

SKRIPSI
ALGORITMA *CLUSTERING K-MEANS* DALAM ANALISIS *CLUSTER*
GAJI BERDASARKAN DATA DEMOGRAFI DENGAN *MECHINE*
***LEARNING* MENGGUNAKAN PYTHON**



NOFI YANTI

205610042

PROGRAM STUDI SISTEM INFORMASI
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA
2024

SKRIPSI

**ALGORITMA *CLUSTERING K-MEANS* DALAM ANALISIS *CLUSTER*
GAJI BERDASARKAN DATA DEMOGRAFI DENGAN *MECHINE*
LEARNING MENGGUNAKAN PYTHON**

Diajukan sebagai salah satu syarat untuk menyelesaikan studi Program

Sarjana

Program Studi Sistem Informasi Fakultas Teknologi Informasi

Universitas Teknologi Digital Indonesia Yogyakarta



Disusun Oleh

NOFI YANTI

NIM : 205610042

**PROGRAM STUDI SISTEM INFORMASI
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

UJIAN SKRIPSI

Judul : Algoritma *Clustering K-Means* Dalam Analisis
Cluster Gaji Berdasarkan Data Demografi Dengan
Mechine Learning Menggunakan Python

Nama : Nofi Yanti

NIM : 205610042

Program Studi : Sistem Informasi

Program : Sarjana

Semester : Gasal

Tahun Akademik : 2023/2024

Telah diperiksa dan disetujui untuk diujikan di hadapan Dewan Penguji Skripsi

Yogyakarta,2024

Dosen Pembimbing,






Deborah Kurtiawati, S.Kom., M.Cs.

NPP:051149

HALAMAN PENGESAHAN
SKRIPSI
ALGORITMA CLUSTERING K-MEANS DALAM ANALISIS
CLUSTER GAJI BERDASARKAN DATA DEMOGRAFI DENGAN
MECHINE LEARNING MENGGUNAKAN PYTHON

Telah dipertahankan di depan Dewan Penguji Skripsi dan dinyatakan
diterima untuk memenuhi sebagian persyaratan guna memperoleh Gelar
Sarjana Sistem Informasi
Program Studi Sistem Informasi
Fakultas Teknologi Informasi
Universitas Teknologi Digital Indonesia
Yogyakarta
Yogyakarta,.....

Dewan Penguji	NIDN	Tandatangan
1. Edi Faizal, S.T., M.Cs.	0512088201	
2. Dr. Asyahri Hadi Nasyuha S. Kom., M.Kom.	0129048601	
3. Deborah Kurniawati, S.Kom., M.Cs.	0511107301	

Mengetahui

Ketua Program Studi Sistem Informasi



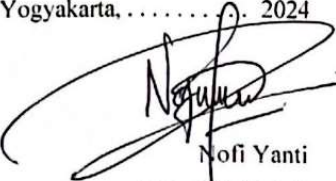
Deborah Kurniawati, S.Kom., M.Cs.

NPP:051149

PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya menyatakan bahwa naskah skripsi ini belum pernah diajukan untuk memperoleh gelar Sarjana Sistem Informasi di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara sah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 2024



Nofi Yanti
NIM : 205610042

HALAMAN PERSEMBAHAN

Puji syukur hanya kepada Allah SWT yang telah melimpahkan rahmat dan karunia-Nya kepada penulis, sehingga dapat menyelesaikan penyusunan laporan akhir ini. Skripsi ini dipersembahkan sebagai ungkapan terima kasih kepada:

1. Kedua orang tua, Bapak Eka Setiawan dan Ibu Junari, yang dengan penuh dedikasi telah mendidik, memberikan dukungan, dan menaruh kepercayaan sepenuhnya pada penulis.
2. Keluarga besar yang senantiasa memberikan dukungan dan semangat, menjadi pilar kuat dalam meraih kesuksesan.
3. Terima kasih kepada Ibu Deborah Kurniawati, S.Kom., M.Cs, atas bimbingan dan semangatnya yang luar biasa, yang menjadi pendorong utama dalam menyelesaikan skripsi ini serta memberikan wejangan, arahan, waktu, dan ilmunya yang berharga.
4. Teman-teman sesama alumni dari SMAN 1 Bolo Bima yang telah memberikan dukungan sejak awal hingga saat ini.
5. Untuk sahabat di kost dan rekan mahasiswa Sistem Informasi angkatan 2020, yang telah berbagi pengalaman dan cerita di setiap langkah perjalanan dari awal perkuliahan hingga akhir.

MOTTO

“Kamu sekalian akan menghadapi ujian dalam harta dan jiwa, dan kamu sekalian pasti akan mendengar banyak hal yang sangat menyakitkan dari orang-orang yang diberikan kitab sebelum kamu dan dari orang-orang yang musyrik. Tetapi jika kamu bersabar dan bertakwa, maka sesungguhnya yang demikian itu termasuk hal yang diutamakan.” (Q.S. Al-Imran: 186)

“Sesungguhnya Allah SWT tidak melihat bentuk tubuh kalian dan tidak pula melihat keindahan wajah kalian, tetapi Allah melihat hati dan amal-amal kalian.”
(HR. Muslim)

“Sebaik-baik manusia adalah yang paling bermanfaat bagi orang lain.” (HR Ath-Thabrani)

KATA PENGANTAR

Bismillahirrahmanirrahim,

Puji syukur hanya bagi Allah SWT atas karunia-Nya yang melimpah, yang telah menjadi penyokong utama dalam penyelesaian tugas akhir ini. Dengan penuh rasa syukur, tugas akhir ini disusun sebagai langkah terakhir dalam perjalanan akademis untuk meraih gelar Sarjana Komputer dari Program Studi Sistem Informasi Universitas Teknologi Digital Indonesia.

Penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, arahan, dan bantuan selama proses penulisan tugas akhir ini terutama kepada orang tua dan keluarga atas doa, dukungan, dan semangat yang telah diberikan sepanjang perjalanan kami dalam menyelesaikan pendidikan ini. Tanpa cinta dan dorongan dari keluarga, pencapaian tugas akhir ini tidak akan menjadi kenyataan.. Tanpa kontribusi dan dorongan dari berbagai pihak, pencapaian ini tidak akan menjadi kenyataan. Oleh karena itu, pada kesempatan ini, saya ingin menyampaikan penghargaan yang tulus kepada:

1. Ibu Sri Redjeki, S.Si., M.Kom., PhD., selaku Rektor di Universitas Teknologi Digital Indonesia.
2. Ibu Deborah Kurniawati, S.Kom., M.Cs., selaku Kepala Program Studi Sistem Informasi dan selaku dosen pembimbing, atas bimbingan, arahan, dan kesabaran yang diberikan selama proses penulisan tugas akhir. Ilmu dan bimbingan yang diterima telah memberikan bekal berharga untuk pengembangan pengetahuan selanjutnya.
3. Bapak Edi Faizal, S.T., M.Cs. dan Dr. Asyahri Hadi Nasyuha S. Kom.,

M.Kom. selaku dosen penguji dalam sidang, telah memberikan masukan yang berharga untuk penyempurnaan penyusunan laporan akhir.

4. Para dosen dari Program Studi Sistem Informasi di Universitas Teknologi Digital Indonesia.
5. Kepada rekan-rekan atas semangat, dukungan batin, dan bantuan yang telah diberikan sepanjang proses penulisan laporan tugas akhir ini.

DAFTAR ISI

SKRIPSI.....	ii
HALAMAN PERSETUJUAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN	v
MOTTO	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR	xiii
INTISARI.....	xiv
<i>ABSTRACT</i>	xv
BAB I_PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup	4
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	5
1.6 Sistematika Penulisan.....	5
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	6
2.1. Tinjauan Pustaka	6
2.2. Dasar Teori.....	10
2.2.1. Gaji.....	10
2.2.2. Demografi Profil	11
2.2.3. Algoritma K-means <i>Clustering</i>	12
2.2.4. Jarak Euclidian	14
2.2.5. <i>Clustering</i>	14
2.2.6. <i>Elbow Method</i>	16
2.2.7. <i>Davies-Bouldin Index</i>	17
2.2.8. Silhouette Score	18
2.2.9. Python	19
BAB III METODE PENELITIAN.....	22
3.1. Bahan.....	22

3.1.1.	Kebutuhan Input.....	22
3.1.2.	Kebutuhan Output.....	22
3.2.	Peralatan.....	23
3.3.	Kebutuhan Perangkat Lunak.....	23
3.4.	Kebutuhan Perangkat Keras.....	23
3.5.	Prosedur dan Pengumpulan Data.....	23
3.6.	Analisis dan Rancangan Sistem.....	26
BAB IV IMPLEMENTASI DAN PEMBAHASAN.....		31
4.1.	Implementasi.....	31
4.1.1.	Data Schema.....	31
4.1.2.	Data <i>Preprocessing</i>	32
4.1.3.	<i>Model Selection</i>	38
4.1.4.	Validasi Model.....	40
4.2.	Pembahasan.....	43
4.2.1.	Hasil Uji <i>Elbow Method</i> dan <i>Silhouette Method</i> pada K-Means.....	44
4.2.2.	Hasil Visualisasi <i>Cluster</i> K-means.....	45
4.2.3.	<i>Cluster</i> 1.....	46
4.2.4.	<i>Cluster</i> 2.....	47
BAB V SIMPULAN DAN SARAN.....		49
5.1	Simpulan.....	49
5.2	Saran.....	50
DAFTAR PUSTAKA.....		52
LAMPIRAN.....		55

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	9
Tabel 4. 1 Atribut Valiariabel Dataset	31
Tabel 4. 2 Hasil Perbandingan Evaluasi Metrik dan Jumlah <i>Cluster</i> k	43
Tabel 4. 3 Hasil Validasi <i>Cluster</i>	43
Tabel 4. 4 Keterangan pada Gambar 4. 20.....	45

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	24
Gambar 3. 2 Alur Proses Analisis.....	26
Gambar 4. 1 Kode Menghapus Atribut.....	32
Gambar 4. 2 Kode Membersihkan Karakter.....	33
Gambar 4. 3 <i>Set Lower Case</i>	33
Gambar 4. 4 Cek Duplikat.....	34
Gambar 4. 5 Hasil Data Cleaning.....	35
Gambar 4. 6 Kode Membuat Atribut <i>Gender</i>	35
Gambar 4. 7 Kode Transformasi Atribut.....	36
Gambar 4. 8 Data Transformasi.....	37
Gambar 4. 9 Kode Remove Outliers.....	37
Gambar 4. 10 Kode Data Scalling.....	38
Gambar 4. 11 Kode <i>Elbow Method</i>	38
Gambar 4. 12 Kode K-Means <i>Clustering</i>	39
Gambar 4. 13 K-Means Centroids.....	40
Gambar 4. 14 Code DBI Uji K.....	40
Gambar 4. 15 Code Silhouette Uji K.....	41
Gambar 4. 16 Code DBI Uji <i>Clustering</i>	42
Gambar 4. 17 Code Silhouette Uji <i>Clustering</i>	42
Gambar 4. 18 Hasil Uji <i>Elbow Method</i>	44
Gambar 4. 19 Hasil Uji <i>Silhouette Method</i>	44
Gambar 4. 20 Visualisasi K-Means.....	45
Gambar 4. 21 <i>Cluster 1</i>	46
Gambar 4. 22 <i>Cluster 2</i>	47

INTISARI

Penelitian ini mengadopsi metode K-Means *Clustering* dalam analisis gaji berdasarkan data demografis, dengan tujuan memahami pola pengelompokan gaji karyawan berdasarkan faktor-faktor demografis yang relevan. Metode K-Means dipilih karena kemampuannya dalam mengelompokkan data secara efisien dan efektif, serta kemampuannya dalam menangani data *unsupervised learning*. Dalam konteks ini, K-Means dapat membantu mengidentifikasi pola pengelompokan gaji yang tidak terlihat secara langsung oleh analisis tradisional.

Proses analisis dimulai dengan persiapan data dengan jumlah 3000 baris data, yang melibatkan pembersihan, transformasi data menjadi numerik, penghapusan outlier dan penskalaan. Langkah ini untuk memastikan kualitas data yang digunakan dalam analisis. Setelah itu, dilakukan pemilihan model dan klastering menggunakan algoritma K-Means dengan python. Proses ini melibatkan penentuan jumlah *cluster* yang optimal dengan metode *elbow* atau siku, dan pembentukan *cluster* menggunakan centroid dan pengelompokan objek berdasarkan jarak minimum ke centroid.

Hasil evaluasi menunjukkan bahwa K-Means dengan 2 *cluster* optimal dimana *cluster* 1 terdapat 1518 data dengan rata-rata gaji Rp 7jt dan *cluster* 2 terdapat 1482 baris data dengan rata-rata gaji Rp 8jt. Dua *cluster* ini menghasilkan *Silhouette Score* sebesar 0.9972 mendekati 1 dan *Davies-Bouldin Index* sebesar 0.00178 mendekati nol, menunjukkan kualitas klastering yang sangat baik dan k=2 adalah nilai k yang paling optimal.

Kata kunci: *K-Means Clustering, Analisis Gaji, Davies-Bouldin Index, Silhouette Score*

ABSTRACT

This study adopts the K-Means Clustering method in salary analysis based on demographic data, with the aim of finding out the Clustering pattern of employee salaries based on relevant demographic factors. The K-Means method was chosen due to its ability to Cluster data efficiently and effectively, as well as its ability to handle unsupervised learning data. In this context, K-Means can help identify salary Clustering patterns that cannot be seen directly by traditional analysis.

The analysis process began with data preparation of 3000 rows of data, which involved cleaning, transformation of the data to numeric, removal of outliers and scaling. This step was to ensure the quality of the data used in the analysis. After that, model selection and clustering using the K-Means algorithm with python were performed. This process involves determining the optimal number of clusters using the elbow method, and forming clusters using centroids and grouping objects based on the minimum distance to the centroid.

The evaluation results show that K-Means with 2 clusters is optimal where cluster 1 has 1518 data with an average salary of Rp 7 million and cluster 2 has 1482 rows of data with an average salary of Rp 8 million. These two clusters produce a Silhouette Score of 0.9972 close to 1 and a Davies-Bouldin Index of 0.00178 close to zero indicating excellent clustering quality and $k=2$ is the most optimal k value.

Keywords: *K-Means Clustering, Salary Analysis, Davies-Bouldin Index , Silhouette Score*