

**SEGMENTASI WILAYAH JAWA TIMUR BERDASARKAN  
INDEKS PEMBANGUNAN MANUSIA DENGAN  
MENGUNAKAN ANALISIS FAKTOR DAN *K-MEANS***

**SKRIPSI**

**Diajukan untuk memenuhi salah satu syarat kelulusan  
di Program Studi Sains Data**



**Disusun Oleh:**

**NINE ALVARIOATI VAROA ANSORI**

**20083010001**

**PROGRAM STUDI SAINS DATA  
FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR**

**SURABAYA**

**2024**

# LEMBAR PENGESAHAN

## SEGMENTASI WILAYAH JAWA TIMUR BERDASARKAN INDEKS PEMBANGUNAN MANUSIA DENGAN MENGGUNAKAN ANALISIS FAKTOR DAN K-MEANS

### SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data  
pada : Selasa, 14 Mei 2024

**Program Studi S-1 Sains Data**  
**Fakultas Ilmu Komputer**  
**Universitas Pembangunan Nasional Veteran Jawa Timur**  
**Surabaya**

Oleh :

**NINE ALVARIQATI VARQA ANSORI**

NPM. 20083010001

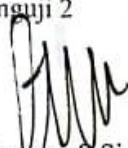
Disetujui oleh Tim Penguji Skripsi :

Penguji 1



Tresna Maulana Fahrudin, S.ST., MT  
NIP 199305012022031007

Penguji 2



Trimono, S.Si., M.Si  
NIP. 199509082022031003

Pembimbing 1



Amri Muhaimin, S.Stat., M.Stat., M.S.  
NIP 211199 50 723270

Pembimbing 2



Aviolla Ferza Damaliana, S.Si., M.Stat  
NIP 19940802 202203 2015

Fakultas Ilmu Komputer  
Dekan,

Prof. Dr. Ir. Novirina Hendrasarie, M.T.  
NIP. 19681126 199403 2 001

Mengetahui,

Koordinator Program Studi Sains Data  
Fakultas Ilmu Komputer

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU.  
NIP. 198012052005011002

Surabaya, Mei 2024

## SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Nine Alvariqati Varqa Ansori  
NPM : 20083010001  
Program Studi : Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

### **Segmentasi Wilayah Jawa Timur Berdasarkan Indeks Pembangunan Manusia Dengan Menggunakan Analisis Faktor dan *K-Means***

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.

Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 29 – Mei – 2024

Hormat Saya



Nine Alvariqati Varqa Ansori  
NPM. 20083010001



## ABSTRAK

### SEGMENTASI WILAYAH JAWA TIMUR BERDASARKAN INDEKS PEMBANGUNAN MANUSIA DENGAN MENGGUNAKAN ANALISIS FAKTOR DAN *K-MEANS*

Nama Mahasiswa / NPM : Nine Alvariqati Varqa Ansori / 20083010001  
Program Studi : Sains Data, FASILKOM, UPN Veteran Jatim  
Dosen Pembimbing 1 : Amri Muhaimin, S.Stat., M.Stat., M.S.  
Dosen Pembimbing 2 : Aviolla Terza Damaliana, S.Si., M.Stat

#### Abstrak

Indeks Pembangunan Manusia (IPM) merupakan upaya untuk mencapai Tujuan Pembangunan Berkelanjutan (TPB) dengan menggunakan komponen IPM seperti kesehatan, pendidikan, dan kehidupan yang layak. Menurut Badan Pusat Statistik (BPS), masih ada beberapa daerah di Provinsi Jawa Timur yang memiliki nilai IPM rendah. Berdasarkan hal tersebut, perlu dilakukan analisis dengan teknik klustering untuk mengetahui lebih jauh karakteristik masing-masing kabupaten dan kota di Jawa Timur. Data yang digunakan meliputi data Indeks Pembangunan Manusia (IPM), Umur Harapan Hidup (UHH), Rata-rata Lama Sekolah (RLS), Harapan Lama Sekolah (HLS), pengeluaran perkapita riil di sesuaikan, ketimpangan pendapatan, Indeks Pembangunan Gender (IPG), upah minimum kota, Produk Domestik Regional Bruto (PDRB), jumlah penduduk miskin, tingkat pengangguran terbuka, jumlah penduduk, dan tingkat partisipasi angkatan kerja. Menggunakan pemodelan *K-Means* yang dibandingkan dengan metode *Spectral* Klustering dan *DBSCAN* Hasil penelitian ini menunjukkan nilai *silhouette K-Means* sebesar 0,43 metode *spectral* klustering memiliki nilai *silhouette* sebesar 0,36 dan metode *DBSCAN* memiliki nilai *silhouette* sebesar 0,36. Model terbaik adalah *K-Means* klustering yang membentuk sebanyak 3 kluster optimum.

**Kata Kunci : Klustering, *K-Means*, Analisis Faktor, Indeks Pembangunan Manusia**

## **ABSTRACT**

### **SEGMENTATION OF EAST JAVA REGION BASED ON HUMAN DEVELOPMENT INDEX USING FACTOR ANALYSIS AND K-MEANS**

*Student Name / NPM* : Nine Alvariqati Varqa Ansori / 20083010001  
*Study Program* : Sains Data, FASILKOM, UPN Veteran Jatim  
*Advisor 1* : Amri Muhaimin, S.Stat., M.Stat., M.S.  
*Advisor 2* : Aviolla Terza Damaliana, S.Si., M.Stat

#### **Abstract**

*The Human Development Index (HDI) is an effort to achieve the Sustainable Development Goals (SDGs) by using HDI components such as health, education, and decent living. According to the Central Bureau of Statistics (BPS), there are still several regions in East Java Province that have low HDI values. Based on this, it is necessary to analyze with clustering techniques to find out more about the characteristics of each district and city in East Java. The data used includes data on Human Development Index (HDI), Life Expectancy (UHH), Average Years of Schooling (RLS), Expected Years of Schooling (HLS), adjusted real per capita expenditure, income inequality, Gender Development Index (HDI), city minimum wage, Gross Regional Domestic Product (GRDP), number of poor people, open unemployment rate, population, and labor force participation rate. Using K-Means modeling which is compared with the Spectral clustering and DBSCAN methods, the results of this study show that the K-Means silhouette score is 0.43, the spectral clustering method has a silhouette score of 0.36 and the DBSCAN method has a silhouette score of 0.36. The best model is K-Means clustering which forms as many as 3 optimum clusters.*

**Keyword:** *Klastering, K-Means, Factor Analysis, Human Developing Indeks*

## KATA PENGANTAR

Puji syukur kepada Tuhan Alam Semesta atas limpahan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Segmentasi Wilayah Jawa Timur Berdasarkan Indeks Pembangunan Manusia Dengan Menggunakan Analisis Faktor Dan *K-Means*” sesuai target yang diinginkan. Proposal skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Jurusan Sains Data. Pada kesempatan ini penulis mengucapkan banyak terima kasih yang sebesar-besarnya kepada

1. Allah SWT atas segala berkah, rezeki, dan juga cinta kasih yang selama ini diberikan kepada penulis sehingga proposal skripsi ini dapat terselesaikan;
2. Keluarga tercinta Abi, Mama, Kakak, dan Adik yang selalu menjadi support system dan senantiasa mendoakan agar penulis diberikan kemudahan dalam menyelesaikan penulisan proposal skripsi serta studi pada jenjang sarjana di Universitas Pembangunan Nasional “Veteran” Jawa Timur;
3. Bapak Amri Muhaimin, S.Stat., M.Stat., MS selaku dosen pembimbing I yang senantiasa memberikan wawasan baru akan topik yang penulis pilih serta bersabar dalam memberikan arahan, kritik, dan juga saran kepada penulis;
4. Ibu Aviolla Terza Damaliana, S.Si., M.Stat selaku dosen pembimbing II yang senantiasa bersabar dan supportive dalam menemani penulis untuk menyelesaikan proposal ini;
5. Sahabat dan teman-teman yang senantiasa memberi dorongan, doa, dan cinta kasihnya sehingga penulis menyusun proposal skripsi ini dengan Bahagia.

Penulis menyadari bahwa adanya keterbatasan wawasan, refrensi, serta pengalaman, sehingga kritik dan saran akan dibutuhkan di masa mendatang. Akhirnya besar harapan penulis agar proposal skripsi ini dapat memberikan manfaat bagi penulis dan pihak yang membutuhkan.

Surabaya, Mei - 2024

Penulis

## DAFTAR ISI

LEMBAR PENGESAHAN .....	<b>Error! Bookmark not defined.</b>
SURAT PERNYATAAN .....	<b>Error! Bookmark not defined.</b>
ABSTRAK .....	v
<i>ABSTRACT</i> .....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI .....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL .....	xi
DAFTAR LAMPIRAN .....	xii
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah .....	4
1.3. Batasan Masalah .....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA .....	6
2.1. Dasar Teori .....	6
2.1.1. Indeks Pembangunan Manusia (IPM) .....	6
2.1.2. Dimensi Pembentuk IPM .....	6
2.1.3. <i>Principal Component Analysis (PCA)</i> .....	8
2.1.4. Analisis Faktor.....	9
2.1.5. Klastering .....	12
2.1.6. <i>K-Means</i> Klastering.....	12
2.1.7. Nilai <i>silhouette</i> .....	13
2.2. Penelitian Terdahulu.....	13
BAB III METODOLOGI PENELITIAN .....	17
3.1. Variabel Penelitian dan Sumber Data.....	17
3.2. Langkah Analisis .....	17
3.3. Diagram Alir Penelitian.....	21
3.4. Jadwal Penelitian .....	21

BAB IV HASIL DAN PEMBAHASAN.....	22
4.1. <i>Preprocessing Data</i> .....	22
4.2. <i>Exploratory Data Analysis</i> .....	26
4.3. <i>Principal Component Analysis (PCA)</i> .....	28
4.4. Faktor Analisis.....	30
4.5. Klastering .....	33
4.6. Interpretasi Hasil <i>Klastering</i> .....	36
4.6. <i>Dashboard</i> .....	37
BAB V PENUTUP .....	39
5.1. Kesimpulan.....	39
5.2. Saran .....	40
DAFTAR PUSTAKA.....	41
LAMPIRAN .....	45
BIODATA PENULIS.....	51



## DAFTAR GAMBAR

Gambar 3. 1. Diagram Alur Penelitian .....	21
Gambar 4. 1. Hasil Pemeriksaan Data <i>Null</i> .....	22
Gambar 4. 2. Hasil Pemeriksaan Data Redundan.....	23
Gambar 4. 3. Hasil Pemeriksaan <i>Outlier</i> .....	23
Gambar 4. 4. Visualisasi <i>Heatmap</i> .....	27
Gambar 4. 5. <i>Code Proses PCA</i> .....	28
Gambar 4. 6. <i>Syntax</i> dan Hasil Nilai Eigen pada Tahap PCA.....	29
Gambar 4. 7. Hasil Uji Korelasi pada Proses Analisis Faktor.....	30
Gambar 4. 8. <i>Syntax</i> dan Hasil Uji Barlettes .....	31
Gambar 4. 9. Hasil Uji KMO .....	31
Gambar 4. 10. Grafik Banyaknya Klaster pada <i>K-Means</i> yang Dapat Terbentuk Berdasarkan Nilai Silhouette .....	34
Gambar 4. 11. Grafik Banyaknya Klaster pada <i>Spectral</i> Klastering yang Dapat Terbentuk Berdasarkan Nilai Silhouette .....	34
Gambar 4. 12. Grafik Banyaknya Klaster pada <i>Spectral</i> Klastering yang Dapat Terbentuk Berdasarkan Nilai Silhouette .....	35
Gambar 4. 13. Visualisasi <i>K-Means</i> Klastering .....	36
Gambar 4. 14. <i>Boxplot</i> klaster pada masing masing PC.....	37
Gambar 4. 15. <i>Dashboard</i> perkembangan IPM.....	38

## **DAFTAR TABEL**

Tabel 2. 1. Tabel Penelitian Terdahulu.....	14
Tabel 3. 1. Tabel Variabel Penelitian .....	17
Tabel 3. 2. Tabel Jadwal Penelitian.....	21
Tabel 4. 1. Hasil Uji IQR.....	24
Tabel 4. 2. Tabel Statistika Deskriptif.....	26
Tabel 4. 3. Tabel Hasil PCA.....	30
Tabel 4. 4. Tabel Hasil Analisis Faktor .....	32
Tabel 4. 5. Hasil Nilai Komunaliti .....	33

## DAFTAR LAMPIRAN

Lampiran 1. Hasil uji plagiasi .....	45
Lampiran 2. Data penelitian .....	46
Lampiran 3. Lanjutan Data Penelitian .....	47
Lampiran 4. <i>Source Code</i> yang digunakan untuk analisis.....	48
Lampiran 5. Hasil <i>Dashboarding</i> .....	49
Lampiran 6. Visualisasi hasil klastering DBSCAN dan <i>Spectral</i> Klastering.....	50