

**SEGMENTASI PENDUDUK MISKIN DI INDONESIA
MENGUNAKAN ALGORITMA *K-MEANS***

SKRIPSI

**Untuk Memenuhi Persyaratan
Memperoleh Gelar Sarjana**

Oleh:

**Agita Vidiasti Rivallinata
NIM 19.51.0013**



PROGRAM STUDI S-1 SISTEM INFORMASI

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
PPKIA PRADNYA PARAMITA
MALANG
2023**

LEMBAR PERSETUJUAN

**SEGMENTASI PENDUDUK MISKIN DI INDONESIA
MENGUNAKAN ALGORITMA *K-MEANS***

Oleh:
Agita Vidiasti Rivallinata
NIM 19.51.0013

**Telah disetujui oleh pembimbing untuk diujikan
Pada tanggal 9 Mei 2023**

Menyetujui,
Komisi Pembimbing

Pembimbing I



Dr. Tb. Mohammad Akhriza, S.Si., MMSI., Ph.D
NIK 00.12.01.002

Pembimbing II



Dr. Dwi Safiroh Utsalina, S.Kom., MMSI
NIK 00.12.01.016

Mengetahui,

Ketua Program Studi S-1 Sistem Informasi



Dr. Dwi Safiroh Utsalina, S.Kom., MMSI
NIK 00.12.01.016

LEMBAR PENGESAHAN

SEGMENTASI PENDUDUK MISKIN DI INDONESIA
MENGUNAKAN ALGORITMA *K-MEANS*

Skripsi oleh Agita Vidiasti Rivallinata
Telah dipertahankan di depan dewan penguji
Pada tanggal 19 Mei 2023

Dewan Penguji,
Ketua



Dr. Rahayu Widayanti, S.E., M.M., MMSI
NIK 00.12.01.006

Anggota I



Dr. Tb. Mohammad Akhriza, S.Si., MMSI., Ph.D
NIK 00.12.01.002

Anggota II



Dr. Mochamad Husni, S.PdI., M.M., MMSI
NIK 01.09.10.036

Mengetahui,
Ketua Program Studi
S-1 Sistem Informasi



Dr. Dwi Saifurrah Utsalina, S.Kom., MMSI
NIK 00.12.01.016

Mengesahkan,
Ketua STMIK PPKIA
Pradnya Paramita



Dr. Tb. Mohammad Akhriza, S.Si., MMSI., Ph.D
NIK 00.12.01.002

PERNYATAAN ORISINALITAS SKRIPSI

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam Naskah SKRIPSI ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu Perguruan Tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis dikutip dalam naskah ini disebutkan dalam sumber kutipan dan daftar pustaka.

Apabila ternyata di dalam naskah SKRIPSI ini dapat dibuktikan terdapat unsur-unsur PLAGIASI, saya bersedia SKRIPSI ini digugurkan dan gelar akademik yang saya peroleh (SARJANA) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No 20 Tahun 2003, pasal 25 ayat 2 dan pasal 70).

Malang, 29 Maret 2023

Mahasiswa,

Materai 10000

Nama : Agita Vidiasti R.

Nim : 19.51.0013

ABSTRAK

Rivallinata, Agita Vidiasti. 2023. *Segmentasi Penduduk Miskin di Indonesia Menggunakan Algoritma K-Means*. Skripsi, Program Studi Sistem Informasi STMIK PPKIA Pradnya Paramita. Pembimbing: (I) Dr. Tb. Mohammad Akhriza, S.Si., MMSI., Ph.D, (II) Dr. Dwi Safiroh Utsalina, S.Kom., MMSI

Kata Kunci: *K-Means, Clustering, Silhoutte Coefficient*

Dalam konteks pandemi Covid-19, kebijakan pemerintah untuk menekan penyebaran virus telah berdampak pada pertumbuhan perekonomian dan meningkatnya jumlah penduduk miskin di Indonesia. Pengelompokan tingkat kemiskinan wilayah-wilayah di Indonesia menjadi penting sebagai informasi yang dapat digunakan untuk menentukan kebijakan dan penyaluran bantuan yang tepat. Tujuan dilakukannya penelitian ini yaitu untuk melakukan pengelompokan pada data kemiskinan di Indonesia berdasarkan atribut GKM, GKNM, IkdK dan IKpK. Oleh karena itu, dibutuhkan teknik data mining yang dapat mengelompokkan tingkat kemiskinan wilayah-wilayah di Indonesia dengan algoritma *K-Means*. Hasil penelitian menunjukkan adanya 3 *cluster* kemiskinan, yaitu kemiskinan tingkat rendah, sedang, dan tinggi. *Clustering* yang dilakukan pada data sebelum, selama, dan setelah pandemi tanpa menggunakan teknik pengelompokan *binning* tidak menghasilkan pergeseran *cluster*, setelah menggunakan *binning* terjadi pergeseran *cluster* di beberapa provinsi pada tahun 2020 dan 2022. Hasil *clustering* pada data puncak pandemi menuju kehidupan kenormalan tanpa menggunakan *binning* menghasilkan pergeseran dari tinggi menuju rendah, yakni pada provinsi Maluku dan Nusa Tenggara Timur, setelah menggunakan *binning* mengalami pergeseran dari tingkat kemiskinan tinggi ke rendah terjadi pada provinsi Nusa Tenggara Timur dan dari tingkat kemiskinan sedang ke rendah terjadi pada provinsi Bengkulu. Hal ini menunjukkan bahwa *binning* data dapat memberikan informasi yang lebih detail dan membantu dalam memahami perubahan tingkat kemiskinan. Penggunaan *Silhoutte Coefficient* sebagai metrik evaluasi *clustering* menunjukkan nilai antara 0,54-0,59, yang mengindikasikan *clustering* yang terbentuk memiliki interpretasi yang baik dan mendekati nilai 1.

KATA PENGANTAR

Puji syukur kami panjatkan kehadiran Tuhan Yang Maha Esa atas segala rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi dengan judul **“SEGMENTASI PENDUDUK MISKIN DI INDONESIA MENGGUNAKAN ALGORITMA *K-MEANS*”** dengan baik sebagai syarat untuk memperoleh gelar Sarjana Komputer.

Penulis menyadari tanpa adanya dukungan dan kerja sama dari berbagai pihak, kegiatan skripsi ini tidak akan dapat berjalan baik. Untuk itu, kami ingin menyampaikan rasa terima kasih kepada:

1. Tuhan Yang Maha Esa yang telah memberikan petunjuk dan hidayah dalam pembuatan skripsi ini.
2. Bapak Dr. Tb. Mohammad Akhriza, S.Si., MMSI., Ph.D, selaku Ketua STMIK PPKIA Pradnya Paramita Malang, juga selaku dosen pembimbing I penyusunan skripsi ini atas segala bimbingan, arahan, serta saran yang diberikan kepada penulis sehingga penyusunan ini dapat diselesaikan dengan baik.
3. Ibu Dr. Dwi Safiroh Utsalina, S.Kom., MMSI, selaku Ketua Program Studi S-1 Sistem Informasi, juga selaku dosen pembimbing II penyusunan skripsi ini atas segala bimbingan, arahan, serta saran yang diberikan kepada penulis sehingga penyusunan ini dapat diselesaikan dengan baik.
4. Bapak Dr. Mochamad Husni S.PdI., MM., MMSI, selaku Dosen Pendamping Akademik yang selalu mengingatkan untuk berprogress sekaligus membimbing dan memberi saran sehingga skripsi ini cepat terselesaikan.
5. Seluruh jajaran Dosen dan Staff STMIK PPKIA Pradnya Paramita Malang dalam membimbing, serta memberikan ilmu yang bermanfaat.
6. Orang tua dan teman-teman yang telah memberikan doa, semangat dan dukungan.
7. Semua pihak yang telah membantu dan mendukung penyelesaian skripsi ini yang tidak dapat disebutkan satu persatu.

Penulis menyadari atas ketidaksempurnaan penyusunan Skripsi ini, penulis mengharapkan adanya kritik dan saran yang membangun demi kemajuan penulis. Semoga skripsi ini berguna bagi pembaca secara umum dan penulis secara khusus. Akhir kata, penulis ucapkan banyak terima kasih.

Malang, 29 Maret 2023

Penulis,

Agita Vidiasti Rivallinata

DAFTAR ISI

SKRIPSI.....	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
ABSTRAK.....	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	4
1.5 Kontribusi Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Data Mining	5
2.2 Metode Clustering	7
2.3 Algoritma K-Means dan Metode Elbow	7
2.4 Penerapan K-Means untuk Segmentasi Penduduk Miskin di Indonesia.	11
2.5 Evaluasi Clustering.....	13
BAB III METODOLOGI PENELITIAN	15
3.1 Analisis Permasalahan	15
3.2 Solusi yang Diusulkan	15
3.2.1 Seleksi Fitur	16
3.2.2 Pre-processing	17
3.2.3 Menentukan Jumlah Cluster Optimal	19
3.2.4 Binning Data.....	19
3.2.5 Clustering Data Hasil Binning dengan Nilai k yang Diperoleh	19

3.2.6	<i>Clustering</i> dengan Nilai k yang Diperoleh.....	20
3.2.7	Evaluasi dan Hasil Analisis <i>Cluster</i>	20
3.3	Rancangan Eksperimen.....	21
3.3.1	Kebutuhan Perangkat	21
3.3.2	Kebutuhan Data.....	22
BAB IV PENGUJIAN DAN HASIL		24
4.1	Penerapan Metode.....	24
4.1.1	Seleksi Fitur	24
4.1.2	<i>Pre-Processing</i>	24
4.1.3	Menentukan Jumlah <i>Cluster</i> Optimal	36
4.1.4	Melakukan <i>Binning</i> Data.....	41
4.1.5	<i>Clustering</i> Data Hasil <i>Binning</i> dengan Nilai k yang Diperoleh	43
4.1.6	<i>Clustering</i> dengan Nilai k yang Diperoleh.....	48
4.2	Evaluasi dan Hasil Analisis <i>Cluster</i>	54
4.2.1	Evaluasi <i>Cluster</i>	54
4.2.2	Hasil Analisis <i>Cluster</i>	55
BAB V KESIMPULAN DAN SARAN.....		62
5.1	Kesimpulan	62
5.2	Saran	63
DAFTAR PUSTAKA		64

DAFTAR TABEL

No	Nama Tabel	Halaman
2. 1	Pengukuran Koefisien <i>Silhoutte</i>	13
3. 1	<i>Sample</i> Data Mentah GKM.....	22
3. 2	<i>Sample</i> Data Mentah GKNM.....	23
3. 3	<i>Sample</i> data Mentah IkdK.....	23
3. 4	<i>Sample</i> Data Mentah IkpK.....	23
4. 1	Data Tahun 2019.....	25
4. 2	Data Tahun 2020.....	26
4. 3	Data Tahun 2021.....	28
4. 4	Data Tahun 2022.....	30
4. 5	Hasil dari <i>Silhoutte_score</i>	39
4. 6	Hasil <i>Clustering</i> Data Tahun 2019.....	55
4. 7	Hasil <i>Clustering</i> Data Tahun 2020.....	56
4. 8	Hasil <i>Clustering</i> Data Tahun 2021.....	57
4. 9	Hasil <i>Clustering</i> Data Tahun 2022.....	57
4. 10	Perbedaan Hasil <i>Clustering</i> Tanpa <i>Binning</i> dan <i>Binning</i>	59
4. 11	Pergeseran <i>Cluster</i> Data Tahun 2019, 2020, 2022.....	60
4. 12	Pergeseran <i>Cluster</i> Data Tahun 2021 & 2022.....	61
4. 13	Pergeseran <i>Cluster</i> Data <i>Binning</i> Tahun 2021 & 2022.....	61

DAFTAR GAMBAR

No	Nama Gambar	Halaman
2. 1	KDD	6
2. 2	<i>Flowchart</i> Algoritma <i>K-Means</i>	8
3. 1	<i>Framework</i> Penelitian	16
4. 1	<i>Source Code</i> Mengimport <i>Library</i>	32
4. 2	<i>Source Code</i> Mengimport dan Memanggil <i>Dataset</i>	32
4. 3	Hasil Tampilan <i>Sample Dataset</i>	32
4. 4	<i>Source Code</i> Perintah <i>iloc</i>	33
4. 5	Hasil perintah <i>iloc</i> pada data tahun 2019	33
4. 6	<i>Source Code</i> <i>Min-max Normalization</i> dengan data tahun 2019	34
4. 7	Hasil <i>Min-max Normalization</i> Data tahun 2019	34
4. 8	Tampilan <i>Dataframe</i> Baru Hasil Normalisasi Data Tahun 2019	35
4. 9	<i>Source Code</i> Menyimpan <i>File</i> dalam Format <i>csv</i>	35
4. 10	<i>Sample</i> Data Hasil Normalisasi Tahun 2019	35
4. 11	<i>Sample</i> Data Hasil Normalisasi Tahun 2020	35
4. 12	<i>Sample</i> Data Hasil Normalisasi Tahun 2021	36
4. 13	<i>Sample</i> Data Hasil Normalisasi Tahun 2022	36
4. 14	<i>Source Code</i> Menginstall <i>library scikit-learn</i>	36
4. 15	<i>Source Code</i> Mengimport <i>iloc</i> Data Tahun 2019	37
4. 16	Hasil Perintah <i>iloc</i> pada Data Hasil Normalisasi Tahun 2019	37
4. 17	<i>Source Code</i> Menentukan Jumlah <i>Cluster</i> Optimal	38
4. 18	Visualisasi <i>Elbow</i> Data Tahun 2019	39
4. 19	Visualisasi <i>Elbow</i> Data Tahun 2020	40
4. 20	Visualisasi <i>Elbow</i> Data Tahun 2021	40
4. 21	Visualisasi <i>Elbow</i> Data Tahun 2022	41
4. 22	Jarak Interval Bin	42
4. 23	<i>Source Code</i> <i>Binning</i> Data	42
4. 24	<i>Sample</i> Hasil <i>Binning</i> Data Tahun 2019	42
4. 25	<i>Sample</i> Hasil <i>Binning</i> Data Tahun 2020	43
4. 26	<i>Sample</i> Hasil <i>Binning</i> Data Tahun 2021	43

4. 27 <i>Sample Hasil Binning Data Tahun 2022</i>	43
4. 28 <i>Source Code K-Means Clustering dengan Data Binning</i>	44
4. 29 <i>Hasil Clustering Data Binning Tahun 2019</i>	45
4. 30 <i>Hasil Clustering Data Binning Tahun 2020</i>	46
4. 31 <i>Hasil Clustering Data Binning Tahun 2021</i>	47
4. 32 <i>Hasil Clustering Data Binning Tahun 2022</i>	48
4. 33 <i>Clustering Data Tahun 2019</i>	49
4. 34 <i>Hasil Clustering Data Tahun 2019</i>	50
4. 35 <i>Clustering Data Tahun 2020</i>	50
4. 36 <i>Hasil Clustering Data Tahun 2020</i>	51
4. 37 <i>Clustering Data Tahun 2021</i>	52
4. 38 <i>Hasil Clustering Data Tahun 2021</i>	53
4. 39 <i>Clustering Data Tahun 2022</i>	53
4. 40 <i>Hasil Clustering Data Tahun 2022</i>	54

DAFTAR LAMPIRAN

No	Nama Lampiran	Halaman
7.1	Riwayat Hidup	66
7.2	Berita Acara Bimbingan	67