



SKRIPSI

KLASTERING PENCARI KERJA MENGUNAKAN METODE KMEANS & PCA DI DINAS PERINDUSTRIAN DAN TENAGA KERJA KOTA SURABAYA

TRIYATUL DEWI SAFITRI
NPM 21082010007

Dosen Pembimbing

Amalia Anjani Arifiyanti, S.Kom., M.Kom
Agung Brastama Putra, S.Kom., M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2025**



SKRIPSI

**KLASTERING PENCARI KERJA
MENGUNAKAN METODE KMEANS & PCA DI
DINAS PERINDUSTRIAN DAN TENAGA KERJA
KOTA SURABAYA**

TRIYATUL DEWI SAFITRI
NPM 21082010007

Dosen Pembimbing
Amalia Anjani Arifiyanti, S.Kom., M.Kom
Agung Brastama Putra, S.Kom., M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SISTEM INFORMASI
SURABAYA
2025**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

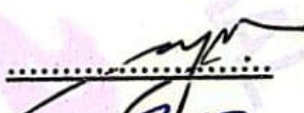
KLASTERING PENCARI KERJA MENGGUNAKAN METODE KMEANS & PCA DI DINAS PERINDUSTRIAN DAN TENAGA KERJA KOTA SURABAYA

Oleh :
TRIYATUL DEWI SAFITRI
NPM. 21082010007

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 09 Mei 2025.

Menyetujui,

Amalia Anjani Arifiyanti, S.Kom., M.Kom
NIP. 19920812 201803 2 001


..... (Pembimbing I)

Agung Brastama Putra, S.Kom., M.Kom
NIP. 19851124 202121 1 003


..... (Pembimbing II)

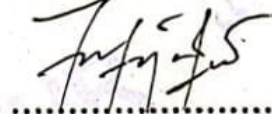
Eka Dyar Wahyuni, S.Kom., M.Kom
NIP. 19841201 202121 2 005


..... (Ketua Penguji)

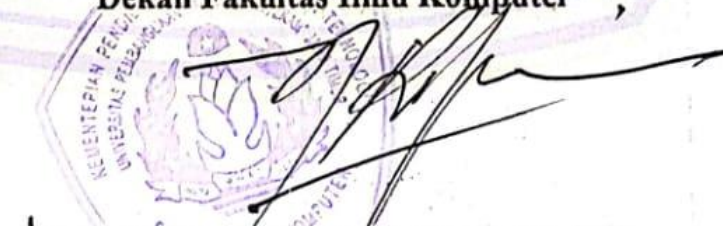
Dhian Satria Yudha Kartika, S.Kom., M.Kom
NPT. 201 19860522 249


..... (Anggota Penguji II)

Seftin Fitri Ana Wati, S.Kom., M.Kom
NPT. 212 19910320 267


..... (Anggota Penguji III)

Mengetahui,
Dekan Fakultas Ilmu Komputer


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

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

**KLASTERING PENCARI KERJA MENGGUNAKAN METODE KMEANS
& PCA DI DINAS PERINDUSTRIAN DAN TENAGA KERJA KOTA
SURABAYA**

Oleh:

TRİYATUL DEWI SAFITRI

NPM. 21082010007

Menyetujui,

Koordinator Program Studi Sistem Informasi

Fakultas Ilmu Komputer


Agung Brastama Putra, S.Kom., M.Kom

NIP. 19851124 202121 1 003

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertandatangan di bawah ini:

Nama Mahasiswa : Triyatul Dewi Safitri
NPM : 21082010007
Program : Sarjana (S1)
Program Studi : Sistem Informasi
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur – unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang – undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 03 Juni 2025
Yang Membuat Pernyataan,



Triyatul Dewi Safitri
NPM. 21082010007

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM : Triyatul Dewi Safitri / 21082010007
Judul Skripsi : Klastering Pencari Kerja Menggunakan Metode Kmeans & PCA Di Dinas Perindustrian dan Tenaga Kerja Kota Surabaya
Dosen Pembimbing : 1. Amalia Anjani Arifiyanti, S.Kom., M.Kom
2. Agung Brastama Putra, S.Kom., M.Kom

Pencari kerja mempunyai latar belakang yang beragam seperti tingkat pendidikan, usia, pengalaman kerja dan kondisi lainnya yang dapat mempengaruhi peluang pencari kerja dalam mendapatkan pekerjaan. Penelitian ini bertujuan untuk mengelompokkan pencari kerja di Kota Surabaya berdasarkan karakteristik yang sama sehingga dapat membantu Dinas Perindustrian dan Tenaga Kerja Kota Surabaya melakukan strategi peningkatan keterampilan pada pencari kerja untuk membantu mendapatkan pekerjaan yang sesuai dengan kebutuhan dan kondisi masing – masing. Metode penelitian menggunakan klaster analisis dengan algoritma K-Means dan PCA, selanjutnya dilakukan uji coba beberapa skenario untuk mendapatkan hasil struktur klaster yang optimal. Hasil penelitian ini menunjukkan perbandingan yang signifikan dari beberapa uji coba menggunakan PCA dan tanpa PCA. Setelah dilakukan evaluasi model dan analisis perbandingan dari struktur klaster didapatkan uji coba terbaik pada penelitian ini menggunakan algoritma K-Means tanpa PCA, Dengan hasil evaluasi berikut, didapatkan hasil *silhouette score* sebesar 0,71 sedangkan *davies bouldin index* sebesar 0,37, dari kedua metode evaluasi ini saling mendukung untuk menunjukkan bahwa struktur klaster yang dihasilkan termasuk dalam kategori kuat. Dengan hasil tersebut penelitian ini didapatkan *insight* bahwa yang mempengaruhi pencari kerja dalam mendapatkan pekerjaan adalah dari latar belakang pendidikan atau jurusan dan usia pencari kerja, sehingga hal ini dapat dijadikan strategi dalam peningkatan program pelatihan pada Dinas Perindustrian dan Tenaga Kerja Kota Surabaya untuk menambah pelatihan dibidang non – teknik khususnya di bidang ilmu hukum, manajemen dan ekonomi agar pencari kerja pada bidang tersebut tidak kesulitan dalam mendapatkan pelatihan dan sertifikasi sesuai bidang mereka dan meminimalisir pencari kerja yang melamar pekerjaan tidak sesuai bidangnya.

Kata kunci : Klasterisasi, Pencari Kerja, K-Means, PCA, *Silhouette Score*, DBI

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Triyatul Dewi Safitri / 21082010007
Thesis Title : Clustering Job Seekers Using Kmeans & PCA
Methods at Surabaya City Industry and Labor
Department
Advisor : 1. Amalia Anjani Arifiyanti, S.Kom., M.Kom
2. Agung Brastama Putra, S.Kom., M.Kom

ABSTRACT

Job seekers have diverse backgrounds such as education level, age, work experience and other conditions that can affect job seekers' chances of getting a job. This research aims to group job seekers in Surabaya City based on similar characteristics so that it can help the Surabaya City Industry and Manpower Office to carry out a strategy to improve skills in job seekers to help get jobs that are in accordance with their respective needs and conditions. The research method uses cluster analysis with K-Means and PCA algorithms, then several scenarios are tested to get the optimal cluster structure results. The results of this study show a significant comparison of several trials using PCA and without PCA. After evaluating the model and analyzing the comparison of the cluster structure, the best trial in this study was obtained using the K-Means algorithm without PCA, with the following evaluation results, the shilhouette score was 0.71 while the davies bouldin index was 0.37, from these two evaluation methods support each other to show that the resulting cluster structure is included in the strong category. With these results, this study obtained insight that what influences job seekers in getting a job is from the educational background or major and the age of the job seeker, so that this can be used as a strategy in improving training programs at the Surabaya City Industry and Manpower Office to add training in non-engineering fields, especially in the fields of law, management and economics so that job seekers in these fields have no difficulty in obtaining training and certification according to their fields and minimizing job seekers who apply for jobs not according to their fields.

Keywords: Clustering, Job Seekers, K-Means, PCA, Shilhouette Score, DBI

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur kehadirat Allah SWT atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Klastering Pencari Kerja Menggunakan Metode Kmeans & PCA Di Dinas Perindustrian dan Tenaga Kerja Kota Surabaya”** dapat terselesaikan dengan baik.

Penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Kedua Orang Tua penulis serta seluruh keluarga yang selalu memberikan doa, dukungan dan semangat, serta menjadi sumber motivasi dalam setiap perjuangan penulis
2. Ibu Amalia Anjani Arifiyanti, S.Kom.,M.Kom selaku Dosen Pembimbing pertama yang telah meluangkan waktunya untuk memberikan bimbingan dengan penuh kesabaran dan perhatian, memberikan arahan, nasehat serta motivasi kepada penulis
3. Bapak Agung Brastama Putra, S.Kom., M.Kom selaku Dosen Pembimbing kedua sekaligus Ketua Program Studi Sistem Informasi Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang telah meluangkan waktu untuk memberikan arahan,bimbingan kepada serta dukungan kepada penulis
4. Bapak Nur Cahyo Wibowo, S.Kom., M.Kom selaku Dosen Wali yang telah memberikan bimbingan, arahan, serta motivasi selama masa perkuliahan, sehingga penulis dapat menyelesaikan pendidikan dengan baik
5. Dosen-dosen Program Studi Sistem Informasi yang telah memberikan ilmu, wawasan, serta bimbingannya kepada penulis selama masa perkuliahan .
6. Pak Adit selaku staff IT Disperinaker Surabaya dan terima kasih kepada Bapak Kepala Bidang Industri Dinas Perindustrian dan Tenaga Kerja Kota Surabaya yang telah memperbolehkan dalam proses permintaan data pencari kerja yang

mendaftar pada *website* ASSIK dan membantu dalam proses penelitian pengumpulan data

7. M.Fauzan Alfauqi, yang menjadi sumber semangat, menemani penulis dalam setiap proses dan perjuangan hingga tugas akhir ini dapat terselesaikan.
8. Teman – teman lapak sambat yang telah menemani, mendukung dan berbagi suka maupun duka selama masa perkuliahan hingga tersusunnya tugas akhir ini.
9. Vioneta Zarazsofia Permatasari yang telah menemani penulis lari setiap sore untuk melepas stress pikiran dari tekanan yang ada dalam hidup dan memberikan keyakinan kepada penulis bahwa penulis dapat menyelesaikannya dengan baik
10. Gading Putri Diniarti, Rhohmah Indah Mekar Sari, Nimas Adjeng Nutfa Rabbaanii yang menemani dari awal perkuliahan dan memberikan semangat untuk penulis dari awal penelitian hingga terselesaikan
11. Shinta Amalia Azhara yang menemani penulis untuk ke perpustakaan dan memberikan motivasi kepada penulis agar lebih cepat dalam menyelesaikan skripsi
12. Fina, Ismi, Afi, Arin yang telah menjadi teman seperjuangan yang selalu memberikan semangat dan menemani dalam setiap langkah.
13. Teman – teman musyawarah yang telah mewujudkan salah satu *wishlist* penulis, yaitu mendaki gunung bersama untuk mencari inspirasi dalam penyusunan skripsi, terima kasih atas pengalaman berharga dan semangat yang diberikan

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, 29 April 2025

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN BEBAS PLAGIASI.....	vii
ABSTRAK	ix
DAFTAR ISI.....	xv
DAFTAR GAMBAR.....	xvii
DAFTAR TABEL	xix
DAFTAR NOTASI.....	xx
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	4
1.4. Tujuan	5
1.5. Sistematika Penulisan	5
1.6. Manfaat Penelitian	6
BAB II TINJAUAN PUSTAKA.....	9
2.1. Penelitian Terdahulu	9
2.2. Landasan Teori.....	19
2.2.1 Pencari Kerja.....	19
2.2.2 Dinas Perindustrian dan Tenaga Kerja Kota Surabaya	19
2.2.3 Data Mining	20
2.2.4 CRISP-DM.....	20
2.2.5 Klastering.....	21
2.2.6 <i>K-Means</i>	22
2.2.7 Python	23
2.2.8 Metode Elbow	24
2.2.9 PCA (<i>Principal Component Analysis</i>)	24
2.2.10 <i>Silhouette Coefficient</i>	24
2.2.11 <i>Davies Bouldin Index (DBI)</i>	25

BAB III METODOLOGI PENELITIAN	27
3.1. Metode Penelitian.....	27
3.1.1 Identifikasi Masalah	27
3.1.2 Studi Literatur	28
3.1.3 Analisis Kebutuhan	28
3.2. Desain Sistem.....	31
3.2.1 <i>Bussiness Understanding</i>	31
3.2.2 <i>Data Understanding</i>	31
3.2.3 <i>Data Preparation</i>	32
3.2.4 <i>Modelling</i>	42
3.2.5 <i>Evaluation</i>	43
3.2.6 <i>Deployment</i>	43
BAB IV HASIL DAN PEMBAHASAN	45
4.1 Hasil Penelitian.....	45
4.1.1 Pengumpulan Data	45
4.1.2 <i>Business Understanding</i>	46
4.1.3 <i>Data Understanding</i>	47
4.1.4 <i>Data Preparation</i>	53
4.1.5 Normalisasi Data	62
4.1.6 Modelling K-Means Tanpa PCA Uji Coba 3 Klaster	63
4.1.7 Modelling K-Means Tanpa PCA Uji Coba 4 Klaster	66
4.1.8 Modelling Kombinasi Seluruh Atribut K-Means Dengan PCA	70
4.1.9 Evaluation Kmeans dengan PCA	81
4.1.10 Perbandingan Hasil Evaluasi Klaster	85
4.1.11 <i>Deployment</i>	86
4.2 Pembahasan	94
BAB V KESIMPULAN DAN SARAN	97
5.1 Kesimpulan.....	97
5.2 Saran	98
DAFTAR PUSTAKA	99
LAMPIRAN	

DAFTAR GAMBAR

Gambar 2. 1 Rumus Euclidian Distance dari Nurahman & Dwi Aulia.....	23
Gambar 3. 1 Alur Tahapan Skripsi	27
Gambar 3. 2 Alur Pengumpulan Data	29
Gambar 3. 3 <i>Business Understanding</i>	31
Gambar 3. 4 <i>Data Understanding</i>	31
Gambar 3. 5 <i>Data Preparation</i>	32
Gambar 3. 7 Hasil <i>Mockup Deployment</i>	43
Gambar 4. 1 Wawancara Di Disperinaker Kota Surabaya.....	45
Gambar 4. 2 Diagram Batang Pencari Kerja Per Tahun	46
Gambar 4. 3 Histogram Usia dan Frekuensi	47
Gambar 4. 4 Grafik <i>Scatter Plot</i> Usia Vs Nama Prodi	48
Gambar 4. 5 Diagram Batang Tingkat Pendidikan	49
Gambar 4. 6 <i>Heatmap</i> Korelasi Fitur	51
Gambar 4. 7 Scatter Plot Usia dengan Pengalaman Kerja	52
Gambar 4. 8 Diagram Batang Top 10 Prodi	53
Gambar 4. 9 <i>Data Preprocessing</i> Pengecekan Data Duplikat	53
Gambar 4. 10 Pengecekan <i>Missing Value</i>	54
Gambar 4. 11 Penanganan <i>Missing Value</i> Pada Atribut Status Pernikahan	54
Gambar 4. 12 Penanganan <i>Missing Value</i> Pada Atribut Nama Prodi.....	55
Gambar 4. 13 Hasil <i>Missing Value</i> Setelah Pengisian	55
Gambar 4. 14 Penghapusan Data yang Tidak Relevan	56
Gambar 4. 15 Pengubahan Isi Data Pengalaman Kerja Sesuai Format	56
Gambar 4. 16 Pengubahan Isi Data Sertifikat Sesuai Format	57
Gambar 4. 17 Pengubahan Nama Atribut Nama Pendidikan.....	58
Gambar 4. 18 Menyimpan Hasil <i>Data Preparation</i>	58
Gambar 4. 19 Pengecekan Tipe Data Atribut yang Digunakan	59
Gambar 4. 20 Pengubahan Tipe Data Atribut Menjadi Numerik	59
Gambar 4. 21 Hasil Tipe Data Setelah Proses Label Encoding.....	60
Gambar 4. 22 Statistik Deskriptif Dataset Pencari Kerja.....	61
Gambar 4. 23 Normalisasi Data	62

Gambar 4. 24 Metode Elbow Menggunakan Kmeans Tanpa PCA 3 Klaster	63
Gambar 4. 25 Kombinasi Seluruh Atribut Kmeans Tanpa PCA 3 Klaster	64
Gambar 4. 26 Hasil Centroid Kmeans Tanpa PCA Uji Coba 3 Klaster	64
Gambar 4. 27 Metode Elbow Menggunakan Kmeans Tanpa PCA 4 Klaster	66
Gambar 4. 28 Kombinasi Seluruh Atribut Kmeans Tanpa PCA 4 Klaster	67
Gambar 4. 29 Hasil Centroid Kmeans Tanpa PCA Uji Coba 4 Klaster	68
Gambar 4. 30 Variansi Kumulatif PCA Menggunakan Normalisasi	70
Gambar 4. 31 Variansi yang Dijelaskan Per-Komponen Utama	70
Gambar 4. 32 Metode Elbow Kmeans dengan PCA Menggunakan Normalisasi	71
Gambar 4. 33 Biplot PCA Menggunakan Normalisasi	72
Gambar 4. 34 Plot Skor Pemuatan PCA Menggunakan Normalisasi	73
Gambar 4. 35 Plot varians Nama Prodi Menggunakan Normalisasi	75
Gambar 4. 36 Plot Varians Tingkat Pendidikan Menggunakan Normalisasi	76
Gambar 4. 37 Hasil Klastering dengan PCA Kmeans	77
Gambar 4. 38 Hasil Centroid Kmeans dengan PCA	80
Gambar 4. 39 Hasil <i>Shilhouette Score</i> Kmeans Tanpa PCA Uji Coba 3 Klaster	81
Gambar 4. 40 Hasil DBI Kmeans Tanpa PCA Uji Coba 3 Klaster	82
Gambar 4. 41 Hasil <i>Shilhouette Score</i> Kmeans Tanpa PCA Uji Coba 4 Klaster	82
Gambar 4. 42 Hasil DBI Kmeans Tanpa PCA Uji Coba 4 Klaster	83
Gambar 4. 43 Hasil Validasi <i>Shilhouette Score</i> Terhadap Kmeans - PCA	83
Gambar 4. 44 Hasil Validasi DBI Terhadap Kmeans - PCA	84
Gambar 4. 45 Hasil Evaluasi Kmeans dengan PCA	84
Gambar 4. 46 Tampilan Awal Dashboard Semua Tahun	86
Gambar 4. 47 Tampilan Lanjutan Isi Dashboard	87
Gambar 4. 48 Tampilan Dashboard Data Tahun 2022	87
Gambar 4. 49 Tampilan Dashboard Data Tahun 2023	88
Gambar 4. 50 Tampilan Dashboard Data Tahun 2024	89
Gambar 4. 51 Tampilan Step Awal Fitur Hasil Klastering	90
Gambar 4. 52 Tampilan Hasil Klastering	91
Gambar 4. 53 Tampilan Lanjutan Hasil Klastering	92
Gambar 4. 54 Lanjutan Fitur Hasil Klaster	93

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	9
Tabel 3. 1 Contoh Dataset Pencari Kerja Pada Website ASSIK.....	30
Tabel 3. 2 Perbaikan Data	32
Tabel 3. 3 Transformasi Data pada Atribut Pengalaman Kerja	33
Tabel 3. 4 Transformasi Data pada Atribut Nama Prodi	34
Tabel 3. 5 Transformasi Data pada Atribut Tingkat Pendidikan	40
Tabel 3. 6 Transformasi Data pada Atribut Status Pernikahan.....	41
Tabel 3. 7 Transformasi Data pada Atribut Status Disabilitas.....	41
Tabel 3. 8 Perbandingan Hasil Evaluasi Klaster.....	85

Halaman ini sengaja dikosongkan

DAFTAR NOTASI

D_e	:	<i>Eucledian Distance</i>
x,y	:	Koordinat objek
(s,t)	:	Koordinat <i>centroid</i>
i	:	Banyak Objek

Halaman ini sengaja dikosongkan