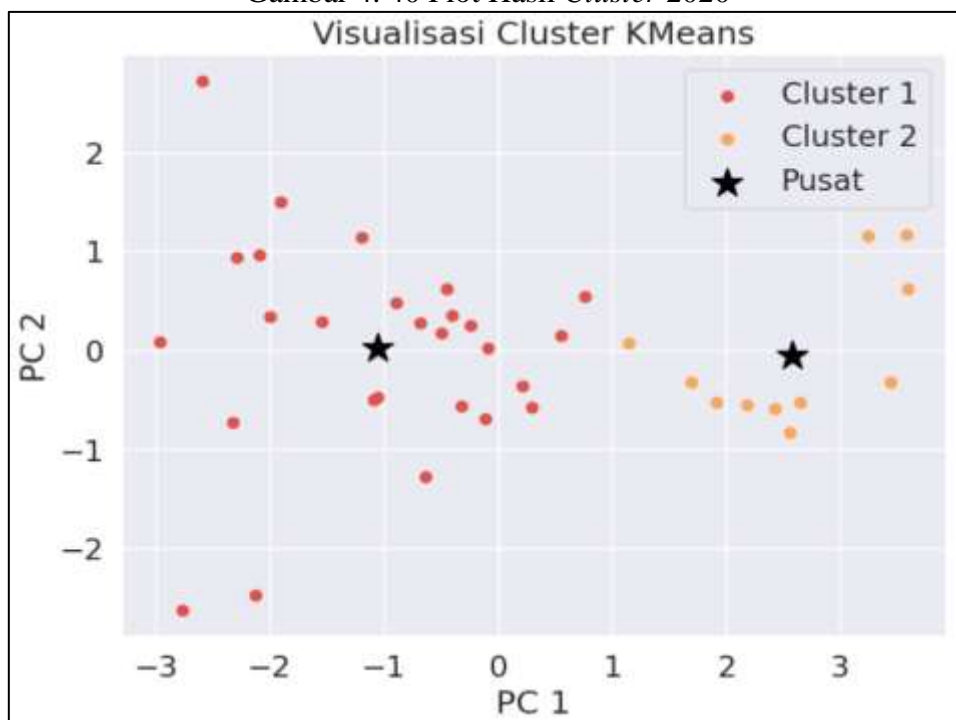
plt.title('Visualisasi Cluster KMeans')
plt.xlabel('PC 1')
plt.ylabel('PC 2')
plt.legend()
plt.show()
```

Gambar 4. 39 Source Code Untuk Plotting Cluster

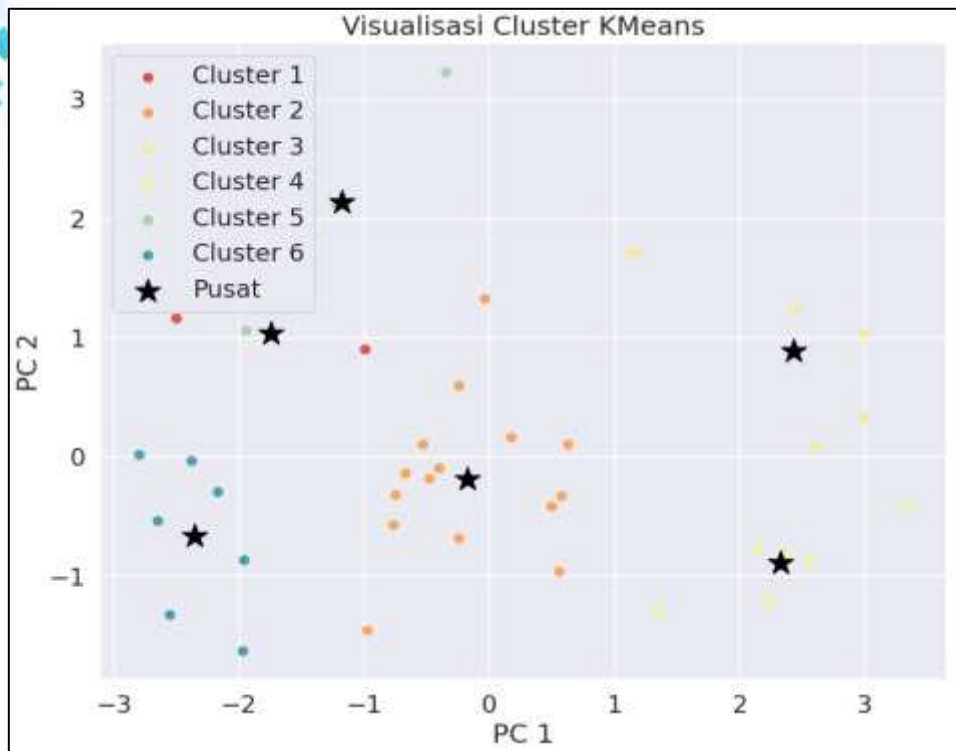
Berikut pada Gambar 4.40 merupakan visualisasi plotting cluster pada tahun 2020, Gambar 4.41 merupakan visualisasi plotting cluster pada tahun 2021, dan Gambar 4.42 merupakan visualisasi plotting cluster pada tahun 2022, dalam visualisasi ini terdapat beberapa titik yang tersebar secara acak di seluruh plot, menunjukkan adanya variasi dalam karakteristik di setiap clusternya. Terdapat overlap antara cluster 1 dan cluster 5 di sebagian wilayah tengah plot, hal ini menunjukkan adanya kemungkinan adanya kesamaan atau keragaman dalam fitur-fitur yang dimiliki oleh kedua cluster tersebut.



Gambar 4. 40 Plot Hasil *Cluster* 2020



Gambar 4. 41 Plot Hasil *Cluster* 2021



Gambar 4. 42 Plot Hasil Cluster 2022

Berikut pada Gambar 4.43 merupakan *source code* untuk menambahkan kolom *cluster* dalam *DataFrame* *df*.

```
[24] # Menambahkan kolom "cluster" ke dalam DataFrame df dengan label klaster hasil pengelompokan dari kmeans_pca
df["cluster"] = kmeans_pca.labels_
# Menampilkan DataFrame df yang sudah diperbarui dengan kolom klaster
df
```

Gambar 4. 43 *Source Code* untuk Menambahkan kolom *cluster*

Pada Gambar 4.44, Gambar 4.45, dan Gambar 4.46 merupakan tampilan *DataFrame* setelah ditambahkan kolom *cluster*.

Kabupaten_kota	Jumlah Penduduk Miskin (ribu) 2020	Tingkat Pengangguran Terbuka (persentase %) 2020	Indeks Pembangunan Manusia (desimal) 2020	Rata-rata Pengeluaran Perkapita (rupiah) 2020	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2020	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2020	cluster
Pacitan	80.02	2.28	69.28	789080	84.43	88.90	0
Ponorogo	86.74	4.45	71.57	1018283	91.15	84.97	0
Trenggalek	81.06	4.11	70.10	863724	80.64	78.81	0
Tulungagung	78.40	4.61	73.15	1032860	96.75	88.45	0
Blitar	108.55	3.82	71.02	805475	85.35	80.36	0
Kediri	179.93	5.24	72.80	897042	98.09	83.59	0

Gambar 4. 44 *DataFrame* dengan Kolom *cluster* 2020

Kabupaten_kota	Jumlah Penduduk Miskin (ribu) 2021	Tingkat Pengangguran Terbuka (persentase %) 2021	Indeks Pembangunan Manusia (desimal) 2021	Rata-rata Pengeluaran Perkapita (rupiah) 2021	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2021	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2021	cluster
Pacitan	84.19	2.04	69.45	689363	76.71	71.70	0
Ponorogo	89.94	4.38	71.81	923418	96.30	87.48	0
Trenggalek	84.89	3.53	70.40	791239	75.78	75.48	0
Tulungagung	78.59	4.91	73.29	1074754	97.40	85.37	0
Blitar	112.82	3.69	71.48	832222	83.33	75.67	0
Kediri	184.49	5.15	73.31	942363	96.35	88.92	0

Gambar 4. 45 *DataFrame* dengan Kolom *cluster* 2021

Kabupaten_Kota	Jumlah Penduduk Miskin (ribu) 2022	Tingkat Pengangguran Terbuka (persentase %) 2022	Indeks Pembangunan Manusia (desimal) 2022	Rata-rata Pengeluaran Perkapita (rupiah) 2022	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2022	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2022	cluster
Pacitan	76.03	3.85	70.19	762721	83.95	72.04	4
Ponorogo	81.80	5.51	72.59	962454	91.50	86.77	1
Trenggalek	76.75	5.37	71.28	859877	81.23	79.27	4
Tulungagung	70.52	6.65	74.14	1200017	96.39	83.60	1
Blitar	101.84	5.45	72.25	946248	96.37	80.11	1
Kediri	169.66	6.03	74.20	1081593	92.17	88.68	1

Gambar 4. 46 DataFrame dengan Kolom cluster 2022

Pada Gambar 4.47 di bawah berikut merupakan *source code* untuk mengimport file geojson koordinat setiap daerah kabupaten/kota yang ada pada Provinsi Jawa Timur.

```
[139] df_geo = gpd.read_file('https://raw.githubusercontent.com/MuhammadWandias/geojson-jatim/main/jawa-timur-simplified-topo.json')
```

Gambar 4. 47 Source Code untuk Import File geojson Titik Koordinat Daerah

Pada Gambar 4.48 di bawah berikut merupakan *source code* untuk mengubah bagian kolom [29:38] dengan kolom ['kabkot'].

```
[142] data_to_replace = df.index[29:38] # Ambil data yang ingin diganti dari df
df_geo['kabkot'][29:38] = data_to_replace.values # Ganti data di df_geo dengan data dari df
df_geo
```

Gambar 4. 48 Source Code untuk replace kolom dari file gojson

Pada Gambar 4.49 di bawah ini merupakan tampilan hasil *replace* kolom dari file geojson yang telah di import sebelum nya.

	id	provinsi	kabkot	geometry
0	35-01	Jawa Timur	Pacitan	POLYGON (((111.27317 -7.93735, 111.29941 -7.941...
1	35-02	Jawa Timur	Ponorogo	POLYGON (((111.75239 -7.80111, 111.75986 -7.813...
2	35-03	Jawa Timur	Trenggalek	POLYGON (((111.75549 -7.90814, 111.75986 -7.926...
3	35-04	Jawa Timur	Tulungagung	POLYGON (((111.79904 -7.83663, 111.79740 -7.848...
4	35-05	Jawa Timur	Blitar	POLYGON (((112.44644 -7.91114, 112.45227 -7.918...
5	35-06	Jawa Timur	Kediri	POLYGON (((112.42221 -7.76685, 112.41838 -7.774...
6	35-07	Jawa Timur	Malang	MULTIPOLYGON (((112.69881 -8.42990, 112.71229 ...
7	35-08	Jawa Timur	Lumajang	POLYGON (((113.29519 -7.92756, 113.30430 -7.926...
8	35-09	Jawa Timur	Jember	MULTIPOLYGON (((113.30102 -8.44710, 113.31669 ...

Gambar 4. 49 DataFrame dari gojson Setelah di Replace

Pada Gambar 4.50 merupakan *source code* untuk menambahkan kolom 'kabkot' untuk menggabungkan data ke df_geo.

```
[29] df_cluster['kabkot'] = df.index # Menambahkan kolom kabkot untuk menggabungkan data dengan df_geo
df_cluster.head()
```

Gambar 4. 50 Source Code untuk Menggabung Kolom 'kabkot' dalam df_geo

Pada Gambar 4.51, Gambar 4.52, dan Gambar 4.53 merupakan tampilan df_geo setelah ditambahkan kolom 'kabkot'.

Kabupaten_Kota	Jumlah Penduduk Risiko (ribu) 2020	Tingkat Pengangguran Terbuka (persentase %) 2020	Indeks Pembangunan Manusia (desimal) 2020	Rata-rata Pengeluaran Perkapita (rupiah) 2020	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2020	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2020	cluster	kabkot
Pacitan	80.82	2.28	69.28	789088	84.43	69.90	0	Pacitan
Ponorogo	86.74	4.45	71.57	1018283	91.15	84.97	0	Ponorogo
Trenggalek	81.06	4.11	70.10	863724	80.84	78.81	0	Trenggalek
Tulungagung	78.40	4.81	73.15	1032860	96.75	88.45	0	Tulungagung
Bitar	108.55	3.82	71.02	885478	86.35	80.38	0	Bitar

Gambar 4. 51 df_cluster 2020

Kabupaten_Kota	Jumlah Penduduk Risiko (ribu) 2021	Tingkat Pengangguran Terbuka (persentase %) 2021	Indeks Pembangunan Manusia (desimal) 2021	Rata-rata Pengeluaran Perkapita (rupiah) 2021	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2021	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2021	cluster	kabkot
Pacitan	84.19	2.04	69.45	689383	78.71	71.70	0	Pacitan
Ponorogo	89.94	4.38	71.81	923419	96.30	87.40	0	Ponorogo
Trenggalek	84.89	3.53	70.40	791239	75.78	75.48	0	Trenggalek
Tulungagung	78.59	4.91	73.28	1074754	87.40	85.37	0	Tulungagung
Bitar	112.82	3.00	71.48	932222	83.33	75.67	0	Bitar

Gambar 4. 52 df_cluster 2021

Kabupaten_Kota	Jumlah Penduduk Risiko (ribu) 2022	Tingkat Pengangguran Terbuka (persentase %) 2022	Indeks Pembangunan Manusia (desimal) 2022	Rata-rata Pengeluaran Perkapita (rupiah) 2022	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2022	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2022	cluster	kabkot
Pacitan	78.93	3.65	70.19	782721	83.95	72.64	4	Pacitan
Ponorogo	81.80	5.51	72.59	962454	91.50	86.77	1	Ponorogo
Trenggalek	76.75	5.37	71.28	858877	81.23	78.27	4	Trenggalek
Tulungagung	70.52	6.65	74.14	1200017	96.39	83.60	1	Tulungagung
Bitar	101.94	5.45	72.25	945248	96.37	80.11	1	Bitar

Gambar 4. 53 df_cluster 2022

Pada Gambar 4.54 merupakan *source code* untuk menggabungkan dua *DataFrame*, yaitu *df_geo* dan *df_cluster*, berdasarkan kolom 'kabkot' dengan metode penggabungan yang digunakan adalah 'inner', artinya hanya baris-baris yang memiliki nilai yang cocok pada kedua *DataFrame* yang akan dimasukkan ke dalam *DataFrame* hasil gabungan.

```
[30] df_merged = pd.merge(df_geo, df_cluster, on='kabkot', how='inner')
```

Gambar 4. 54 *Source Code* untuk Menggabungkan *df_geo* dan *df_cluster*
Pada Gambar 4.55, Gambar 4.56, dan Gambar 4.57 adalah tampilan atau hasil dari penggabungan antara *df_geo* dan *df_cluster*.

#df_merged	id provinsi	kabkot	geoetry	Jumlah Penduduk Risiko (ribu) 2020	Tingkat Pengangguran Terbuka (persentase %) 2020	Indeks Pembangunan Manusia (desimal) 2020	Rata-rata Pengeluaran Perkapita (rupiah) 2020	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2020	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2020	cluster	
0	35-01	Jawa Timur	Pacitan	POLYGON ((111.27317 -7.93735, 111.29941 -7.941...	80.82	2.28	69.28	789088	84.43	69.90	0
1	35-02	Jawa Timur	Ponorogo	POLYGON ((111.75239 -7.80111, 111.75086 -7.813...	86.74	4.45	71.57	1018283	91.15	84.97	0
2	35-03	Jawa Timur	Trenggalek	POLYGON ((111.75549 -7.90814,	81.06	4.11	70.10	863724	80.84	78.81	0

Gambar 4. 55 Hasil Penggabungan *df_geo* dan *df_cluster* 2020

#f_merged											
	id provinsi	kabkot	geoetry	Jumlah Penduduk Miskin (ribu) 2021	Tingkat Pengangguran Terbuka (persentase %) 2021	Indeks Pembangunan Manusia (desimal) 2021	Rata-rata Pengeluaran Perkapita (rupiah) 2021	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2021	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2021	cluster	
0	35-01	Jawa Timur	Pacitan	POLYGON ((111.27317 -7.93735, 111.29941 -7.94111, 111.27317 -7.94111, 111.27317 -7.93735))	84.18	2.04	69.45	669363	70.71	71.70	0
1	35-02	Jawa Timur	Ponorogo	POLYGON ((111.75239 -7.80111, 111.75988 -7.81311, 111.75239 -7.81311, 111.75239 -7.80111))	89.94	4.38	71.81	923419	90.30	87.48	0
2	35-03	Jawa Timur	Trenggalek	POLYGON ((111.75549 -7.90814, 111.75549 -7.90814, 111.75549 -7.90814, 111.75549 -7.90814))	84.89	3.53	70.40	701239	75.78	75.48	0

Gambar 4. 56 Hasil Penggabungan df_geo dan df_cluster 2021

#f_merged											
	id provinsi	kabkot	geoetry	Jumlah Penduduk Miskin (ribu) 2022	Tingkat Pengangguran Terbuka (persentase %) 2022	Indeks Pembangunan Manusia (desimal) 2022	Rata-rata Pengeluaran Perkapita (rupiah) 2022	Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak (persentase %) 2022	Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak (persentase %) 2022	cluster	
0	35-01	Jawa Timur	Pacitan	POLYGON ((111.27317 -7.93735, 111.29941 -7.94111, 111.27317 -7.94111, 111.27317 -7.93735))	76.93	3.85	70.19	762721	83.95	72.64	4
1	35-02	Jawa Timur	Ponorogo	POLYGON ((111.75239 -7.80111, 111.75988 -7.81311, 111.75239 -7.81311, 111.75239 -7.80111))	81.80	5.51	72.59	962454	91.50	86.77	1
2	35-03	Jawa Timur	Trenggalek	POLYGON ((111.75549 -7.90814, 111.75549 -7.90814, 111.75549 -7.90814, 111.75549 -7.90814))	76.75	5.37	71.28	856877	81.23	79.27	4

Gambar 4. 57 Hasil Penggabungan df_geo dan df_cluster 2022

Pada Gambar 4.58 merupakan source code untuk import library folium.

```
import folium
from folium import features
```

Gambar 4. 58 Import Library folium

Pada Gambar 4.59 merupakan source code untuk visualisasi hasil cluster ke dalam bentuk peta dengan menggunakan dua library yaitu geopandas dan folium.

```
[41] # Inisialisasi peta
m = folium.Map(location=[-7.250445, 112.768845], zoom_start=8)
# Mengatur warna untuk setiap cluster
colors = ['orange', 'green']
# Mengiterasi melalui kluster unik yang sudah diurutkan dan menambahkan ke peta
for i, cluster in enumerate(sorted(df_merged['cluster'].unique())):
    # Filter data berdasarkan cluster
    cluster_data = df_merged[df_merged['cluster'] == cluster]

    # Tambahkan setiap baris dalam cluster ke peta
    for _, row in cluster_data.iterrows():
        geo_json = folium.GeoJson(
            row['geometry'].__geo_interface__,
            style_function=lambda feature, color=colors[i % len(colors)]: {
                'fillColor': color,
                'color': color,
                'weight': 0.8,
                'fillOpacity': 0.5,
            })
        geo_json.add_to(m)

# Menambahkan legenda kustom ke dalam peta
legend_html = '''
<div style="position: fixed;
bottom: 50px; left: 50px; width: 150px; height: 160px;
border:2px solid grey; z-index:9999; font-size:14px;
background-color:white;
">
    &nbsp; <b>Legenda</b> <br>
    &nbsp; Cluster 1 &nbsp; <i class="fa fa-square" style="color:orange"></i><br>
    &nbsp; Cluster 2 &nbsp; <i class="fa fa-square" style="color:green"></i><br>
</div>
'''
m.get_root().html.add_child(folium.Element(legend_html))
```

Gambar 4. 59 Source Code untuk visualisasi hasil cluster

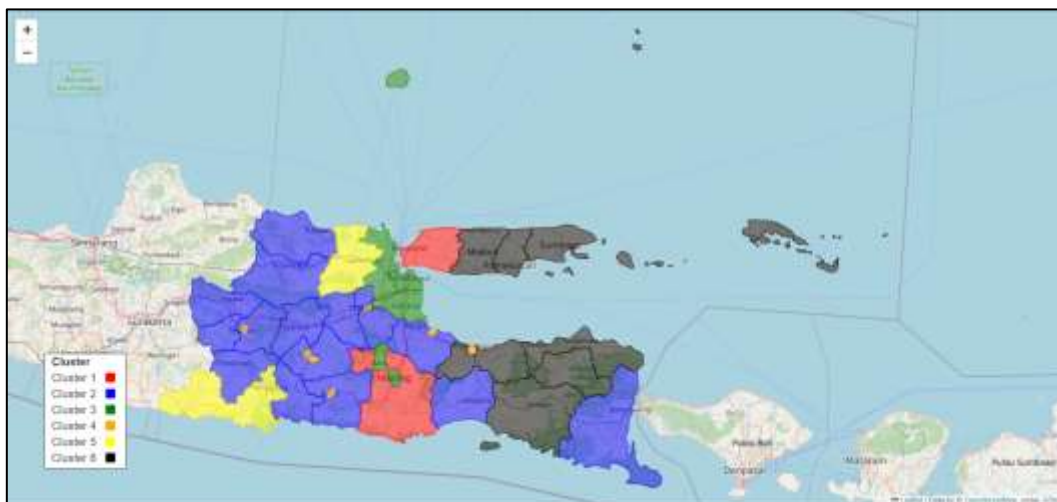
Pada Gambar 5.60, Gambar 5.61, dan Gambar 5.62 di bawah merupakan tampilan dari visualisasi hasil *cluster* ke dalam bentuk peta.



Gambar 4. 60 Hasil Visualisasi 2020



Gambar 4. 61 Hasil Visualisasi 2021



Gambar 4. 62 Hasil Visualisasi 2022

Setelah diketahui jumlah cluster optimum untuk pengelompokan kabupaten dan kota di Jawa Timur, selanjutnya dilakukan identifikasi karakteristik untuk setiap cluster. Hasil

identifikasi karakteristik tiap kelompok cluster disajikan dalam Tabel 4.2 pada tahun 2020, Tabel 4.3 pada tahun 2021, dan Tabel 4.4 pada tahun 2022.

Tabel 4. 2 Identifikasi Karakteristik *Cluster* Tahun 2020

Cluster 2 (Sejahtera)			
Anggota <i>Cluster</i>	Kota Kediri Kota Blitar Kota Malang Kota Probolinggo Kota Pasuruan	Kota Mojokerto Kota Madiun Kota Surabaya Kota Batu Sidoarjo	Mojokerto Jombang Gresik
Karakteristik	<ul style="list-style-type: none"> - Klaster ini memiliki jumlah penduduk miskin lebih rendah dari pada cluster 1 -Tingkat pengangguran terbuka lebih banyak dari klaster 1 -Rata-rata pengeluaran perkapita lebih banyak dari klaster 1 -Akses sanitasi layak lebih tinggi dari klaster 1 -Indeks pembangunan manusia (IPM) lebih tinggi dari klaster 1 -Akses air minum layak stabil. 		
Cluster 1 (Kurang Sejahtera)			
Anggota <i>Cluster</i>	Pacitan Ponorogo Trenggalek Tulungagung Blitar Kediri Malang Lumajang Jember	Banyuwangi Bondowoso Situbondo Probolinggo Pasuruan Nganjuk Madiun Magetan Ngawi	Bojonegoro Tuban Lamongan Bangkalan Sampang Pamekasan Sumenep
Karakteristik	<ul style="list-style-type: none"> -Klaster ini memiliki jumlah penduduk miskin lebih tinggi dari pada cluster 2 -Tingkat pengangguran terbuka lebih sedikit dari klaster 2 -Rata-rata pengeluaran perkapita lebih sedikit dari klaster 2 -Akses sanitasi layak lebih rendah dari klaster 2 -Indeks pembangunan manusia (IPM) lebih rendah dari klaster 2 -Akses air minum layak stabil. 		

Tabel 4. 3 Identifikasi Karakteristik *Cluster* Tahun 2021

Cluster 2 (Sejahtera)			
Anggota <i>Cluster</i>	Kota Kediri Kota Blitar Kota Malang Kota Probolinggo	Kota Pasuruan Kota Mojokerto Kota Madiun Kota Surabaya	Kota Batu Sidoarjo Gresik
Karakteristik	<ul style="list-style-type: none"> - Klaster ini memiliki jumlah penduduk miskin lebih rendah dari pada cluster 1 -Tingkat pengangguran terbuka lebih banyak dari klaster 1 -Rata-rata pengeluaran perkapita lebih banyak dari klaster 1 -Akses sanitasi layak lebih tinggi dari klaster 1 -Indeks pembangunan manusia (IPM) lebih tinggi dari klaster 1 -Akses air minum layak stabil. 		
Cluster 1 (Kurang Sejahtera)			
Anggota <i>Cluster</i>	Pacitan Ponorogo	Banyuwangi Bondowoso	Magetan Ngawi

	Trenggalek Tulungagung Blitar Kediri Malang Lumajang Jember	Situbondo Probolinggo Pasuruan Mojokerto Jombang Nganjuk Madiun	Bojonegoro Tuban Lamongan Bangkalan Sampang Pamekasan Sumenep
Karakteristik	<ul style="list-style-type: none"> -Klaster ini memiliki jumlah penduduk miskin lebih tinggi dari pada cluster 2 -Tingkat pengangguran terbuka lebih sedikit dari klaster 2 -Rata-rata pengeluaran perkapita lebih sedikit dari klaster 2 -Akses sanitasi layak lebih rendah dari klaster 2 -Indeks pembangunan manusia (IPM) lebih rendah dari klaster 2 -Akses air minum layak stabil. 		

Identifikasi Karakteristik *cluster* pada Tabel 4.4 digunakan untuk menjelaskan karakter dari setiap *cluster* yang dihasilkan oleh algoritma *K-Means clustering* dengan melihat dari nilai rata-rata setiap variabel dalam kesejahteraan di Jawa Timur untuk tahun 2022. Dengan jumlah *cluster* sebanyak 6 pada tahun 2022, maka pada deskripsi karakteristiknya untuk indikator Jumlah Penduduk miskin, Tingkat pengangguran terbuka(TPT), dan rata-rata pengeluaran perkapita di bedakan dengan urutan dari skala 1 sampai 6, semakin mendekati angka satu maka semakin besar/banyak, sebaliknya jika semakin mendekati angka enam maka semakin kecil/sedikit, kemudian untuk indikator Indeks pembangunan manusia (IPM), Persentase Rumah Tangga yang Memiliki Akses Sumber Air Minum Layak, dan Persentase Rumah Tangga yang Memiliki Akses Terhadap Layanan Sanitasi Layak di bedakan dengan urutan dari skala 1 sampai 6, semakin mendekati angka enam maka semakin besar/banyak, sebaliknya jika semakin mendekati angka satu maka semakin kecil/sedikit.

Tabel 4. 4 Identifikasi Karakteristik *Cluster* Tahun 2022

Cluster 3 (Paling Sejahtera)			
Anggota <i>Cluster</i>	Sidoarjo Gresik	Kota Malang Kota Surabaya	Kota Batu
Karakteristik	Klaster ini memiliki IPM tertinggi (urutan 2) dan rata-rata pengeluaran per kapita tertinggi (urutan 2). Klaster ini juga memiliki akses terhadap air minum layak dan sanitasi layak yang paling baik (urutan 1). Oleh karena itu, klaster ini dikategorikan sebagai klaster yang paling sejahtera.		
Cluster 4 (Sejahtera)			
Anggota <i>Cluster</i>	Kota Kediri Kota Blitar	Kota Probolinggo Kota Pasuruan	Kota Mojokerto Kota Madiun
Karakteristik	Klaster ini memiliki IPM yang cukup tinggi (urutan 4) dan akses terhadap air minum layak yang cukup baik (urutan 6). Klaster ini juga memiliki tingkat pengangguran terbuka yang rendah (urutan 5). Oleh karena itu, klaster ini dikategorikan sebagai klaster yang sejahtera.		
Cluster 1 (Cukup Sejahtera)			
Anggota <i>Cluster</i>	Malang	Bangkalan	
Karakteristik	Klaster ini memiliki tingkat pengangguran terbuka yang rendah (urutan 2) dan akses terhadap sanitasi layak yang		

	cukup baik (urutan 5). Klaster ini juga memiliki IPM yang sedang (urutan 5). Oleh karena itu, klaster ini dikategorikan sebagai klaster yang cukup sejahtera.		
Cluster 2 (Kurang Kesejahteraan)			
Anggota Cluster	Ponorogo Tulungagung Blitar Kediri Lumajang	Banyuwangi Pasuruan Mojokerto Jombang Nganjuk	Madiun Magetan Ngawi Bojonegoro Tuban
Karakteristik	Klaster ini memiliki IPM yang sedang (urutan 3) dan akses terhadap air minum layak yang cukup baik (urutan 3). Klaster ini juga memiliki rata-rata pengeluaran per kapita yang sedang (urutan 3). Oleh karena itu, klaster ini dikategorikan sebagai klaster kurang sejahtera.		
Cluster 5 (Sangat Kurang Sejahtera)			
Anggota Cluster	Pacitan	Trenggalek	Lamongan
Karakteristik	Klaster ini memiliki jumlah penduduk miskin yang tinggi (urutan 2) dan rata-rata pengeluaran per kapita yang rendah (urutan 5). Klaster ini juga memiliki akses terhadap sanitasi layak yang kurang baik (urutan 6). Oleh karena itu, klaster ini dikategorikan sebagai klaster sangat kurang sejahtera.		
Cluster 6 (Paling Kurang Sejahtera)			
Anggota Cluster	Jember Bondowoso Situbondo	Probolinggo Sampang Pamekasan	Sumenep
Karakteristik	Klaster ini memiliki jumlah penduduk miskin yang paling tinggi (urutan 6) dan rata-rata pengeluaran per kapita yang paling rendah (urutan 6). Klaster ini juga memiliki akses terhadap air minum layak dan sanitasi layak yang kurang baik (urutan 5 dan 4). Oleh karena itu, klaster ini dikategorikan sebagai klaster yang paling kurang sejahtera.		

Pada Gambar 4.63 Merupakan tampilan utama dari pengembangan visualisasi hasil *cluster* Jawa Timur, halaman utama ini memuat keterangan dimana pada tabel yang tertera adalah faktor-faktor yang dapat mempengaruhi tingkat kesejahteraan dan di bawah tabel tersebut juga terdapat penjelasan lebih lanjut untuk setiap faktor yang ada pada tabel.



Gambar 4. 63 Hasil Visualisasi Website Halaman Utama

Pada Gambar 4.64 Merupakan tampilan halaman cluster Jawa Timur dari pengembangan visualisasi hasil *cluster*, halaman ini memuat tampilan peta beserta *cluster* nya dan dibawah peta tersebut terdapat identifikasi karakteristik clusternya dalam bentuk tabel untuk memudahkan pembaca dalam memahami arti setiap *cluster*.



Gambar 4. 64 Hasil Visualisasi Website Halaman Cluster Jawa Timur

KESIMPULAN

Berdasarkan hasil analisis dari penelitian yang sudah dilakukan, didapatkan kesimpulan sebagai berikut:

1. Dari pengolahan data pada tahun 2020, 2022, dan 2022 menggunakan algoritma K-Means, dihasilkan 2 *cluster* pada tahun 2020 dimana *cluster* 2 merupakan *cluster* daerah sejahtera dan *cluster* 1 merupakan *cluster* daerah kurang sejahtera. 2 *cluster* pada tahun 2021 dimana *cluster* 2 merupakan *cluster* daerah sejahtera dan *cluster* 1 merupakan *cluster* daerah kurang sejahtera. 6 *cluster* pada tahun 2022 dimana *cluster* 3 merupakan *cluster* daerah paling sejahtera, *cluster* 4 merupakan *cluster* daerah sejahtera, *cluster* 1 merupakan *cluster* daerah cukup sejahtera, *cluster* 2 merupakan *cluster* daerah kurang sejahtera, *cluster* 5 merupakan *cluster* daerah sangat kurang sejahtera, *cluster* 6 merupakan *cluster* daerah paling kurang sejahtera. *Cluster* 1 pada tahun 2020 terdiri dari 25 Kota/Kabupaten, sedangkan *cluster* 2 terdiri dari 13 Kota/Kabupaten. *Cluster* 1 pada tahun 2021 terdiri dari 27 Kota/Kabupaten, sedangkan *cluster* 2 terdiri dari 11 Kota/Kabupaten. Pada 2022

cluster 1 terdiri dari 2 Kabupaten/Kota. Kemudian *cluster* 2 terdiri dari 15 Kabupaten/Kota. Kemudian *cluster* 3 terdiri dari 5 Kabupaten/Kota. Kemudian *cluster* 4 terdiri dari 6 Kabupaten/Kota. Kemudian *cluster* 5 terdiri dari 3 Kabupaten/Kota. Sedangkan *cluster* 6 terdiri dari 7 Kabupaten/Kota. Setiap *cluster* nya memiliki karakteristiknya masing-masing. Dari hasil penelitian ini diharapkan agar daerah dengan tingkat kesejahteraan rendah dapat ditinjau atau dilakukan pengentasan kemiskinan oleh instansi terkait.

2. Visualisasi dalam bentuk peta dilakukan menggunakan 2 *library* python yaitu *geopandas* dan *folium*. *Geopandas* digunakan untuk memuat data berupa koordinat *polygon* yang membentuk wilayah di setiap daerah, agar dapat diberikan warna sesuai dengan *cluster* masing-masing. *Folium* digunakan sebagai kanvas tampilan peta agar tampilannya lebih interaktif.

DAFTAR PUSTAKA

- Afifah Nur, A. (t.t.). *Journal of Student Development Information System (JoSDIS)*.
- Al Farizi, M. F., Harianto, F., Dewanti, M. S., Siburian, C. A., Mardianto, M. F. F., Amelia, D., & Ana, E. (2023). Pengelompokan Daerah di Jawa Timur Berbasis Indikator Kesejahteraan Masyarakat dengan Pendekatan Analisis Cluster Hierarki. *Inferensi*, 6(2), 141. <https://doi.org/10.12962/j27213862.v6i2.15452>
- Amalia, A., & Cahyono, A. B. (t.t.). *Analisis Pemanfaatan Playwright untuk Pengujian Otomatis Aplikasi Berbasis Web (Studi Kasus: Sistem Manajemen Jaringan)*.
- Angelina M. T. I. Sambu Ua, Diandra Lestriani H, Elizabeth Sonia Kristanty Marpaung, Jesslyn Ong, Michelle Savinka, Putri Nurhaliza, & Rahmi Yulia Ningsih. (2023). Penggunaan Bahasa Pemrograman Python Dalam Analisis Faktor Penyebab Kanker Paru-Paru. *Jurnal Publikasi Teknik Informatika*, 2(2), 88–99. <https://doi.org/10.55606/jupti.v2i2.1742>
- Ayu, D., Dewi, I. C., & Pramita, K. (2019). Analisis Perbandingan Metode Elbow dan Sillhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. Dalam *JURNAL MATRIX* (Vol. 9, Nomor 3).
- Bali, P. N., & Raya Bukit, J. (t.t.). *Implementasi Python Folium dalam Pembangunan Sistem Informasi Peta Interaktif Cagar Budaya Provinsi Bali Oleh : Gde Brahupadhya Subiksa, Made Pasek Agus Ariawan, Ida Bagus Adisimakrisna Peling Implementasi Python Folium dalam Pembangunan Sistem Informasi Peta Interaktif Cagar Budaya Provinsi Bali 1) Gde Brahupadhya Subiksa 2) Made Pasek Agus Ariawan 3) Ida Bagus Adisimakrisna Peling*. 8(2). http://ejournal.ust.ac.id/index.php/Jurnal_Means/
- Christianto, Y., Intan, R., & Adipranata, R. (t.t.). *Penerapan Metode Klasifikasi C4.5 dalam Pembuatan Website Identifikasi untuk Prediksi Kredibilitas Akun pada Media Sosial Instagram*.
- Dirjen, S. K., Riset, P., Pengembangan, D., Dikti, R., Ramadar, P., Saputra, N., & Chusyairi, A. (2017). Terakreditasi SINTA Peringkat 2 Perbandingan Metode Clustering dalam Pengelompokan Data Puskesmas pada Cakupan Imunisasi Dasar Lengkap. *masa berlaku mulai*, 1(3), 1077–1084.
- Dirjen, S. K., Riset, P., Pengembangan, D., Dikti, R., Yaumi, A. S., Zulfikar, Z., & Nugroho, A. (2018). *Terakreditasi SINTA Peringkat 4 Klasterisasi Karakter Konsumen Terhadap Kecenderungan Pemilihan Produk Menggunakan K-Means* (Vol. 3, Nomor 1).
- Faqih, A., Chalik, A. M., Qowy, B. A., Hanafi, F., Nuraminah, A., Studi, P., Komputer, I., Tinggi, S., & Manajemen, I. (2023). *Klasifikasi Mata Terbuka dan Tertutup Berdasarkan Sinyal EEG Menggunakan Artificial Neural Network dan Principal Component Analysis*.

- Ferdyandi, M., Setiawan, N. Y., & Abdurrachman Bachtiar, F. (2022). *Prediksi Potensi Penjualan Makanan Beku berdasarkan Ulasan Pengguna Shopee menggunakan Metode Decision Tree Algoritma C4.5 dan Random Forest (Studi Kasus Dapur Lilis)* (Vol. 6, Nomor 2). <http://j-ptiik.ub.ac.id>
- Hafiz, Y. A., & Sudarmilah, E. (t.t.). *IMPLEMENTASI WEB SCRAPING PADA PORTAL BERITA ONLINE*.
- Handoko, S., Fauziah, F., & Handayani, E. T. E. (2020). IMPLEMENTASI DATA MINING UNTUK MENENTUKAN TINGKAT PENJUALAN PAKET DATA TELKOMSEL MENGGUNAKAN METODE K-MEANS CLUSTERING. *Jurnal Ilmiah Teknologi dan Rekayasa*, 25(1), 76–88. <https://doi.org/10.35760/tr.2020.v25i1.2677>
- Harsono, S., Dwi Prihatin, T., Sadad, A., & Maulina, D. (2023). Penerapan Algoritma K-Means Untuk Pemetaan Biodiversity Kayu Bulat Di Indonesia Application of the K-Means Algorithm for Mapping Roundwood Biodiversity in Indonesia. *Cogito Smart Journal* /, 9(1).
- Haryanto, C., Rahaningsih, N., & Muhammad Basysyar, F. (2023). KOMPARASI ALGORITMA MACHINE LEARNING DALAM MEMREDIKSI HARGA RUMAH. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Nomor 1).
- Hermanto, K., Salim, D., Wu, B., Regina Salim, O., & Gunadi Belinda, R. (t.t.). Penggunaan Python Untuk Menganalisis Pola Penyebaran Covid-19 Di Masa Pandemi. *Journal of Student Development Information System (JoSDIS)*.
- Kosasih, R. (2021). Pengenalan Wajah Menggunakan PCA dengan Memperhatikan Jumlah Data Latih dan Vektor Eigen. *Jurnal Informatika Universitas Pamulang*, 6(1), 1. <https://doi.org/10.32493/informatika.v6i1.7261>
- Kusumo, S., & Somya, R. (2022). Penerapan Web Scraping Deskripsi Produk Menggunakan Selenium Python Dan Framework Laravel. *Jurnal Teknik Informatika dan Sistem Informasi*, 9(4). <http://jurnal.mdp.ac.id>
- Laksono Suryoputro, A., Yulianto Joko Prasetyo, S., Studi Teknik Informatika, P., & Teknologi Informasi Universitas Kristen Satya Wacana Jl Notohamidjodjo Blotongan, F. O. (2023). Prediksi dan visualisasi penyakit COVID-19 menggunakan kombinasi Prophet dan GeoPandas. *AITI: Jurnal Teknologi Informasi*, 20(Agustus), 135–149.
- Luh, N., Sukmayanti, K. A., Kadek, I., Asmarajaya, A., & Sanjaya, K. O. (t.t.). *PEMETAAN NASABAH TABUNGAN PADA LPD DESA ADAT SUMERTA MENGGUNAKAN METODE VISUALISASI DATA*.
- Miralda, V., Zarlis, M., & Irawan, E. (2020). Penerapan Metode K-Means Clustering Untuk Daging Ayam Buras. *Technology and Science (BITS)*, 2(2). <https://www.bps.go.id>.
- Mulia, R. A., & Saputra, N. (t.t.). *ANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI KESEJAHTERAAN MASYARAKAT KOTA PADANG*.
- Muttaqin, M. R., & Defriani, M. (2020). Algoritma K-Means untuk Pengelompokan Topik Skripsi Mahasiswa. *ILKOM Jurnal Ilmiah*, 12(2), 121–129. <https://doi.org/10.33096/ilkom.v12i2.542.121-129>
- Novianto, D., Sugihartono, T., Luhur Pangkalpinang Jl Jenderal Sudirman, A., Selindung Baru, K., Gabek, K., & Pangkal Pinang, K. (t.t.). *Sistem Deteksi Kualitas Buah Jambu Air Berdasarkan Warna Kulit Menggunakan Algoritma Principal Component Analysis (Pca) dan K-Nearest Neighbor (K-NN)*.
- Nurohmah, Y., Mayasari, R., & Nurina Sari, B. (2023). OPTIMALISASI PERFORMA K-MEANS CLUSTERING DENGAN PCA DALAM ANALISIS TINGKAT KEMISKINAN DI JAWA BARAT. Dalam *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Nomor 3).

- Nyoman, I., & Adiputra, M. (2021). CLUSTERING PENYAKIT DBD PADA RUMAH SAKIT DHARMA KERTI MENGGUNAKAN ALGORITMA K-MEANS. *INSERT: Information System and Emerging Technology Journal*, 2(2), 99.
- Optimasi, I., Gridsearchcv, H., Sistem, P., Serangan..., P., Maisat, Z., Darmawan, E., Fauzan Dianta, A., & Korespondensi, *. (2023). Implementasi Optimasi Hyperparameter GridSearchCV Pada Sistem Prediksi Serangan Jantung Menggunakan SVM. *Online) Teknologi: Jurnal Ilmiah Sistem Informasi*, 13(1), 8–15. <https://doi.org/10.26594/teknologi.v13i1.3098>
- Pahlevi, R., Prayogi, M. W., Rasywir, E., & Pratama, Y. (2022). Penerapan Algoritma K-Means Pada Penyebaran Covid-19 Di Provinsi Jambi. *Journal of Computer System and Informatics (JoSYC)*, 4(1), 229–237. <https://doi.org/10.47065/josyc.v4i1.2550>
- Penerapan Metode Clustering Dengan Algoritma K-Means Untuk Analisa Persebaran Varian Covid-19 (Studi Kasus Kelurahan Antapani Kidul)*. (t.t.).
- Rahma Salsabila, A., Daffa, M., Kandias, M., Maulana, H., & Wahyuni, E. D. (t.t.). *Prosiding Seminar Nasional Teknologi dan Sistem Informasi (SITASI) 2022 IMPLEMENTASI TEKNIK WEB SCRAPING UNTUK MENAMPILKAN DATA TIM ENGLISH PREMIER LEAGUE IMPLEMENTATION OF WEB SCRAPING TECHNIQUE TO SHOWING ENGLISH PREMIER LEAGUE TEAM DATA*. <http://sitasi.upnjatim.ac.id/40>
- Regina Lo, Alfred Edbert Yunanto, Rasya Nuhaifa Movia, Lambert Aditama Soehardjianto, Ferdinand Wangsa, Natalie Asdyo Lidjaja, & Rahmi Yulia Ningsih. (2023). Penggunaan Bahasa Pemrograman Python dalam Menganalisis Hubungan Kualitas Kopi dengan Lokasi Pertanian Kopi. *Jurnal Publikasi Teknik Informatika*, 2(2), 100–109. <https://doi.org/10.55606/jupti.v2i2.1752>
- Ritonga, A. S., & Muhandhis, I. (t.t.). TEKNIK DATA MINING UNTUK MENGLASIFIKASIKAN DATA ULASAN DESTINASI WISATA MENGGUNAKAN REDUKSI DATA PRINCIPAL COMPONENT ANALYSIS (PCA). Dalam *Jurnal Ilmiah Edutic* (Vol. 7, Nomor 2).
- Riziq sirfatullah Alfarizi, M., Zidan Al-farish, M., Taufiqurrahman, M., Ardiansah, G., & Elgar, M. (2023). PENGGUNAAN PYTHON SEBAGAI BAHASA PEMROGRAMAN UNTUK MACHINE LEARNING DAN DEEP LEARNING. Dalam *Karimah Tauhid* (Vol. 2, Nomor 1).
- Rizquina, A. Z., & Ratnasari, C. I. (2023). Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(4), 377–383. <https://doi.org/10.47233/jteksis.v5i4.913>
- Samosir, H., Amin, M., & Harahap, I. R. (2021). Penerapan Data Mining untuk Klasifikasi Produk Merk Bata Menggunakan Algoritma K-Means. *JUTSI (Jurnal Teknologi dan Sistem Informasi)*, 1(2), 161–166. <https://doi.org/10.33330/jutsi.v2i1.1163>
- Sartika, D., & Saluza, I. (t.t.). *Penerapan Metode Principal Component Analysis (PCA) Pada Klasifikasi Status Kredit Nasabah Bank Sumsel Babel Cabang KM 12 Palembang Menggunakan Metode Decision Tree*.
- Srijani, N. (t.t.). *PERAN UMKM (USAHA MIKRO KECIL MENENGAH) DALAM MENINGKATKAN KESEJAHTERAAN MASYARAKAT 1) 2)*.
- Wicaksono, D. A., & Susetyo, Y. A. (2023a). CLUSTERING ZONASI DAERAH RAWAN BENCANA ALAM DI PROVINSI SUMATERA BARAT MENGGUNAKAN ALGORITMA K-MEANS DAN LIBRARY GEOPANDAS. *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, 4(2), 426–438. <https://doi.org/10.35870/jimik.v4i2.225>

- Wicaksono, D. A., & Susetyo, Y. A. (2023b). CLUSTERING ZONASI DAERAH RAWAN BENCANA ALAM DI PROVINSI SUMATERA BARAT MENGGUNAKAN ALGORITMA K-MEANS DAN LIBRARY GEOPANDAS. *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, 4(2), 426–438. <https://doi.org/10.35870/jimik.v4i2.225>
- Winarta, A., & Kurniawan, W. J. (2021). OPTIMASI CLUSTER K-MEANS MENGGUNAKAN METODE ELBOW PADA DATA PENGGUNA NARKOBA DENGAN PEMROGRAMAN PYTHON. *Jurnal Teknik Informatika Kaputama (JTIK)*, 5(1).
- Zuhal, N. K. (2022). Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering. Dalam *Universitas Nusantara PGRI Kediri. Kediri* (Vol. 1).