

**KLASTERISASI DATA MAHASISWA UNDIKSHA
MENGUNAKAN ALGORITMA K-MEANS DAN
DBSCAN SERTA ANALISIS PELUANG KELULUSAN
MAHASISWA PADA SUATU *CLUSTER***



**OLEH
GUSTI AYU GITA MULYA SARI
NIM 1915101011**

**PROGRAM STUDI ILMU KOMPUTER
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN KEJURUAN
UNIVERSITAS PENDIDIKAN GANESHA
SINGARAJA**

2023

**KLASTERISASI DATA MAHASISWA UNDIKSHA
MENGUNAKAN ALGORITMA K-MEANS DAN
DBSCAN SERTA ANALISIS PELUANG KELULUSAN
MAHASISWA PADA SUATU *CLUSTER***

SKRIPSI

Diajukan Kepada

Universitas Pendidikan Ganesha

Untuk memenuhi salah satu persyaratan Dalam Menyelesaikan

Program Sarjana Ilmu Komputer

Oleh

GUSTI AYU GITA MULYA SARI

NIM 1915101011

PROGRAM STUDI ILMU KOMPUTER

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK DAN KEJURUAN

UNIVERSITAS PENDIDIKAN GANESHA

SINGARAJA

2023

SKRIPSI

**DIAJUKAN UNTUK MELENGKAPI TUGAS
DAN MEMENUHI SYARAT-SYARAT UNTUK
MENCAPAI GELAR SARJANA KOMPUTER**

Menyetujui

Pembimbing I,



I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs.
NIP. 198910262019031004

Pembimbing II,



I Ketut Purnamawan, S.Kom., M.Kom.
NIP.197905112006041004

Skripsi oleh Gusti Ayu Gita Mulya Sari
telah dipertahankan di depan dewan penguji
pada tanggal 10 Oktober 2023

Dewan Penguji,



I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs.
NIP. 198910262019031004

(Ketua)



I Ketut Purnamawan, S.Kom., M.Kom.
NIP. 197905112006041004

(Anggota)



Putu Hendra Suputra, S.Kom., M.Cs.
NIP. 198212222006041001

(Anggota)



Dr. Luh Joni Erayati Dewi, S.T., M.Pd.
NIP. 197606252001122001

(Anggota)

Diterima oleh Panitia Ujian Fakultas Teknik dan Kejuruan
Universitas Pendidikan Ganesha
Guna memenuhi syarat-syarat untuk mencapai gelar sarjana

Pada:

Hari : Selasa.....
Tanggal : 10 - oktober.....

Mengetahui,

Ketua Ujian,



Made Windu Antara Kesiman, S.T., M.Sc., Ph.D.
NIP. 198211112008121001

Sekretaris Ujian,

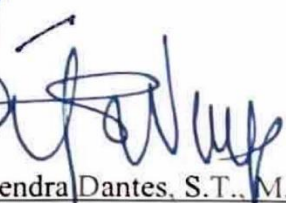


I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs.
NIP. 198910262019031004

Mengesahkan,

Dekan Fakultas Teknik dan Kejuruan




Dr. Kadek Rihendra Dantes, S.T., M.T.
NIP. 197912012006041001

PERNYATAAN

Dengan ini saya menyatakan bahwa karya tulis yang berjudul **“Klasterisasi Data Mahasiswa Undiksha Menggunakan Algoritma K-Means dan DBSCAN serta Analisis Peluang Kelulusan Mahasiswa pada Suatu Cluster”** beserta keseluruhan isinya adalah benar-benar hasil karya sendiri dan tidak melakukan penjiplakan dan pengutipan dengan cara yang tidak sesuai dengan etika yang berlaku pada buku pedoman skripsi. Jika dikemudian hari ditemukan adanya pelanggaran saya siap menanggung resiko atau sanksi yang dijatuhkan kepada saya.

Singaraja, 10 Oktober 2023

Yang membuat pernyataan,



Gusti Ayu Gita Mulya Sari

NIM. 1915101011



MOTTO

"Be high in the sky but still have your feet on the ground"

PRAKATA

Puji syukur peneliti panjatkan kehadapan Ida Sang Hyang Widhi Wasa/Tuhan Yang Maha Esa karena berkat rahmat-Nya lah sehingga memungkinkan peneliti menyelesaikan skripsi dengan judul **“Klasterisasi Data Mahasiswa Undiksha Menggunakan Algoritma K-Means dan DBSCAN serta Analisis Peluang Kelulusan Mahasiswa pada Suatu Cluster”** sebagai salah satu syarat untuk memperoleh gelar sarjana Ilmu Komputer di Universitas Pendidikan Ganesha. Dalam menyelesaikan skripsi ini, peneliti banyak mendapatkan motivasi dan bantuan baik berupa moral maupun material dari berbagai pihak. Untuk itu dalam kesempatan ini peneliti mengucapkan rasa terima kasih kepada :

1. Bapak Dr. Kadek Rihendra Dantes, S.T., M.T., selaku Dekan Fakultas Teknik dan Kejuruan atas motivasi dan fasilitas yang diberikan sehingga peneliti dapat menyelesaikan studi sesuai dengan rencana.
2. Bapak Putu Hendra Suputra, S.Kom., M.Cs., selaku Ketua Jurusan Teknik Informatika atas motivasi dan fasilitas yang diberikan sehingga peneliti dapat menyelesaikan studi ini dengan tepat waktu.
3. Bapak I Nyoman Saputra Wahyu Wijaya, S.Kom., M.Cs., selaku Koordinator Prodi Ilmu Komputer atas motivasi dan fasilitas yang diberikan kepada peneliti dalam penyelesaian skripsi ini, serta selaku pembimbing I yang telah memberikan banyak motivasi, bimbingan, arahan, petunjuk skripsi, dengan penuh kesabaran kepada peneliti dalam menyelesaikan skripsi ini sehingga peneliti dapat menyelesaikan skripsi dengan tepat waktu.
4. Bapak I Ketut Purnamawan, S.Kom., M.Kom., selaku pembimbing II yang telah banyak memberikan motivasi, semangat, bimbingan, arahan dengan penuh kesabaran kepada peneliti dalam menyelesaikan skripsi ini sehingga peneliti dapat menyelesaikan skripsi dengan tepat waktu.
5. Ibu Dr. Luh Joni Erawati Dewi, S.T., M.Pd., selaku penguji II yang telah banyak memberikan masukan dan arahan dengan penuh kesabaran kepada

peneliti dalam menyelesaikan skripsi ini sehingga peneliti dapat menyelesaikan skripsi dengan tepat waktu.

6. Seluruh staf dosen di lingkungan Jurusan Teknik Informatika/ Program Studi Ilmu Komputer yang telah memberikan ilmu, pengalaman, motivasi, serta semangat selama menjalani studi di Universitas Pendidikan Ganesha
7. Untuk teman seperjuangan peneliti Prima, Risma, Subiksa, Agung, Sindu, Andika, Soma dan teman-teman lainnya yang tidak bisa peneliti sebutkan satu persatu, terima kasih telah memberikan semangat, dukungan, bantuan sehingga peneliti sampai di titik ini dan dapat menyelesaikan skripsi dengan tepat waktu.

Semoga Ida Sang Hyang Widhi Wasa/Tuhan Yang Maha Esa senantiasa memberikan karunianya dan membantu peneliti dalam menyelesaikan skripsi ini. Peneliti menyadari sepenuhnya bahwa apa yang disajikan dalam skripsi ini masih jauh dari kata sempurna mengingat keterbatasan peneliti. Peneliti mengharapkan segala masukan, kritik dan saran yang bersifat membangun dari berbagai pihak. Peneliti berharap apa yang dibuat oleh peneliti dapat bermanfaat dan berharga bagi kita semua, khususnya untuk mengembangkan dunia pendidikan

Singaraja, 10 Oktober 2023
Peneliti

DAFTAR ISI

PRAKATA	i
ABSTRAK.....	iii
DAFTAR ISI	v
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian.....	4
1.4 Batasan Masalah.....	5
1.5 Manfaat Penelitian.....	6
BAB II LANDASAN TEORI.....	8
2.1 Penelitian Terkait	8
2.2 <i>Data mining</i>	14
2.3 <i>Clustering</i>	17
2.4 K-Means	20
2.5 DBSCAN.....	22
2.6 <i>Silhouette coefficient</i>	23
2.7 <i>Interquartile Range (IQR)</i>	25
2.8 <i>Min-max Normalization</i>	27
2.9 <i>Elbow Method</i>	28
2.10 <i>Principal Component Analysis (PCA)</i>	29
2.11 <i>Cluster Variance</i>	31
2.12 Probabilitas.....	32
2.13 <i>Softmax</i>	33
BAB III METODOLOGI PENELITIAN	35
3.1 Kerangka Penelitian	35

3.2 Studi Literatur	36
3.3 Pengumpulan Data	36
3.4 Deskripsi <i>Dataset</i>	37
3.5 <i>Data Preprocessing</i>	38
3.6 K-Means <i>Clustering</i>	40
3.7 DBSCAN <i>Clustering</i>	42
3.8 Visualisasi dan Interpretasi Cluster	45
3.9 <i>Cluster Validation</i>	45
BAB IV HASIL DAN PEMBAHASAN	47
4.1 <i>Dataset Preprocessing</i>	47
4.1.1 <i>Data Cleaning</i>	47
4.1.2 <i>Data Integration</i>	53
4.1.3 <i>Data Transformation</i>	56
4.1.4 <i>Outliers Removing</i>	57
4.1.5 <i>Data Normalization</i>	60
4.1.6 <i>Feature Selection</i>	67
4.1.7 <i>Data Labeling</i>	71
4.2 Proses <i>Clustering</i>	72
4.2.1 K-Means <i>Clustering</i>	72
4.2.1 DBSCAN <i>Clustering</i>	80
4.3 Interpretasi <i>Cluster</i>	86
4.3.1 Denormalisasi Data <i>Centroid</i> K-Means	87
4.3.2 Interpretasi Hasil <i>Cluster</i> K-Means	88
4.3.3 Interpretasi Hasil <i>Cluster</i> DBSCAN	94
4.3.4 Analisis <i>Cluster Variance</i>	101
4.3.5 Analisis Probabilitas Kelulusan Mahasiswa Peluang Sederhana ...	103
4.3.6 Analisis Probabilitas Kelulusan Mahasiswa Metode <i>Softmax</i>	105
4.4 <i>Cluster Validation</i>	111
BAB V PENUTUP	113
5.1 Kesimpulan	113

5.2 Saran.....	115
DAFTAR PUSTAKA.....	116
LAMPIRAN	120
RIWAYAT HIDUP	130



DAFTAR TABEL

Tabel 2.1 Rentang Nilai <i>Silhouette Coefficient</i>	24
Tabel 3.1 Deskripsi <i>Dataset</i> Mahasiswa.....	37
Tabel 4.1 Contoh Data Mentah Non-Akademik dari UPT TIK Undiksha	48
Tabel 4.2 Contoh Data Mentah Akademik dari UPT TIK Undiksha.....	50
Tabel 4.3 Hasil Eliminasi Data Tabel 4.1	51
Tabel 4.4 Hasil Perubahan Struktur Tabel Data Tabel 4.2	52
Tabel 4.5 Hasil dari integrasi Tabel 4.3 dan Tabel 4.4	54
Tabel 4.6 Hasil <i>One-Hot Encoding</i> Fitur Status Beasiswa dan Kota Asal	57
Tabel 4.7 Data Sebelum Normalisasi.....	61
Tabel 4.8 Data Setelah Normalisasi	64
Tabel 4.9 Sampel Data dari Hasil Implementasi PCA.....	70
Tabel 4.10 Percobaan Nilai <i>Random Seed</i>	77
Tabel 4.11 <i>Initial Random Centroid</i> Terbaik.....	77
Tabel 4.12 K-Means <i>Centroids</i> Setelah Denormalisasi Data.....	87
Tabel 4.13 <i>Centroids</i> Akhir K-Means.....	89
Tabel 4.14 Sampel Data <i>Cluster Noise</i>	95
Tabel 4.15 Rata-Rata <i>Features</i> pada Setiap <i>Cluster</i>	97
Tabel 4.16 Analisis Probabilitas Menggunakan Peluang Sederhana.....	103
Tabel 4.17 Analisis Probabilitas Menggunakan Perhitungan <i>Softmax</i>	107
Tabel 4.18 Hasil <i>Silhouette Coefficient</i> Metode K-Means dan DBSCAN	111

DAFTAR GAMBAR

Gambar 2.1 Serangkaian Proses KDD	15
Gambar 2.2 <i>Cluster</i> yang Terbentuk dari Proses <i>Clustering</i>	17
Gambar 2.3 <i>Hierarchical Clustering</i>	18
Gambar 2.4 <i>Partitional Clustering</i>	19
Gambar 2.5 <i>Density-based Clustering</i>	20
Gambar 2.6 <i>Principal Component Analysis</i>	26
Gambar 2.7 Grafik Metode <i>Elbow</i>	29
Gambar 3.1 <i>Framework</i> Penelitian	35
Gambar 3.2 Tahap <i>Data Preprocessing</i>	39
Gambar 3.3 Alur Implementasi Algoritma K-Means	41
Gambar 3.4 Alur Implementasi algoritma DBSCAN	44
Gambar 3.5 Alur implementasi <i>Silhouette Coefficient</i>	46
Gambar 4.1 Visualisasi Seluruh Fitur pada Data Mahasiswa	58
Gambar 4.2 Visualisasi Fitur penghasilan_ortu dengan Boxplot	59
Gambar 4.3 Visualisasi Hasil Penghapusan Outliers	60
Gambar 4.4 Visualisasi <i>Silhouette Coefficient</i> K-Means tanpa PCA	68
Gambar 4.5 Visualisasi <i>Silhouette Coefficient</i> DBSCAN tanpa PCA	68
Gambar 4.6 <i>Variance</i> Data pada <i>Principal Component</i>	70
Gambar 4.7 Hasil Metode <i>Elbow</i>	73
Gambar 4.8 Algoritma Fungsi <i>kmeans_clustering</i>	75
Gambar 4.9 Visualisasi Hasil <i>Cluster</i> dengan K-Means	78
Gambar 4.10 Informasi <i>Centroids</i> Akhir <i>Cluster</i>	79
Gambar 4.11 Sampel Hasil K-Means <i>Clustering</i>	80
Gambar 4.12 Algoritma Fungsi <i>dbscan_clustering</i>	82
Gambar 4.13 Visualisasi Hasil KNN 4-dist	84
Gambar 4.14 Visualisasi Hasil <i>Cluster</i> dengan DBSCAN	85
Gambar 4.15 Sampel Hasil <i>Cluster</i> dengan DBSCAN	86
Gambar 4.16 Perbandingan IPS Mahasiswa Berdasarkan Hasil <i>Cluster</i>	92
Gambar 4.17 Jumlah Anggota <i>Cluster</i> K-Means	93
Gambar 4.18 Jumlah Anggota <i>Cluster</i> DBSCAN	94

Gambar 4.19 Perbandingan IPS Mahasiswa Berdasarkan Hasil *Cluster*..... 100
Gambar 4.20 *Cluster Variance* 102



DAFTAR LAMPIRAN

Lampiran 1. <i>Source Code</i> Program.....	121
---	-----

