

**ANALISIS CLUSTERING TERHADAP SOSIAL-EKONOMI  
KABUPATEN/KOTA DI PROVINSI JAWA TIMUR  
MENGUNAKAN METODE PCA DAN HIERARCHICAL  
CLUSTERING**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**RIFQI HILAL BHAHARI**

**20.11.3434**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2024**

**ANALISIS CLUSTERING TERHADAP SOSIAL-EKONOMI  
KABUPATEN/KOTA DI PROVINSI JAWA TIMUR  
MENGUNAKAN METODE PCA DAN HIERARCHICAL  
CLUSTERING**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi S1 Informatika



disusun oleh

**RIFQI HILAL BHAHARI**

**20.11.3434**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2024**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**ANALISIS CLUSTERING TERHADAP SOSIAL-EKONOMI  
KABUPATEN/KOTA DI PROVINSI JAWA TIMUR MENGGUNAKAN  
METODE PCA DAN HIERARCHICAL CLUSTERING**

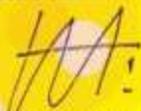
yang disusun dan diajukan oleh

**Rifqi Hilal Bhahari**

**20.11.3434**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 29/07/2024

Dosen Pembimbing,



Kusnawi, S.Kom, M. Eng.

NIK. 190302112

**HALAMAN PENGESAHAN**

**SKRIPSI**

**ANALISIS CLUSTERING TERHADAP SOSIAL-EKONOMI  
KABUPATEN/KOTA DI PROVINSI JAWA TIMUR MENGGUNAKAN  
METODE PCA DAN HIERARCHICAL CLUSTERING**

yang disusun dan diajukan oleh

**Rifqi Hilal Bhahari**

**20.11.3434**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 29/07/2024

**Susunan Dewan Penguji**

**Nama Penguji**

**Tonny Hidayat, S.Kom., M.Kom., Ph.D.**  
NIK. 190302182

**Ali Mustopa, M.Kom.**  
NIK. 190302192

**Kusnawi, S.Kom, M. Eng.**  
NIK. 190302112

**Tanda Tangan**



Skrripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 29/07/2024

**DEKAN FAKULTAS ILMU KOMPUTER**



**Hanif Al Fatta, S.Kom., M.Kom., Ph.D.**  
NIK. 190302096

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Rifqi Hilal Bhahari  
NIM : 20.11.3434

Menyatakan bahwa Skripsi dengan judul berikut:

### **Analisis Clustering Terhadap Sosial-Ekonomi Kabupaten/Kota Di Provinsi Jawa Timur Menggunakan Metode PCA Dan Hierarchical Clustering**

Dosen Pembimbing : Kusnawi, S.Kom, M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 29 Juli 2024

Yang Menyatakan,



Rifqi Hilal Bhahari

## HALAMAN PERSEMBAHAN

Segala puji dan syukur saya panjatkan kepada Allah SWT atas berkat, rahmat, dan hidayah-Nya yang telah memberikan kekuatan dan petunjuk sehingga saya dapat menyelesaikan skripsi ini dengan baik. Saya juga ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan kontribusi dalam penelitian dan penyusunan naskah ini. Skripsi ini saya persembahkan kepada:

1. Kedua Orang tua saya, Bapak Mujawid dan Ibu Hawa Yuswanti, saudara saya yang senantiasa memberikan dukungan, doa, serta kasih sayang sehingga saya dapat menempuh pendidikan dan menyelesaikan skripsi ini dengan baik.
2. Bapak Kusnawi, S.Kom., M.Eng. selaku pembimbing yang telah sabar dan tulus membimbing serta memberikan ilmu pengetahuan yang berharga kepada saya.
3. Kepada teman-teman saya, mengucapkan terima kasih yang sebesar-besarnya. Terima kasih atas kehadiran dan dukungan kalian, terutama di saat-saat sulit. Kalian selalu ada untuk memberikan semangat dan dorongan, sehingga saya bisa bangkit dan menyelesaikan skripsi ini. Ucapan terima kasih ini saya sampaikan kepada kalian yang telah menganggap saya sebagai teman.
4. Dengan penuh rasa syukur dan terima kasih, saya persembahkan skripsi ini kepada Windah Basudara. Terima kasih atas streaming-streaming yang telah menemani hari-hari saya di sela-sela mengerjakan skripsi ini. Setiap canda tawa dan momen seru yang tercipta dalam setiap streaming-mu telah membantu saya melewati masa-masa sulit dan memberikan energi positif untuk terus melangkah. Semoga kesuksesan dan kebahagiaan selalu menyertaimu.
5. Seluruh pihak keluarga besar Universitas Amikom Yogyakarta yang tidak dapat saya sebutkan satu per satu, atas segala bantuan dan dukungan yang telah diberikan.

Terima kasih atas semua doa, dukungan, serta bimbingan yang telah diberikan. Semoga hasil penelitian ini dapat bermanfaat bagi banyak pihak.

## KATA PENGANTAR

Segala puji dan syukur saya panjatkan kepada Allah SWT atas rahmat, berkah, dan hidayah-Nya, sehingga saya dapat menyelesaikan skripsi dengan judul “Analisis Clustering Terhadap Sosial-Ekonomi Kabupaten/Kota Di Provinsi Jawa Timur Menggunakan Metode PCA Dan Hierarchical Clustering” sebagai syarat untuk menyelesaikan Program Sarjana (S1) pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selama proses penyusunan skripsi ini, berbagai hambatan dan tantangan telah dihadapi. Namun, berkat bantuan dan bimbingan dari berbagai pihak, baik secara moral maupun spiritual, semua rintangan tersebut dapat dilalui. Pada kesempatan ini, saya ingin mengucapkan terima kasih kepada:

1. Bapak M. Suyanto, Prof., Dr., M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Kusnawi, S.Kom., M.Eng. yang dengan sabar dan tulus telah membimbing serta memberikan ilmu kepada saya.
3. Seluruh jajaran dosen Program Studi S1 Informatika Fakultas Ilmu Komputer yang telah memberikan bekal ilmu pengetahuan selama masa perkuliahan.
4. Semua pihak yang tidak dapat saya sebutkan satu per satu yang telah membantu sehingga penelitian ini bisa dilaksanakan.

Saya menyadari bahwa laporan skripsi ini masih jauh dari kata sempurna, oleh karena itu kritik dan saran yang membangun sangat saya harapkan. Akhir kata, semoga tulisan ini dapat memberikan manfaat bagi pembaca dan khususnya bagi penulis sendiri.

Yogyakarta, 24 Juli 2024

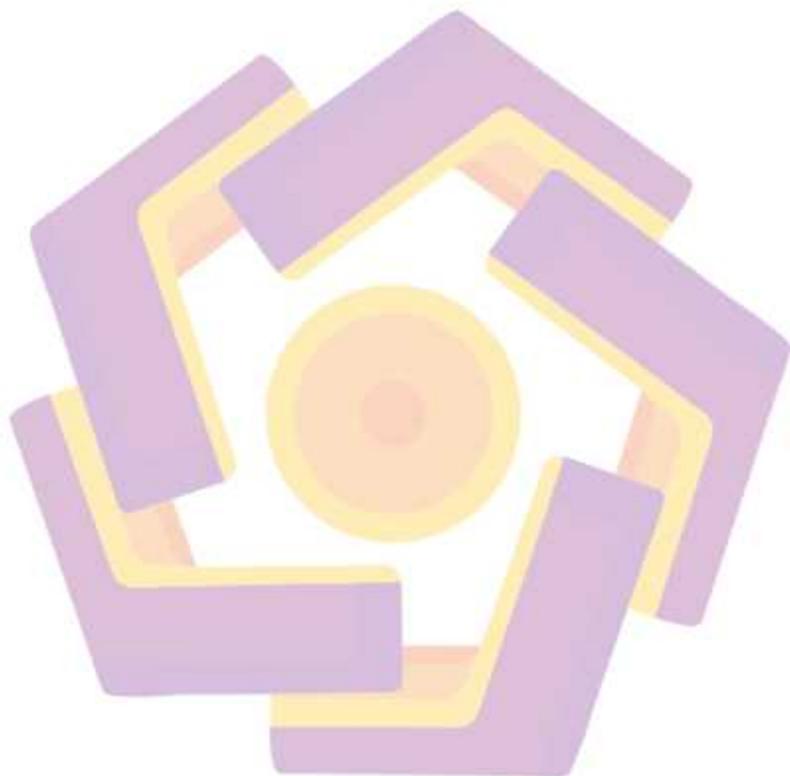
Penulis

## DAFTAR ISI

HALAMAN JUDUL .....	I
HALAMAN PERSETUJUAN .....	II
HALAMAN PENGESAHAN .....	III
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	IV
HALAMAN PERSEMBAHAN .....	V
KATA PENGANTAR.....	VI
DAFTAR ISI.....	VII
DAFTAR TABEL.....	X
DAFTAR GAMBAR.....	XI
DAFTAR LAMBANG DAN SINGKATAN .....	XII
DAFTAR ISTILAH .....	XIII
INTISARI .....	XIV
<i>ABSTRACT</i> .....	XV
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	4
1.6 Sistematika Penulisan .....	5
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>6</b>
2.1 Studi Literatur .....	6
2.2 Dasar Teori .....	14
2.2.1 Sosial Ekonomi .....	14
2.2.2 Clustering .....	14
2.2.3 Penanganan Outlier .....	16
2.2.4 Scaling Data (normalisasi) .....	16
2.2.5 Hierarchical Clustering .....	17
2.2.6 PCA(Principal Component Analysis) .....	20

2.2.7	Silhouette Index .....	22
2.2.8	Davies Boildin Index .....	23
2.2.9	Machine Learning .....	24
2.2.10	Python .....	24
<b>BAB III METODE PENELITIAN .....</b>		<b>25</b>
3.1	Gambaran Umum.....	25
3.2	Alur Penelitian.....	26
3.2.1	Proses Pengumpulan Data.....	26
3.2.2	Exploratory Data Analysis (EDA).....	28
3.2.3	Pre-processing Data .....	29
3.2.3.1	Penanganan Outlier.....	29
3.2.3.2	Scaling Data(normalisasi).....	31
3.2.4	Penerapan Model .....	32
3.2.4.1	PCA(Principal Component Analysis).....	32
3.2.4.2	Analisis Hierarchical Clustering .....	34
3.2.5	Hasil Model.....	36
3.2.6	Evaluasi Model .....	36
3.3	Alat dan Bahan.....	39
3.3.1.	Data Penelitian .....	39
3.3.2.	Software .....	39
3.3.3.	Library.....	39
3.3.4.	Hardware.....	39
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>41</b>
4.1	Exploratory Data Analysis (EDA).....	41
4.2	Pre-processing Data .....	43
4.2.1	Penanganan Outlier.....	43
4.2.1	Scaling Data (normalisasi).....	45
4.3	Penerapan Model .....	46
4.3.1	Principal Component Analysis (PCA).....	46
4.3.2	Hierarchical Clustering .....	48
4.3	Hasil Model .....	48

4.4 Evaluasi Model.....	51
<b>BAB V PENUTUP.....</b>	<b>53</b>
5.1 Kesimpulan.....	53
5.2 Saran.....	54
<b>REFERENSI.....</b>	<b>55</b>



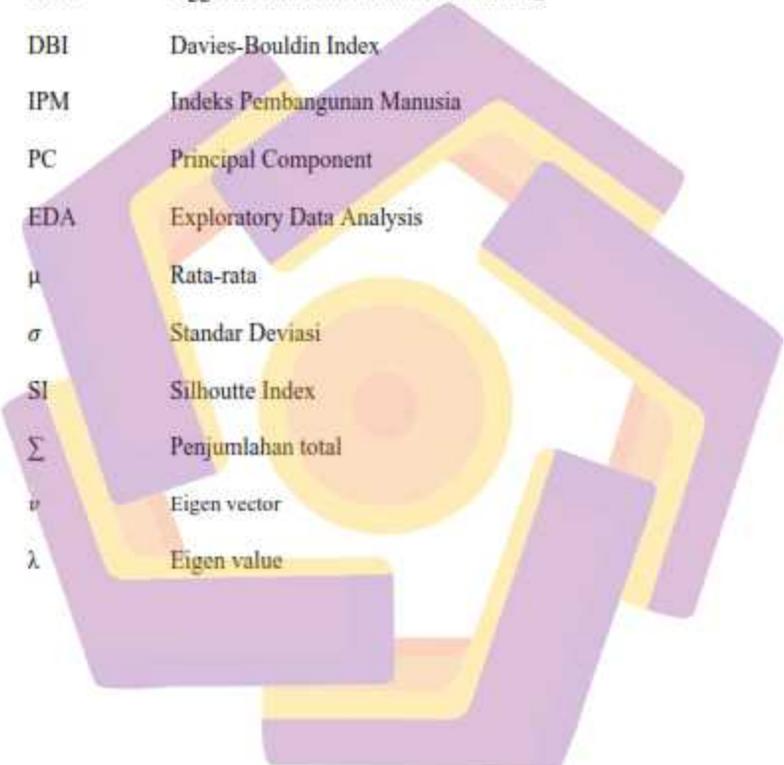
## DAFTAR TABEL

Tabel 1. 1 Kabupaten dan Kota di Provinsi Jawa Timur .....	1
Tabel 2. 1 Perbandingan Penelitian .....	10
Tabel 3. 1 Variabel Penelitian .....	27
Tabel 3. 2 Dataset .....	27
Tabel 3. 3 Outlier pada variabel <i>poorpeople_percentage</i> .....	31
Tabel 3. 4 Hasil <i>scaling</i> pada variabel <i>poorpeople_percentage</i> .....	32
Tabel 3. 5 Data yang telah di standarisasi .....	32
Tabel 3. 6 Hasil <i>Hierarchical Clustering</i> dari beberapa data .....	36
Tabel 4. 1 Hasil Statistik Deskriptif .....	41
Tabel 4. 2 Data setelah dilakukan <i>scaling</i> .....	46
Tabel 4. 3 Data frame hasil PCA .....	47
Tabel 4. 4 Hasil pengelompokkan dan karakteristik setiap cluster .....	49

## DAFTAR GAMBAR

Gambar 2. 1 Contoh Clustering .....	15
Gambar 2. 2 Contoh dendrogram (Mladenović, dkk, 2023).....	18
Gambar 2. 3 Model konseptual PCA untuk tahap seleksi fitur (Sumber.Kotu & Deshpande, 2015).....	21
Gambar 2. 4 Contoh Scree Plot (Halder, dkk. 2022).....	22
Gambar 3. 1 Alur Penelitian.....	26
Gambar 3. 2 Data sebelum dilakukan penanganan outlier.....	29
Gambar 4. 1 Perbandingan antar variabel sosial ekonomi.....	42
Gambar 4. 2 Heatmap korelasi antara berbagai variable .....	43
Gambar 4. 3 Data sebelum dilakukan penanganan outlier.....	44
Gambar 4. 4 Data setelah dilakukan penanganan outlier.....	45
Gambar 4. 5 Komponen PCA yang dihasilkan dari train model.....	46
Gambar 4. 6 Scree plot untuk menentukan komponen.....	47
Gambar 4. 7 Denogram untuk <i>Hierarchical Clustering</i> .....	48
Gambar 4. 8 Hasil pembagian cluster menggunakan <i>Hierarchical Clustering</i> .....	48
Gambar 4. 9 Rata-rata untuk setiap indikator pada masing-masing cluster.....	49
Gambar 4. 10 Visualisasi hasil klaster berdasarkan peta Jawa Timur.....	51
Gambar 4. 11 Hasil evaluasi model.....	52

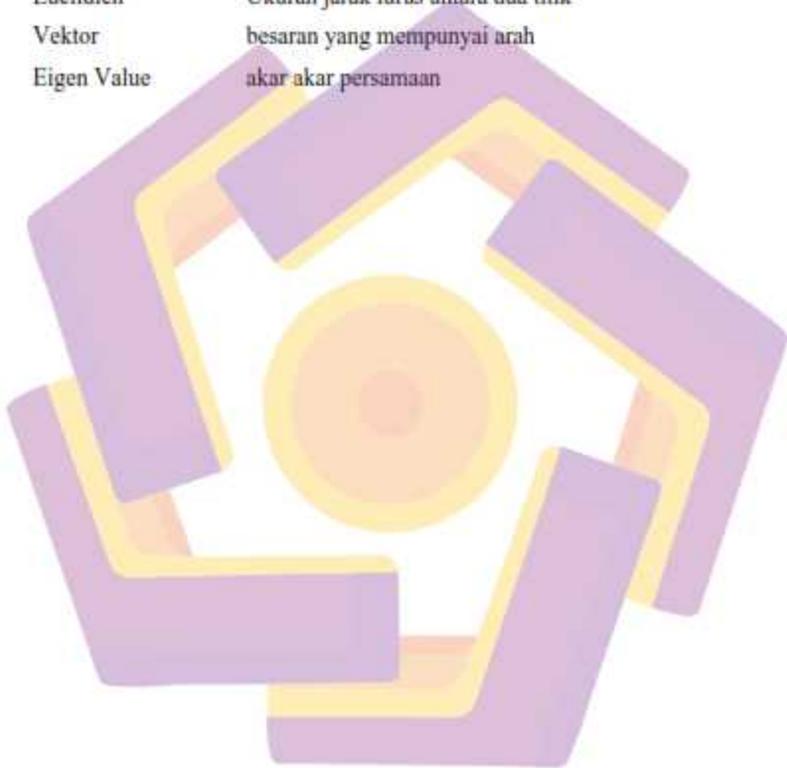
## DAFTAR LAMBANG DAN SINGKATAN



PCA	Principal Component Analysis
PDRB	Produk Domestik Regional Bruto
AHC	<i>Agglomerative Hierarchical Clustering</i>
DBI	Davies-Bouldin Index
IPM	Indeks Pembangunan Manusia
PC	Principal Component
EDA	Exploratory Data Analysis
$\mu$	Rata-rata
$\sigma$	Standar Deviasi
SI	Silhouette Index
$\Sigma$	Penjumlahan total
$v$	Eigen vector
$\lambda$	Eigen value

## DAFTAR ISTILAH

Persentil	Ukuran statistik untuk memahami distribusi data
Outlier	Nilai yang berbeda secara signifikan dari nilai-nilai lainnya
Mean	Rata-rata
Eucliden	Ukuran jarak lurus antara dua titik
Vektor	besaran yang mempunyai arah
Eigen Value	akar akar persamaan



## INTISARI

Penelitian ini bertujuan untuk menganalisis kondisi sosial-ekonomi Kabupaten/Kota di Provinsi Jawa Timur menggunakan metode Principal Component Analysis (PCA) dan *Hierarchical Clustering*. Jawa Timur, sebagai salah satu provinsi dengan jumlah penduduk terbesar di Indonesia, memiliki tantangan dalam pemerataan pembangunan dan kesejahteraan sosial-ekonomi. Penelitian ini menggunakan data sosial-ekonomi yang mencakup persentase penduduk miskin, GDP regional, harapan hidup, rata-rata lama sekolah, pengeluaran per kapita, dan tingkat pengangguran dari 38 Kabupaten/Kota di Jawa Timur pada tahun 2023.

Metode PCA digunakan untuk mereduksi dimensi data, sehingga memudahkan analisis dan visualisasi. Selanjutnya, data yang telah direduksi dianalisis menggunakan *Hierarchical Clustering* untuk mengelompokkan Kabupaten/Kota berdasarkan kesamaan karakteristik sosial-ekonomi. Hasil clustering dievaluasi menggunakan Silhouette Index dan Davies-Bouldin Index (DBI) untuk menilai kualitas pengelompokan.

Hasil penelitian menunjukkan bahwa metode clustering dapat mengidentifikasi empat cluster utama yang memiliki karakteristik sosial-ekonomi yang berbeda. Cluster dengan kondisi sosial-ekonomi terbaik memiliki GDP regional, harapan hidup, rata-rata lama sekolah, dan pengeluaran per kapita yang tinggi, serta tingkat pengangguran yang rendah. Sebaliknya, cluster dengan kondisi terburuk menunjukkan persentase penduduk miskin dan tingkat pengangguran yang tinggi. Dengan pemahaman yang lebih mendalam mengenai kondisi sosial-ekonomi di setiap cluster, pemerintah dan pemangku kepentingan dapat mengembangkan kebijakan yang lebih efektif dan tepat sasaran untuk meningkatkan kesejahteraan masyarakat di seluruh Kabupaten/Kota di Jawa Timur.

**Kata kunci:** PCA, *Hierarchical Clustering*, Sosial Ekonomi, Jawa Timur, Silhouette Index, Davies-Bouldin Index.

## ABSTRACT

*This study aims to analyze the socio-economic conditions of districts/cities in East Java Province using the Principal Component Analysis (PCA) and Hierarchical Clustering methods. East Java, as one of the provinces with the largest population in Indonesia, has challenges in equitable development and socio-economic welfare. This study uses socio-economic data that includes the percentage of poor people, regional GDP, life expectancy, average years of schooling, per capita expenditure, and unemployment rate from 38 districts/cities in East Java in 2023.*

*The PCA method is used to reduce the dimensions of the data, thus facilitating analysis and visualization. Furthermore, the reduced data was analyzed using Hierarchical Clustering to group districts/cities based on similarities in socio-economic characteristics. The clustering results were evaluated using the Silhouette Index and Davies-Bouldin Index (DBI) to assess the quality of the clustering.*

*The results show that the clustering method can identify four main clusters that have different socio-economic characteristics. Clusters with the best socio-economic conditions have high regional GDP, life expectancy, average years of schooling, and per capita expenditure, as well as low unemployment rates. In contrast, clusters with the worst conditions show a high percentage of poor people and a high unemployment rate. With a deeper understanding of the socio-economic conditions in each cluster, the government and stakeholders can develop more effective and targeted policies to improve the welfare of people in all districts/cities in East Java.*

**Keyword:** *PCA, Hierarchical Clustering, Socio-Economic, East Java, Silhouette Index, Davies-Bouldin Index.*