

**PENGELOMPOKAN DATA PENGANGGURAN DI
INDONESIA MENGGUNAKAN ALGORITMA *K-MEANS*
*CLUSTERING***

SKRIPSI

**Untuk Memenuhi Persyaratan
Memperoleh Gelar Sarjana**

Oleh:

**Novita Dwi Octaviana Ibrahim
NIM 19.51.0016**



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

**PENGELOMPOKAN DATA PENGANGGURAN DI
INDONESIA MENGGUNAKAN ALGORITMA *K-MEANS*
*CLUSTERING***

Oleh:
Novita Dwi Octaviana Ibrahim
NIM 19.51.0016

**Telah disetujui oleh pembimbing untuk diujikan
Pada tanggal 09 Mei 2023**

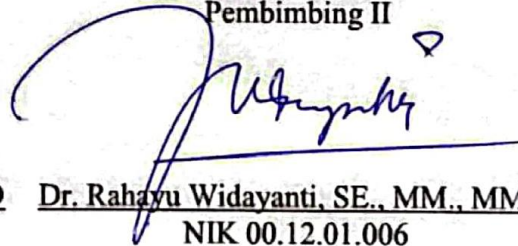
**Menyetujui,
Komisi Pembimbing**

Pembimbing I



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

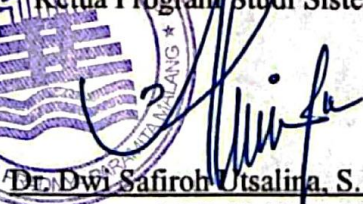
Pembimbing II



Dr. Rahayu Widayanti, SE., MM., MMSI
NIK 00.12.01.006



**Mengetahui,
Ketua Program Studi Sistem Informasi**

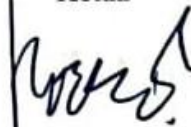

Dr. Dwi Safiroh Utsalina, S. Kom., MMSI
NIK 00.12.01.016

LEMBAR PENGESAHAN

**PENGELOMPOKAN DATA PENGANGGURAN DI
INDONESIA MENGGUNAKAN ALGORITMA *K-MEANS*
*CLUSTERING***

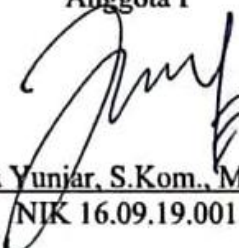
**Skripsi oleh Novita Dwi Octaviana Ibrahim
Telah dipertahankan di depan dewan penguji
Pada tanggal 19 Mei 2023**

Dewan Penguji,
Ketua



Dr. Mochamad Husni, S.Pd.I., M.M., MMSI
NIK 01.09.10.036

Anggota I



Eka Yunjar, S.Kom., MMSI
NIK 16.09.19.001

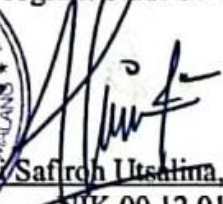
Anggota II



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

Mengetahui,

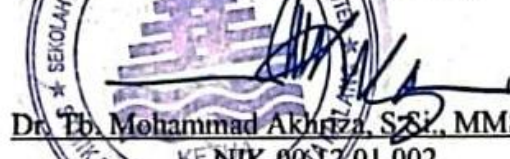
Ketua Program Studi S1-Sistem Informasi



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

Mengesahkan,

Ketua STMIK PPKIA
Pradnya Paramita Malang



Dr. Tb. Mohammad Akhriza, S.Si., MMSI Ph. D
KE 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 : Novita Dwi O. I.
Nim : 19.51.0016

ABSTRAK

Novita Dwi Octaviana Ibrahim. 2023. *Pengelompokan Data Pengangguran di Indonesia Menggunakan Algoritma K-Means Clustering*. Skripsi, Program Studi Sistem Informasi STMIK PPKIA Pradnya Paramita. Pembimbing: (I) Dr. Tb. Mohammad Akhriza, S.Si., MMSI Pd. D, (II) Dr. Rahayu Widayanti, SE., MM., MMSI

Kata Kunci: *Algorithm K-Means, Clustering, Elbow, Silhouette Coefficient*

Pengangguran merupakan masalah yang dihadapi oleh negara berkembang maupun maju, termasuk Indonesia. Menurut data BPS tingkat pengangguran dari tahun 2019 sampai 2021 mengalami peningkatan akibat adanya pandemi covid-19 yang terjadi di Indonesia. Pemerintah mengeluarkan kebijakan PSBB untuk mencegah persebaran covid-19. Penurunan pertumbuhan ekonomi membuat pelaku usaha melakukan efisiensi untuk mencegah kerugian. Akibatnya, pelaku usaha melakukan tindakan PHK sehingga berdampak pada peningkatan jumlah pengangguran. Penelitian ini bertujuan untuk mengcluster data pengangguran setiap provinsi di Indonesia tahun 2019 dan 2021. Hasil *clustering* menggunakan metode *K-Means* nilai *k* di optimasi dengan metode *elbow* dan di evaluasi menggunakan *silhouette coefficient*, dapat disimpulkan tahun 2019 terdapat 5 *cluster*, yaitu *cluster 0* ada 2 provinsi, *cluster 1* ada 16 provinsi, *cluster 2* ada 3 provinsi, *cluster 3* ada 12 provinsi, *cluster 4* ada 1 provinsi dan tahun 2021 terdapat 4 *cluster*, yaitu *cluster 0* ada 21 provinsi, *cluster 1* ada 3 provinsi, *cluster 2* dan *cluster 3* ada 5 provinsi. Terdapat perpindahan provinsi pada tahun 2019 dan 2021 yaitu provinsi Kalimantan Barat, NTT, NTB, dan Sulawesi Barat dari level 3 berpindah ke level 1 terdampak pengangguran. Provinsi Sulawesi Selatan, Maluku, Maluku Utara, Aceh, Kalimantan Utara, Kalimantan Timur, Banten, Jambi, Riau, dan Sulawesi Utara dari yang memiliki level 5 berpindah ke level 3 terdampak pengangguran. Provinsi Kalimantan Timur, DKI Jakarta, dan Kep. Riau dari yang memiliki level 5 berpindah ke level 4 terdampak pengangguran. Hasil evaluasi menetapkan tahun 2019 nilai $k=5$ dengan *silhouette* 0.36 dan 2021 nilai $k=4$ dengan *silhouette* 0.35 dapat dikatakan kriteria struktur *cluster* lemah.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah memberikan kesehatan, rahmat, hidayah serta kemudahan sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul **“PENGELOMPOKAN DATA PENGANGGURAN DI INDONESIA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING”** sebagai salah satu syarat untuk menyelesaikan Program Sarjana (S1) Jurusan Sistem Informasi STMIK Pradnya Paramita Malang.

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

1. Allah SWT, yang telah memberikan petunjuk dan hidayah dalam pembuatan skripsi dan penyusunan laporan sehingga dapat berjalan dengan baik dari awal hingga akhir.
2. Kedua orang tua yang telah memberikan doa, dukungan, dan motivasi sehingga penulis dapat menyelesaikan SKRIPSI ini.
3. Bapak Dr. Tb. Mohammad Akhriza, S.Si., MMSI Ph. D, selaku Ketua STMIK PPKIA Pradnya Paramita Malang, juga selaku dosen pembimbing I penyusun skripsi yang senantiasa meluangkan waktunya untuk memberikan bimbingan dan arahan serta saran yang diberikan kepada penulis sehingga penyusunan ini dapat diselesaikan dengan baik.
4. Ibu Dr. Rahayu Widayanti, SE., MM., MMSI, Selaku Dosen Pembimbing II penyusun skripsi yang senantiasa meluangkan waktunya untuk memberikan bimbingan dan arahan serta saran yang diberikan kepada penulis sehingga penyusunan ini dapat diselesaikan dengan baik.
5. Bapak Dr. Mochamad Husni, S.Pd.I., M.M., MMSI, selaku dosen wali yang telah memberikan semangat dan motivasi sehingga penulis dapat menyelesaikan dengan baik.
6. Ibu Dr. Dwi Safiroh Utsalina, S. Kom., MMSI, selaku Ketua Program Studi S1-Sistem Informasi STMIK PPKIA Pradnya Paramita Malang.
7. Seluruh Ibu dan Bapak Dosen dan Para staff STMIK PPKIA Pradnya Paramita Malang dalam membimbing, serta memberikan ilmu yang bermanfaat.

8. Kakak, Sepupu, Paman, Bibi yang telah memberikan semangat dan motivasi sehingga penulis dapat menyelesaikan SKRIPSI ini.
9. Teman-teman yang selalu memberikan semangat dan dukungan selama penyelesaian skripsi ini.
10. Dan seluruh pihak yang telah membantu dan mendukung lancarnya pembuatan Skripsi dari awal hingga akhir yang tidak bisa di sebutkan satu persatu.

Penulis mengharapkan saran dan kritik dari berbagai pihak yang bersifat membangun demi penyempurnaan skripsi ini. Semoga skripsi ini berguna bagi pembaca secara umum dan penulis secara khusus. Akhir kata, penulis ucapkan banyak terima kasih.

Malang, 29 Maret 2023

Penulis,

Novita Dwi Octaviana I

DAFTAR ISI

SKRIPSI.....	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN ORISINALITAS SKRIPSI	iv
ABSTRAK	v
KATA PENGANTAR	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan Penelitian	6
1.4 Batasan Masalah	6
1.5 Kontribusi Penelitian	6
BAB II TINJAUAN PUSTAKA.....	8
2.1 Data Mining	8
2.2 Metode <i>Clustering</i>	9
2.3 Metode <i>K-Means</i> dan <i>Elbow</i>	10
2.4 Penerapan <i>K-Means</i> untuk Pengelompokan Penduduk Pengangguran ..	15
2.5 Evaluasi <i>Clustering</i>	17
BAB III METODOLOGI PENELITIAN	19
3.1 Analisis Permasalahan	19
3.2 Solusi yang Diusulkan	20
3.2.1 Seleksi Data.....	20

3.2.2	Pre-processing	22
3.2.3	Menjalankan <i>K-Means</i> dengan Metode <i>Elbow</i>	23
3.2.4	<i>Clustering</i> dengan Nilai <i>k</i> yang Diperoleh.....	24
3.2.5	Evaluasi dan Analisis Hasil <i>Cluster</i>	24
3.3	Eksperimen	25
3.3.1	Kebutuhan Perangkat	25
3.3.2	Kebutuhan Data.....	25
BAB IV PENGUJIAN DAN HASIL		29
4.1	Seleksi Data	29
4.2	Pre-processing.....	29
4.3	Menjalankan <i>K-Means</i> dengan Metode <i>Elbow</i>	32
4.4	<i>Clustering</i> dengan Nilai <i>k</i> yang Diperoleh.....	38
4.5	Evaluasi dan Analisis Hasil <i>Cluster</i>	42
BAB V KESIMPULAN DAN SARAN.....		48
5.1	Kesimpulan	48
5.2	Saran	49
DAFTAR PUSTAKA		50

DAFTAR TABEL

No	Nama Tabel	Halaman
2. 1	Pengukuran <i>Silhouette Coefficient</i>	18
3. 1	Sampel Data Jumlah Penduduk.....	21
3. 2	Sampel Data Tingkat Partisipasi Angkatan Kerja.....	22
3. 3	Sampel Data Rata-rata lama sekolah penduduk umur 15 tahun ke atas	22
3. 4	Data Tahun 2019	26
3. 5	Data Tahun 2021	27
4. 3	Hasil <i>Clustering</i> Data Tahun 2019.....	41
4. 4	Hasil <i>Clustering</i> Tahun 2021	42
4. 5	Nilai Rata-Rata <i>Cluster</i> Tahun 2019.....	44
4. 6	Nilai Rata-Rata <i>Cluster</i> Tahun 2021	45

DAFTAR GAMBAR

No	Nama Gambar	Halaman
2. 1	<i>Flowchart</i> Algoritma <i>K-Means</i>	12
3. 1	Framework Penelitian	20
4. 1	Hasil Proses Normalisasi Data <i>Min-Max</i> Pada Tahun 2019	31
4. 2	Hasil Proses Normalisasi Data <i>Min-Max</i> Pada Tahun 2021	31
4. 3	Instalasi <i>numpy scikit-learn matplotlib</i>	32
4. 4	Tampilan <i>Import Library</i>	32
4. 5	Tampilan Data Tahun 2019.....	33
4. 6	Tampilan Data Variabel Berupa <i>Array</i> Tahun 2019.....	33
4. 7	Normalisasi Data Tahun 2019.....	34
4. 8	Normalisasi Data Tahun 2021.....	34
4. 9	<i>Cluster</i> Dengan Metode <i>Elbow</i>	35
4. 10	Hasil Nilai <i>k</i> Data Tahun 2019.....	36
4. 11	Hasil Nilai <i>k</i> Data Tahun 2021.....	36
4. 12	Grafik <i>Silhouette Coefficient</i> Tahun 2019.....	37
4. 13	Grafik <i>Silhouette Coefficient</i> Tahun 2021.....	37
4. 14	Tampilan Hasil <i>cluster K-Means</i> Data Tahun 2019.....	38
4. 15	Tampilan Hasil <i>cluster K-Means</i> Data Tahun 2021.....	38
4. 16	Hasil Penambahan Kolom <i>Cluster</i> Tahun 2019.....	39
4. 17	Hasil Penambahan Kolom <i>Cluster</i> Tahun 2021.....	40
4. 18	Hasil Koefisien Korelasi tahun 2019	42

DAFTAR LAMPIRAN

No	Nama Lampiran	Halaman
7.1	Riwayat Hidup	53
7.2	Berita Acara Bimbingan	54