

**PENERAPAN ALGORITMA K-MEANS UNTUK  
SEGMENTASI DAERAH DI JAWA TIMUR BERDASARKAN  
INDIKATOR KESEJAHTERAAN MASYARAKAT**

**SKRIPSI**

Diajukan untuk memenuhi persyaratan  
dalam memperoleh gelar Sarjana Komputer  
Program Studi Sistem Informasi



**Disusun Oleh:**

**M.KANDIAS HAPPY MAULANA**

**19082010100**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
SURABAYA  
2024**

**SKRIPSI**


**PENERAPAN ALGORITMA K-MEANS UNTUK SEGMENTASI  
DAERAH DI JAWA TIMUR BERDASARKAN INDIKATOR  
KESEJAHTERAAN MASYARAKAT**

**Disusun Oleh:**  
**M.KANDIAS HAPPY MAULANA**  
**19082010100**

**Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Skripsi  
Program Studi Sistem Informasi Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional "Veteran" Jawa Timur  
pada Tanggal 12 Juli 2024**

**Pembimbing :**

1.

  
**Nur Cahyo Wibowo, S.Kom, M.Kom.**  
**NIP. 19790317 2021211 002**

2.

  
**Abdul Rezha Efrat Najaf, S.Kom, M.Kom.**  
**NIP. 19940929 2022031 008**

**Tim Penguji :**


1.

  
**Asif Faroqi, S.Kom., M.Kom.**  
**NIP. 19870519 2018031 001**

2.

  
**Dhian Satria Yudha Kartika, S.Kom, M.Kom.**  
**NPT. 2 0119 86 052224 9**

3.

  
**Seftin Fitri Ana Wati, S.kom., M.Kom.**  
**NPT. 2 1219 91 032026 7**

**Mengetahui,**

**Dekan Fakultas Ilmu Komputer  
Universitas Pembangunan Nasional "Veteran" Jawa Timur**

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

**LEMBAR PENGESAHAN**

**PENERAPAN ALGORITMA K-MEANS UNTUK SEGMENTASI  
DAERAH DI JAWA TIMUR BERDASARKAN INDIKATOR  
KESEJAHTERAAN MASYARAKAT**


**Disusun Oleh:**  
**M.KANDIAS HAPPY MAULANA**  
**19082010100**

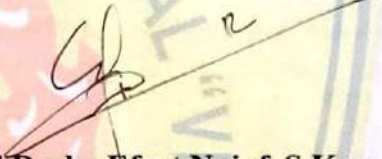
**Telah disetujui mengikuti Ujian Negara Lisan Gelombang Juli  
Periode 2024 pada Tanggal 12 Juli 2024**

**Menyetujui,**


**Dosen Pembimbing 1**

**Dosen Pembimbing 2**

  
**Nur Cahyo Wibowo, S.Kom, M.Kom.**  
**NIP. 19790317 2021211 002**

  
**Abdul Rezha Efrat Najaf, S.Kom, M.Kom.**  
**NIP. 19940929 2022031 008**

**Mengetahui,**  
**Ketua Program Studi Sistem Informasi**  
**Fakultas Ilmu Komputer**

  
**Agung Brastama Putra, S.Kom., M.Kom.**  
**NIP. 19851124 2021211 003**



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN**  
**UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"**  
**JAWA TIMUR**  
**FAKULTAS ILMU KOMPUTER**

**KETERANGAN REVISI**

Kami yang bertanda tangan di bawah ini menyatakan bahwa mahasiswa berikut:

Nama : M.Kandias Happy Maulana

NPM : 19082010100

Program Studi : Sistem Informasi

Telah mengerjakan revisi Ujian Negara Lisan Skripsi pada tanggal 23 Juli 2024 dengan judul:

**"PENERAPAN ALGORITMA K-MEANS UNTUK SEGMENTASI  
DAERAH DI JAWA TIMUR BERDASARKAN INDIKATOR  
KESEJAHTERAAN MASYARAKAT"**

Oleh karenanya mahasiswa tersebut di atas dinyatakan bebas revisi Ujian Negara Lisan Skripsi dan diijinkan untuk membukukan laporan SKRIPSI dengan judul tersebut.

Surabaya, 23 Juli 2024

Dosen penguji yang memeriksa skripsi:

1. Asif Faroqi, S.Kom., M.Kom.  
NIP. 19870519 2018031 001


2. Dhian Satria Yudha Kartika, S.Kom, M.Kom.  
NPT. 2 0119 86 052224 9

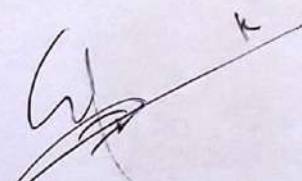
3. Seftin Fitri Ana Wati, S.kom., M.Kom.  
NPT. 2 1219 91 032026 7

Mengetahui,

Dosen Pembimbing 1

Dosen Pembimbing 2

  
Nur Cahyo Wibowo, S.Kom, M.Kom.  
NIP. 19790317 2021211 002

  
Abdul Rezha Efrat Najaf, S.Kom, M.Kom.  
NIP. 19940929 2022031 008



**KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN  
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"  
JAWA TIMUR  
FAKULTAS ILMU KOMPUTER**

**SURAT PERNYATAAN**

Saya yang bertanda tangan di bawah ini:

Nama : M.Kandias Happy Maulana

NPM : 19082010100

Program Studi : Sistem Informasi

Menyatakan bahwa Judul Skripsi / Tugas Akhir sebagai berikut:

**"PENERAPAN ALGORITMA K-MEANS UNTUK SEGMENTASI  
DAERAH DI JAWA TIMUR BERDASARKAN INDIKATOR  
KESEJAHTERAAN MASYARAKAT"**

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir / Penelitian orang lain dan juga bukan merupakan Produk / 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 terbukti benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah di kemudian hari.

Surabaya, 23 Juli 2024

Hormat Saya,



**M.Kandias Happy Maulana**

**NPM. 19082010100**

Judul : PENERAPAN ALGORITMA K-MEANS UNTUK  
SEGMENTASI DAERAH DI JAWA TIMUR  
BERDASARKAN INDIKATOR KESEJAHTERAAN  
MASYARAKAT

Pembimbing 1 : Nur Cahyo Wibowo, S.Kom, M.Kom

Pembimbing 2 : Abdul Rezha Efrat Najaf, S.Kom, M.Kom

---

### **ABSTRAK**

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 .

#### **Kata kunci:**

Klustering, Kesejahteraan, *Principal Component Analysis*, K-Means, Segmentasi

## **KATA PENGANTAR**

Segala puji syukur penulis panjatkan atas kehadiran Allah SWT karena atas berkat dan hidayahNya penulis dapat menyelesaikan skripsi yang berjudul “PENERAPAN ALGORITMA K-MEANS UNTUK SEGMENTASI DAERAH DI JAWA TIMUR BERDASARKAN INDIKATOR KESEJAHTERAAN MASYARAKAT” sebagai salah satu syarat untuk menyelesaikan program studi Sistem Informasi Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Dalam proses menyelesaikan skripsi ini, tidak lepas dari dukungan, dorongan, dan bantuan dari berbagai pihak. Oleh karena itu penulis menyampaikan terima kasih kepada:

1. Kedua orang tua, Bapak Djoko Walujo, S.T dan Ibu Zumaroh, S.Sos yang senantiasa memberikan doa dan dukungan dalam segala hal, tidak terkecuali skripsi.
2. Kakek-nenek dan saudara yang senantiasa memberikan motivasi serta doa untuk kelancaran skripsi ini.
3. Ibu Amalia Anjani Arifiyanti, S.Kom, M.Kom selaku promotor seminar proposal yang telah memberikan arahan dan masukan, dalam hal teknis serta penulisan.
4. Bapak Nur Cahyo Wibowo, S.Kom, M.Kom selaku pembimbing I yang telah memberikan arahan dan masukan, dalam hal teknis serta penulisan.
5. Bapak Abdul Rezha Efrat Najaf, S.Kom, M.Kom selaku pembimbing II yang telah memberikan arahan dan masukan, dalam hal teknis serta penulisan.

6. Bapak Agung Brastama Putra, S.Kom, M.Kom selaku dosen wali yang selalu memberikan motivasi dan arahan selama periode studi.
7. Seluruh dosen Sistem Informasi yang telah banyak memberikan ilmu yang sangat bermanfaat selama kuliah sebagai bekal kedepannya.
8. Seluruh teman-teman Sistem Informasi angkatan 2019 yang telah memberi semangat untuk menyelesaikan skripsi ini, terutama untuk teman-teman kelas A.
9. Seluruh teman ngopi dan bercanda tawa yang telah memberi semangat untuk menyelesaikan skripsi ini yaitu Muhammad Daffa, Suhri Ainur Rifky, Brian Firmansyah Kartono Soebari, Catur Arpal Perkasa, Talenta Eklesia Noor, dan kawan-kawan lainnya.
10. Seluruh bolo-bolo Warkop Teras 28 yang telah memberi semangat serta menghibur dikala penat untuk menyelesaikan skripsi ini.
11. Seluruh pihak yang tidak bisa saya sebutkan satu persatu karena telah memberikan bantuan bahkan sekecil apapun.

Semoga Allah SWT membalas kebaikan pihak-pihak diatas. Dalam pengerjaan skripsi ini penulis menyadari masih banyak sekali kekurangan. Semoga skripsi ini bermanfaat bagi semua pihak.



## DAFTAR ISI

ABSTRAK .....	i
KATA PENGANTAR .....	ii
DAFTAR ISI.....	iv
DAFTAR TABEL.....	vi
DAFTAR GAMBAR .....	vii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah .....	4
1.3 Batasan Masalah .....	4
1.4 Tujuan Penelitian .....	5
1.5 Sistematika Penulisan .....	5
BAB II TINJAUAN PUSTAKA.....	8
2.1 Dasar Teori.....	8
2.1.1 Badan Pusat Statistik (BPS).....	8
2.1.2 <i>Cross Industry Standard Process for Data Mining (CRISP-DM)</i> .....	8
2.1.3 Segmentasi .....	9
2.1.4 <i>Principal Component Analysis (PCA)</i> .....	9
2.1.5 <i>Clustering</i> .....	9
2.1.6 <i>K-Means</i> .....	10
2.1.7 Metode Elbow .....	11
2.1.8 <i>Silhouette Coefficient</i> .....	12
2.1.9 Python .....	12
2.1.10 Pandas .....	13
2.1.11 Geopandas.....	13
2.1.12 Folium .....	14
2.1.13 Sklearn .....	14
2.2 Penelitian Terdahulu .....	15
BAB III METODOLOGI PENELITIAN.....	18
3.1 Analisis Kebutuhan .....	18
3.1.1 Kebutuhan Data.....	18
3.1.2 Kebutuhan Software dan Hardware .....	19
3.2 Cross Industry Standard Process for Data Mining (CRISP-DM) .....	19

BAB IV HASIL DAN PEMBAHASAN .....	24
4.1 Implementasi Kebutuhan .....	24
4.1.1 Kebutuhan Data.....	24
4.2 Import Data .....	25
4.3 Klasterisasi .....	26
4.3.1 <i>Data Preparation</i> .....	26
4.3.2 <i>Modeling</i> .....	35
4.3.3 <i>Evaluation</i> .....	38
BAB V PENUTUP.....	55
5.1 Kesimpulan .....	55
5.2 Saran .....	56
DAFTAR PUSTAKA .....	57
LAMPIRAN.....	65

## DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu .....	15
Tabel 4. 1 Data Indikator Kesejahteraan Masyarakat .....	24
Tabel 4. 2 Identifikasi Karakteristik <i>Cluster</i> Tahun 2020.....	50
Tabel 4. 3 Identifikasi Karakteristik <i>Cluster</i> Tahun 2021.....	51
Tabel 4. 4 Identifikasi Karakteristik <i>Cluster</i> Tahun 2022.....	52

## DAFTAR GAMBAR

Gambar 3. 1 Alur Metode Penelitian .....	18
Gambar 3. 2 Model Analitik CRISP-DM .....	20
Gambar 4. 1 Website BPS.....	25
Gambar 4. 2 Contoh Data yang Akan Diolah .....	26
Gambar 4. 3 <i>Sourcecode</i> utnuk standarisasi data df .....	26
Gambar 4. 4 <i>Load Library</i> .....	26
Gambar 4. 5 Import Dataset .....	27
Gambar 4. 6 <i>Sourcecode</i> untuk menerapkan PCA.....	27
Gambar 4. 7 Hasil Penerapan PCA 2020 .....	28
Gambar 4. 8 Hasil Penerapan PCA 2021 .....	28
Gambar 4. 9 Hasil Penerapan PCA 2022 .....	29
Gambar 4. 10 Penetapan Komponen PCA .....	29
Gambar 4. 11 Penerapan PCA dengan Jumlah Komponen Optimal .....	29
Gambar 4. 12 Source Code Untuk Ekstrak Beban PCA .....	30
Gambar 4. 13 Hasil Ekstrak Beban PCA 2020 .....	30
Gambar 4. 14 Hasil Ekstrak Beban PCA 2021 .....	30
Gambar 4. 15 Hasil Ekstrak Beban PCA 2022 .....	30
Gambar 4. 16 Source Code untuk Membuat data frame Ekstrak Beban PCA.....	31
Gambar 4. 17 Data Frame Ekstrak Beban PCA 2020 .....	31
Gambar 4. 18 Data Frame Ekstrak Beban PCA 2021 .....	31
Gambar 4. 19 Data Frame Ekstrak Beban PCA 2022 .....	31
Gambar 4. 20 <i>Source Code</i> untuk heatmap PCA Loadings .....	31
Gambar 4. 21 Heatmap PCA Loadings 2020.....	32
Gambar 4. 22 Heatmap PCA Loadings 2021 .....	32
Gambar 4. 23 Heatmap PCA Loadings 2022.....	33
Gambar 4. 24 <i>Source Code</i> untuk Plot 2 Dimensi .....	33
Gambar 4. 25 Hasil Visualisasi 2 Dimensi 2020 .....	34
Gambar 4. 26 Hasil Visualisasi 2 Dimensi 2021 .....	34
Gambar 4. 27 Hasil Visualisasi 2 Dimensi 2022 .....	35
Gambar 4. 28 <i>Source Code</i> untuk Menentukan jumlah K Optimal .....	36
Gambar 4. 29 Hasil Visualisasi Metode <i>Elbow</i> 2020 .....	36

Gambar 4. 30 Hasil Visualisasi Metode <i>Elbow</i> 2021 .....	37
Gambar 4. 31 Hasil Visualisasi Metode <i>Elbow</i> 2022 .....	37
Gambar 4. 32 <i>Source code</i> untuk Perhitungan <i>Silhouette Coefficient</i> .....	38
Gambar 4. 33 Hasil Perhitungan <i>Silhouette Coefficient</i> 2020.....	38
Gambar 4. 34 Hasil Perhitungan <i>Silhouette Coefficient</i> 2021.....	39
Gambar 4. 35 Hasil Perhitungan <i>Silhouette Coefficient</i> 2022.....	39
Gambar 4. 36 <i>Source Code</i> untuk penerapan K-Means 2020.....	40
Gambar 4. 37 <i>Source Code</i> untuk penerapan K-Means 2021 .....	40
Gambar 4. 38 <i>Source Code</i> untuk penerapan K-Means 2022.....	40
Gambar 4. 39 <i>Source Code</i> Untuk Plotting <i>Cluster</i> .....	41
Gambar 4. 40 Plot Hasil <i>Cluster</i> 2020 .....	42
Gambar 4. 41 Plot Hasil <i>Cluster</i> 2021 .....	42
Gambar 4. 42 Plot Hasil <i>Cluster</i> 2022 .....	43
Gambar 4. 43 <i>Source Code</i> untuk Menambahkan kolom <i>cluster</i> .....	43
Gambar 4. 44 <i>DataFrame</i> dengan Kolom <i>cluster</i> 2020.....	43
Gambar 4. 45 <i>DataFrame</i> dengan Kolom <i>cluster</i> 2021.....	44
Gambar 4. 46 <i>DataFrame</i> dengan Kolom <i>cluster</i> 2022.....	44
Gambar 4. 47 <i>Source Code</i> untuk Import File geojson Titik Koordinat Daerah ..	44
Gambar 4. 48 <i>Source Code</i> untuk <i>replace</i> kolom dari file gojson.....	44
Gambar 4. 49 <i>DataFrame</i> dari gojson Setelah di <i>Replace</i> .....	45
Gambar 4. 50 <i>Source Code</i> untuk Menggabung Kolom 'kabkot' dalam df_geo..	45
Gambar 4. 51 df_cluster 2020.....	45
Gambar 4. 52 df_cluster 2021 .....	46
Gambar 4. 53 df_cluster 2022.....	46
Gambar 4. 54 <i>Source Code</i> untuk Menggabungkan df_geo dan df_cluster .....	46
Gambar 4. 55 Hasil Penggabungan df_geo dan df_cluster 2020.....	47
Gambar 4. 56 Hasil Penggabungan df_geo dan df_cluster 2021 .....	47
Gambar 4. 57 Hasil Penggabungan df_geo dan df_cluster 2022.....	47
Gambar 4. 58 Import <i>Library</i> folium .....	48
Gambar 4. 59 <i>Source Code</i> untuk visualisasi hasil <i>cluster</i> .....	48
Gambar 4. 60 Hasil Visualisasi 2020.....	49
Gambar 4. 61 Hasil Visualisasi 2021 .....	49

Gambar 4. 62 Hasil Visualisasi 2022 .....	50
Gambar 4. 63 Hasil Visualisasi <i>Website</i> Halaman Utama .....	54
Gambar 4. 64 Hasil Visualisasi <i>Website</i> Halaman <i>Cluster</i> Jawa Timur .....	54