

**KLASTERISASI PEGAWAI PT JASAMARGA TOLLROAD  
OPERATOR MENGGUNAKAN METODE K-MEANS  
BERDASARKAN DATA KEHADIRAN**

**SKRIPSI**

**Oleh :  
Ananda Setyawati  
202010225243**



**PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BHAYANGKARA JAKARTA RAYA  
2024**

## LEMBAR PERSETUJUAN PEMBIMBING

Judul Skripsi : Klasterisasi Pegawai PT Jasamarga Tollroad  
Operator Menggunakan Metode K-Means  
Berdasarkan Data Kehadiran

Nama Mahasiswa : Ananda Setyawati

Nomor Pokok Mahasiswa : 202010225243

Program Studi / Fakultas : Informatika / Ilmu Komputer

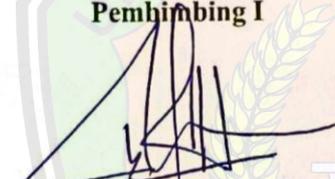
Tanggal Lulus Ujian Skripsi : 26 Juli 2024

Jakarta, 26 Juli 2024

MENYETUJUI,

Pembimbing I

Pembimbing II

  
( Dr. Mujiono Sadikin, S.T., M.T. )

NIDN : 0406127002

  
( Fata Nidaul Khasanah, S.Kom., M.Eng )

NIDN : 0327059202

Ketua Program Studi

  
( Ahmad Fathurrozi, S.E., M.M.S.I )

NIP : 2012786

Program Studi Informatika  
Fakultas Ilmu Komputer  
Universitas Bhayangkara Jakarta Raya  
2024

## LEMBAR PENGESAHAN

Judul Skripsi : Klasterisasi Pegawai PT Jasamarga Tollroad  
Operator Menggunakan Metode K-Means  
Berdasarkan Data Kehadiran

Nama Mahasiswa : Ananda Setyawati

Nomor Pokok Mahasiswa : 202010225243

Program Studi/Fakultas : Informatika/Ilmu Komputer

Tanggal Lulus Ujian : 26 Juli 2024

Skripsi

Jakarta, 02 Agustus 2024  
MENGESAHKAN,

Ketua Tim Penguji : Wowon Priatna, S.T., M.T.I  
NIDN. 0429118007

Penguji I : Aida Fitriyani, S.Kom., M.M.S.I  
NIDN. 0302078508

Penguji II : Dr. Mujiono Sadikin, S.T., M.T  
NIDN. 0406127002

MENGETAHUI,

Ketua Program Studi Informatika

Dekan Fakultas Ilmu Komputer

Ahmad Fathurrozi, S.E., M.M.S.I  
NIP. 2012486

Dr. Dra. Tyastuti Sri Lestari, M.M.  
NIP. 1408206



UNIVERSITAS BHAYANGKARA JAKARTA RAYA  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Ananda Setyawati  
NPM : 202010225243  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Judul Akhir : Klasterisasi Pegawai PT Jasamarga  
Tollroad Operator Menggunakan  
Metode K – Means Berdasarkan Data  
Kehadiran

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Jakarta, 26 Juli 2024  
Ananda Setyawati



## ABSTRAK

**Ananda Setyawati, 202010225243.** Klasterisasi pegawai PT Jasamarga Tollroad Operator menggunakan metode *K-Means* berdasarkan data kehadiran. Bekasi: Fakultas Ilmu Komputer. Universitas Bhayangkara Jakarta Raya. 2024

Tingkat kehadiran pegawai merupakan indikator penting bagi produktivitas dan kinerja perusahaan. Penelitian ini menggunakan data historis kehadiran pegawai untuk mengidentifikasi pola kehadiran dan mengelompokkan pegawai ke dalam beberapa klaster berdasarkan pola tersebut. Metode *K-Means* digunakan untuk klasterisasi dengan langkah-langkah analisis yang meliputi pra-pemrosesan data, pemilihan jumlah klaster optimal, implementasi ke algoritma *K-Means* serta mengevaluasi hasil clustering. Pemilihan jumlah klaster optimal dilakukan menggunakan metode *Silhouette*, yang menunjukkan jumlah klaster optimal adalah 7. Evaluasi klaster menggunakan *Davies-Bouldin Index* menghasilkan nilai sebesar 0.6284960627229907. Evaluasi ini membantu memahami karakteristik masing-masing klaster dan hasil penelitian tersebut memberikan wawasan yang berguna bagi manajemen PT Jasamarga Tollroad Operator dalam mengoptimalkan manajemen kehadiran pegawai. Dengan pemahaman yang lebih baik tentang pola kehadiran pegawai, perusahaan dapat mengambil langkah-langkah efektif untuk meningkatkan produktivitas dan kinerja organisasi.

**Kata Kunci :** *Klasterisasi, Kehadiran Pegawai, K-Means, Metode Silhouette, Davies – Bouldin Index, dan Manajemen Produktivitas.*

## ABSTRACT

**Ananda Setyawati, 202010225243.** *Clustering of PT Jasamarga Tollroad Operator employees using the K-Means method based on attendance data. Bekasi: Faculty of Computer Science. Bhayangkara University, Greater Jakarta. 2024*

*Employee attendance is an important indicator of a company's productivity and performance. This research uses historical employee attendance data to identify attendance patterns and group employees into clusters based on these patterns. K-Means method is used for clustering with analysis steps that include data pre-processing, selection of optimal number of clusters, implementation to K-Means algorithm as well as evaluating clustering results. Selection of the optimal number of clusters was done using the Silhouette method, which showed the optimal number of clusters was 7. Evaluation of the clusters using the Davies-Bouldin Index resulted in a value of 0.6284960627229907. This evaluation helps understand the characteristics of each cluster and the results of the study provide useful insights for the management of PT Jasamarga Tollroad Operator in optimizing employee attendance management. With a better understanding of employee attendance patterns, the company can take effective measures to improve productivity and organizational performance.*

**Keywords :** *Clustering, Employee Attendance, K-Means, Silhouette Method, Davies - Bouldin Index, and Productivity Management.*

## LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIK

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Ananda Setyawati  
NPM : 202010225243  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty-Free Right*), atas karya ilmiah saya yang berjudul :

Klasterisasi Pegawai PT Jasamarga Tollroad Operator Menggunakan Metode K – Means Berdasarkan Data Kehadiran

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta  
Pada tanggal : 26 Juli 2024  
Yang Menyatakan



Ananda Setyawati

## KATA PENGANTAR

Segala puji bagi Allah SWT, yang telah melimpahkan rahmat dan hidayah nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **“Klasterisasi Pegawai PT Jasamarga Tollroad Operator Menggunakan Metode K-Means Berdasarkan Data Kehadiran”**. Skripsi ini ditulis sebagai prasyarat untuk memperoleh gelar sarjana pada program studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya. Saya mengucapkan terima kasih kepada semua pihak yang telah mendukung dan mendoakan selama proses penyusunan skripsi ini.

1. Bapak Irjen Pol (Purn) Dr. Drs. Bambang Karsono, S.H., M.M. selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Dr. Dra. Tyastuti Sri Lestari, M.M. selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I. selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
4. Bapak Dr. Mujiono Sadikin, S.T., M.T. selaku Dosen Pembimbing I dalam penyusunan skripsi yang selalu memberikan arahan dan bimbingan dengan sangat baik.
5. Ibu Fata Nidaul Khasanah, S.Kom., M.Eng selaku Dosen Pembimbing II dalam penyusunan skripsi yang selalu memberikan arahan dan bimbingan dengan sangat baik.
6. Bapak Mukhlis S.Kom., M.T selaku dosen Pembimbing Akademik

Universitas Bhayangkara Jakarta Raya Program Studi Informatika  
Fakultas Ilmu Komputer.

7. Kepada Orang Tua saya yang telah memberikan dukungan moral maupun material, mendoakan serta memberikan semangat kepada penulis selama penulisan skripsi berlangsung.
8. Detha Ardiantama Tarigan selaku support sistem saya yang selalu memberikan semangat dan membantu selama penulisan skripsi berlangsung.
9. Teman – teman Angkatan 2020 Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya yang menjadi tempat bertukar ilmu dan bertukar informasi akademik maupun non akademik serta berjuang bersama untuk menyelesaikan perkuliahan hingga mendapatkan gelar sarjana.

Penyusunan skripsi ini masih terdapat kekurangan dan masih jauh dari sempurna, oleh karena itu saya dengan hati terbuka menerima masukan baik berupa kritik dan saran membangun untuk memperbaiki kekurangan yang ada. Semoga skripsi ini dapat bermanfaat baik bagi saya pribadi maupun bagi orang lain yang membacanya.

Jakarta, 26 Juli 2024

Hormat Saya,



Ananda Setyawati

## DAFTAR ISI

<b>LEMBAR PERSETUJUAN PEMBIMBING .....</b>	<b>i</b>
<b>LEMBAR PENGESAHAN .....</b>	<b>ii</b>
<b>LEMBAR PERNYATAAN BUKAN PLAGIASI .....</b>	<b>iii</b>
<b>ABSTRAK .....</b>	<b>iv</b>
<b>ABSTRACT .....</b>	<b>v</b>
<b>LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI .....</b>	<b>vi</b>
<b>KATA PENGANTAR .....</b>	<b>vii</b>
<b>DAFTAR ISI .....</b>	<b>ix</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR .....</b>	<b>xiii</b>
<b>BAB I PENDAHULUAN .....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	2
1.3 Rumusan Masalah .....	3
1.4 Tujuan Penelitian .....	4
1.5 Manfaat Penelitian .....	4
1.6 Batasan Masalah .....	5
1.7 Sistematika Penulisan .....	5
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>7</b>
2.1 Penelitian Terdahulu .....	7
2.2 Pengertian Data Mining .....	9
2.3 Pengertian Dataset .....	10
2.4 Pengertian Klasterisasi .....	11

2.5	Pengertian Kehadiran atau Absensi .....	11
2.6	Pengertian Data Histori.....	12
2.7	Pengertian Algoritma K-Means .....	12
2.8	Pengertian Metode Silhouette .....	14
2.9	Pengertian Davies – Bouldin Index .....	15
2.10	Pengertian Python .....	15
2.11	Pengertian Google Colab .....	16
2.12	Knowledge Discovery in Database (KDD) .....	16
<b>BAB III METODOLOGI PENELITIAN .....</b>		<b>19</b>
3.1	Lokasi Penelitian.....	19
3.2	Kerangka Penelitian .....	19
3.3	Tahapan Penelitian .....	19
3.3.1	<i>Selection</i> Data .....	19
3.3.2	<i>Preprocessing</i> Data.....	20
3.3.3	<i>Transformation</i> Data.....	21
3.3.4	<i>Data Mining</i> .....	21
3.3.5	<i>Evaluasi dan Interpretation</i> .....	22
3.4	Jenis dan Pendekatan Penelitian .....	23
3.5	Rancangan Jadwal Penelitian.....	24
3.6	Penelitian yang Sudah Dilakukan .....	25
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>27</b>
4.1	<i>Data Selection</i> .....	27
4.2	<i>Preprocessing</i> Data.....	27
4.3	<i>Transformation</i> Data.....	31
4.4	<i>Data Mining</i> .....	31

4.5	Tahapan Hasil .....	47
4.5.1	Visualisasi Distribusi Data .....	47
4.5.2	Hasil Distribusi K – Means .....	50
4.5.3	Hasil Davies-Bouldin Index .....	52
<b>BAB V PENUTUP .....</b>		<b>54</b>
5.1	Kesimpulan .....	54
5.2	Saran .....	55
<b>DAFTAR PUSTAKA .....</b>		<b>57</b>
<b>LAMPIRAN .....</b>		<b>60</b>



## DAFTAR TABEL

<b>Tabel 2. 1</b> Penelitian Terdahulu.....	7
<b>Tabel 3. 1</b> Jadwal Penelitian.....	24
<b>Tabel 3. 2</b> Data Mentah .....	25
<b>Tabel 3. 3</b> Data yang sudah di Cleansing .....	26
<b>Tabel 4. 1</b> Hasil Missing Values .....	28
<b>Tabel 4. 2</b> Hasil Preprocessing .....	29
<b>Tabel 4. 3</b> Tahapan Encode Label Manual.....	30
<b>Tabel 4. 4</b> Centroid Awal .....	33
<b>Tabel 4. 5</b> Hasil Persamaan Jarak Euclidean.....	34
<b>Tabel 4. 6</b> Centroid Baru .....	35
<b>Tabel 4. 7</b> Hasil K - Means.....	36
<b>Tabel 4. 8</b> Data Contoh Untuk Pengujian Model .....	37
<b>Tabel 4. 9</b> Perhitungan Jarak Iterasi 1 .....	41
<b>Tabel 4. 10</b> Centroid Baru dari Iterasi 1 .....	42
<b>Tabel 4. 11</b> Perhitungan Jarak Iterasi 2 .....	43
<b>Tabel 4. 12</b> Centroid Baru dari Iterasi 2 .....	43
<b>Tabel 4. 13</b> Perhitungan Jarak Iterasi 3 .....	45
<b>Tabel 4. 14</b> Centroid Baru dari Iterasi 3 .....	45
<b>Tabel 4. 15</b> Distribusi Cluster K - Means.....	51

## DAFTAR GAMBAR

<b>Gambar 2. 1</b>	Contoh Perhitungan Persamaan 2.1 .....	14
<b>Gambar 2. 2</b>	Contoh Perhitungan Persamaan 2.2 .....	14
<b>Gambar 2. 3</b>	Tahapan Knowledge Discovery in Database .....	17
<b>Gambar 3. 1</b>	Tahapan Modelling K - Means .....	22
<b>Gambar 3. 2</b>	Bagan Kerangka Penelitian .....	23
<b>Gambar 4. 1</b>	Paket Modul yang Diperlukan .....	28
<b>Gambar 4. 2</b>	Source Code Encode Label Kategorikal.....	30
<b>Gambar 4. 3</b>	Source Code Pemilihan Fitur Relevan.....	31
<b>Gambar 4. 4</b>	Source Code Standardisasi Fitur.....	31
<b>Gambar 4. 5</b>	Source Code Silhouette Method .....	32
<b>Gambar 4. 6</b>	Hasil Silhouette Method .....	32
<b>Gambar 4. 7</b>	Source Code Menentukan Centroid Awal .....	33
<b>Gambar 4. 8</b>	Source Code Persamaan Jarak Euclidean .....	34
<b>Gambar 4. 9</b>	Source Code Klasterisasi .....	35
<b>Gambar 4. 10</b>	Hasil Visualisasi Data K - Means.....	47
<b>Gambar 4. 11</b>	Visualisasi data Nama.....	47
<b>Gambar 4. 12</b>	Visualisasi data nama_jabatan .....	48
<b>Gambar 4. 13</b>	Visualisasi data status kehadiran .....	48
<b>Gambar 4. 14</b>	Visualisasi data late_reason .....	49
<b>Gambar 4. 15</b>	Visualisasi data location_work .....	49
<b>Gambar 4. 16</b>	Visualisasi data Kmeans_Cluster .....	50
<b>Gambar 4. 17</b>	Source Code Distribusi Data Algoritma .....	51
<b>Gambar 4. 18</b>	Source Code Evaluasi dengan DBI.....	53

## DAFTAR LAMPIRAN

<b>Lampiran 1.</b> Cek Plagiarisme .....	61
<b>Lampiran 2.</b> Biodata Mahasiswa .....	62
<b>Lampiran 3.</b> Kartu Bimbingan Skripsi .....	63
<b>Lampiran 4.</b> Surat Balasan Izin Riset .....	66
<b>Lampiran 5.</b> Surat Rekomendasi Pembimbing .....	67
<b>Lampiran 6.</b> Data Mentah Kehadiran Pegawai .....	68

