



UNIVERSITAS PERTAHANAN RI

**KOMPARATIF KLASTERISASI ALGORITMA *K-MEANS* DAN
GAUSSIAN MIXTURE MODELS UNTUK KESIAPAN
SEKOLAH PERWIRA PRAJURIT KARIER TENTARA
NASIONAL INDONESIA (SEPA PK TNI) DI UNIVERSITAS
PERTAHANAN REPUBLIK INDONESIA**

Made Hayu Aretha Adira Pangastuti

320200401012

**Skripsi yang Ditulis untuk Memenuhi Sebagian Persyaratan dalam
Mendapatkan Gelar Sarjana**

**FAKULTAS SAINS DAN TEKNOLOGI PERTAHANAN
PROGRAM STUDI INFORMATIKA**

BOGOR 2024

LEMBAR PERSETUJUAN SKRIPSI

Nama : Made Hayu Aretha Adira Pangastuti
NIM : 320200401012
Program Studi : Informatika
Judul Skripsi : Komparatif Klasterisasi Algoritma *K-Means* dan
Gaussian Mixture Models untuk Kesiapan Sekolah
Perwira Prajurit Karier Tentara Nasional Indonesia
(SEPA PK TNI) di Universitas Pertahanan RI

Pembimbing I



Nadiza Lediwara, S.T., M.Eng

Penata Muda Tk.I III/b

NIP. 198802092022032001

Tanggal: 23 - 07 - 2024

Pembimbing II



Aditya Adiprabowo, S.T., M.Sc

Penata Muda Tk.I III/b

NIP. 199603212022031005

Tanggal: 23 - 07 - 2024

Mengetahui,

Kepala program Studi

Informatika




Adam Mardamsyah, M.Han

Kolonel Inf NRP. 11940019450871

Tanggal: 23 - 07 - 2024

Dekan

Fakultas Sains dan Teknologi Pertahanan








Prof. Dr. Ir. Muhamad Asvial, M.Eng

Pembina Utama Muda IV/c

Tanggal: 25-07-2024

LEMBAR PENGESAHAN SKRIPSI

Nama : Made Hayu Aretha Adira Pangastuti
 NIM : 320200401012
 Program Studi : Informatika
 Judul Skripsi : Komparatif Klasterisasi Algoritma *K-Means* dan *Gaussian Mixture Models* untuk Kesiapan Sekolah Perwira Prajurit Karier Tentara Nasional Indonesia (SEPA PK TNI) di Universitas Pertahanan RI

No.	Nama	Tanda Tangan	Tanggal
1.	Dosen Pembimbing I: Nadiza Lediwara, S.T., M.Eng Penata Muda Tk.I III/b NIP. 198802092022032001		23/2024 /07
2.	Dosen Pembimbing II: Aditya Adiprabowo, S.T., M.Sc Penata Muda Tk.I III/b NIP. 199603212022031005		23/2024 /07
3.	Dosen Penguji I: Jonson Manurung, S.T., M.Kom PPPK Golongan X NIP 198003232024211005		23 Juli 2024
4.	Dosen Penguji II: Anindito, S.Kom., S.S., MTI., CHFI NIDN. 4719057801		19/24 /7
5.	Dosen Penguji III: Dr. Hondor Saragih, S.T., Msi (Han) Letkol Laut (KH) NRP. 14633/P		18/24 /07

PERNYATAAN ORISINALITAS

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya atau bagian karya yang pernah diajukan untuk memperoleh gelar kesarjanaan jenjang apapun di suatu Perguruan Tinggi; dan sepanjang sepengetahuan saya juga tidak terdapat istilah, frasa, kalimat, paragraf, subbab atau bab dari karya yang pernah ditulis atau diterbitkan; kecuali yang secara tertulis diajukan dalam naskah ini dan disebutkan dalam Daftar Referensi.

Apabila di kemudian hari terbukti bahwa terdapat plagiat dalam skripsi ini, saya bersedia menerima sanksi sesuai ketentuan peraturan / undang-undang yang berlaku.

Bogor, 25 Juli 2024



Made Hayu Aretha Adira P

**LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH
UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Pertahanan Republik Indonesia, saya yang bertanda tangan di bawah ini:

Nama : Made Hayu Aretha Adira Pangastuti
NIM : 320200401012
Program Studi : Informatika
Fakultas : Fakultas Sains dan Teknologi Pertahanan
Jenis Karya : Skripsi

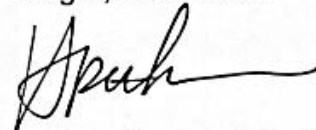
demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pertahanan Republik Indonesia **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty Free Right*)** atas karya ilmiah saya yang berjudul:

PENGEMBANGAN SISTEM INFORMASI MANAJEMEN RUANGAN DI UNIVERSITAS PERTAHANAN RI.

berserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Pertahanan Republik Indonesia berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik Hak Cipta/Karya Intelektual dari skripsi ini.

Demikian pernyataan ini saya buat dengan kesadaran penuh tanpa paksaan dari pihak manapun.

Bogor, 25 Juli 2024



Made Hayu Aretha Adira P

NIM. 320200401012

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Komparatif Klasterisasi Algoritma K-Means dan Gaussian Mixture Models untuk Kesiapan Sekolah Perwira Prajurit Karier Tentara Nasional Indonesia (SEPA PK TNI) di Universitas Pertahanan Republik Indonesia” sesuai dengan waktu yang telah direncanakan. Skripsi ini disusun guna memenuhi sebagian persyaratan dalam memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Sains dan Teknologi Pertahanan di Universitas Pertahanan Republik Indonesia (Unhan RI).

Penyusunan dan penyelesaian skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan berbagai pihak baik secara langsung maupun tidak langsung. Oleh karena itu pada kesempatan ini dengan kerendahan dan ketulusan hati penulis ingin menyampaikan rasa terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. Penata Muda Tk.I III/b Nadiza Lediwara, S.T., M.Eng, selaku Dosen Pembimbing I, yang telah memberikan bimbingan, arahan, dan dukungan yang sangat berarti selama proses penyusunan skripsi ini;
2. Penata Muda Tk.I III/b Aditya Adiprabowo, S.T., M.Sc, selaku Pembimbing II, yang telah memberikan masukan, saran, dan kritik yang konstruktif untuk kesempurnaan skripsi ini;
3. Prof. Dr. Ir. Muhammad Asvial, M. Eng., selaku Dekan Fakultas Sains dan Teknologi Pertahanan (FSTP) Unhan RI;
4. Kepala Program Studi Informatika Unhan RI, Kolonel Inf Adam Mardamsyah, M. Han yang telah memberikan izin dan dukungan untuk pelaksanaan penelitian ini;
5. Ibu Ajeng Hidayati, S.Kom., M.Kom selaku Dosen Pembimbing Akademik yang telah mengarahkan penulis sedari awal pendidikan hingga sekarang;

6. Segenap Dosen dan Staf Program Studi Informatika FSTP yang telah memberikan ilmu, bantuan, dan fasilitas selama masa studi penulis. Terima kasih juga kepada Pelatih dan Pengasuh serta seluruh Civitas Akademika Universitas Pertahanan Republik Indonesia yang telah memberikan bekal bimbingan, arahan, dan nasehat selama penulis menempuh pendidikan di Universitas Pertahanan Republik Indonesia;
7. Kepada keluarga tercinta, Ayahanda Wayan Nuriada, Ibunda Sri Wulansari, dan Abang Gede Satria Aji Pangestu yang telah memberikan dukungan moral, materiil, dan spiritual yang tiada henti serta segala pengorbanan, kasih sayang, doa yang senantiasa mengiringi langkah penulis. Tanpa cinta, dorongan, dan ketulusan hati mereka, penulis tidak akan mampu mencapai titik ini;
8. Kepada 24 teman Program Studi Informatika *cohort* 1 terutama untuk keluarga asuh Kosaraju yang tidak bisa penulis sebutkan namanya satu persatu, yang telah bersama kurang lebih selama 4 tahun menjalani setiap momen baik dalam menghadapi tantangan maupun merayakan keberhasilan kegiatan akademik dan non-akademik serta memberikan kenangan berharga dan pengalaman yang tak terlupakan;
9. Saudara asuh bimbingan Penata Muda Tk.I III/b Nadiza Lediwara, S.T., M.Eng dan Penata Muda Tk.I III/b Aditya Adiprabowo, S.T., M.Sc yang selalu memberikan motivasi, inspirasi, dan berbagi informasi kepada penulis dalam penyelesaian skripsi;
10. Rekan-rekan Tongkrongan Kami dan Calon Orang Sukses yang tidak bisa penulis sebutkan namanya satu persatu, Saudari Riyanti Putri, Saudara Rendy Cahya Aditya, dan Farrel Tajusalatin Widjanarko yang telah memberikan

dukungan secara moral, memotivasi, dan memberikan doa dalam penyelesaian skripsi ini;

11. Semua pihak yang tidak dapat penulis sebutkan satu per satu.

Akhir kata, izinkan penulis mengakhiri tulisan ini dengan mengutip sloka Veda yang meneguhkan hati, pikiran, dan iman penulis dalam penelitian ini hingga akhir yang terambil dari; **Bhagavad Gita (18.62) Berserahlah sepenuhnya kepada-Nya, Sang Pengendali Tertinggi. Atas karunia-Nya, engkau akan mencapai kedamaian transendental dan keabadian yang tertinggi.**

Semoga kebaikan yang diberikan oleh semua pihak mendapatkan balasan yang berlipat ganda dari Tuhan Yang Maha Esa. Penulis menyadari skripsi ini masih jauh dari sempurna dan sangat mengharapkan kritik serta saran yang membangun untuk perbaikan di masa mendatang. Semoga skripsi ini bermanfaat bagi pembaca dan menjadi referensi dalam pengembangan ilmu, khususnya di bidang klasterisasi data dan analisis kesiapan personil. Penulis berharap hasil penelitian ini dapat mendukung program peningkatan kualitas sumber daya manusia dan membantu pengambilan keputusan strategis di Universitas Pertahanan Republik Indonesia.

Bogor, 25 Juli 2024



Made Hayu Aretha Adira P

ABSTRAK

KOMPARATIF KLASTERISASI ALGORITMA *K-MEANS* DAN *GAUSSIAN MIXTURE MODELS* UNTUK KESIAPAN SEKOLAH PERWIRA PRAJURIT KARIER TENTARA NASIONAL INDONESIA (SEPA PK TNI) DI UNIVERSITAS PERTAHANAN REPUBLIK INDONESIA

Tujuan dari penelitian ini adalah untuk mengevaluasi seberapa efektif algoritma klasterisasi *K-Means* dan *Gaussian Mixture Models (GMM)* dalam mengukur kesiapan siswa Sekolah Perwira Prajurit Karier TNI di Universitas Pertahanan Republik Indonesia. Indeks Prestasi Kumulatif (IPK) dan nilai kesegaran jasmani adalah beberapa metrik akademik dan fisik yang dikumpulkan. Pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan persebaran (*deployment*) adalah bagian dari metode CRISP-DM (*Cross-Industry Standar Process for Data Mining*) yang digunakan dalam penelitian ini. Berdasarkan indeks validasi *Davies-Bouldin* dan *Calinski-Harabasz*, hasil evaluasi menunjukkan bahwa algoritma *K-Means* menghasilkan klaster yang lebih kompak dan terpisah dengan baik dengan nilai yang lebih tinggi dalam hal separasi dan kepadatan klaster dibandingkan dengan *GMM*. Model klastering yang telah dievaluasi telah siap untuk tahap *deployment*, yang menunjukkan bahwa model siap untuk digunakan dalam tugas-tugas yang dilakukan di Unhan RI mengenai pengelompokan kesiapan dan pembagian personel. Penelitian ini akan membantu menentukan program di Universitas Pertahanan RI yang dapat meningkatkan kualitas sumber daya manusia dan membantu pengambilan keputusan strategis. Diharapkan penerapan model ini akan membantu pengembangan dan pengelolaan sumber daya manusia dalam upaya untuk meningkatkan kesiapan dan efektivitas pertahanan nasional.

Kata kunci : *k-means, gaussian mixture models, davies-bouldin index, calinski-harabasz, CRISP-DM*

ABSTRACT

COMPARATIVE CLUSTERING OF K-MEANS ALGORITHM AND GAUSSIAN MIXTURE MODELS FOR READINESS OF THE SCHOOL OF CAREER SOLDIER OFFICERS (SEPA PK TNI) AT THE DEFENSE UNIVERSITY OF THE REPUBLIC OF INDONESIA TO SUPPORT NATIONAL SECURITY

The purpose of this study is to evaluate how effective the K-Means clustering algorithm and Gaussian Mixture Models (GMM) are in measuring the readiness of TNI's School of Career Soldier Officers (SEPA PK) students at the Republic of Indonesia Defense University. Grade Point Average (GPA) and physical fitness scores are some of the academic and physical metrics collected. Business understanding, data understanding, data preparation, modeling, evaluation, and deployment are part of the CRISP-DM (Cross-Industry Standar Process for Data Mining) method used in this study. Based on the Davies-Bouldin and Calinski-Harabasz validation indices, the evaluation results show that the K-Means algorithm produces more compact and well-separated clusters with higher values in terms of separation and cluster density compared to GMM. The evaluated clustering model is ready for the deployment stage, which indicates that the model is ready to be used in the tasks performed at the Indonesia Defense University regarding readiness clustering and personnel assignment. This research will help determine programs at the Indonesia Defense University that can improve the quality of human resources and help strategic decision making. It is expected that the application of this model will help the development and management of human resources in an effort to improve the readiness and effectiveness of national defense.

Keywords : *k-means, gaussian mixture models, davies-bouldin index, calinski-harabasz, CRISP-DM*

DAFTAR ISI

LEMBAR PERSETUJUAN SKRIPSI.....	ii
LEMBAR PENGESAHAN SKRIPSI	iii
PERNYATAAN ORISINALITAS.....	iv
LEMBAR PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS	v
KATA PENGANTAR	vi
ABSTRAK	ix
<i>ABSTRACT</i>	x
DAFTAR ISI	xi
DAFTAR GAMBAR	xiv
DAFTAR TABEL	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Landasan Teori.....	7
2.1.1 Komponen Utama.....	7
2.1.1.1 Tentara Nasional Indonesia (TNI).....	8
2.1.1.2 Sekolah Perwira Prajurit Karier TNI.....	10
2.1.2 Data Mining	11
2.1.2.1 CRISP-DM.....	13

2.1.3	Klasterisasi	15
2.1.3.1	Kluster-Means (K-Means)	16
2.1.3.2	Gaussian Mixture Models (GMM)	17
2.1.3.3	Indeks Davies-Bouldin (Davies-Bouldin Index).....	19
2.1.3.4	Indeks Calinski-Harabasz (Calinski-Harabasz Index)	20
2.2	Hasil Penelitian Terdahulu	21
2.3	Kerangka Pemikiran.....	25
BAB III METODOLOGI PENELITIAN		27
3.1	Metode dan Desain Penelitian	27
3.1.1	Pemahaman Bisnis (Business Understanding)	27
3.1.2	Pemahaman Data (Data Understanding)	27
3.1.3	Persiapan Data (Data Preparation).....	30
3.1.4	Pemodelan (Modelling).....	31
3.1.5	Evaluasi (Evaluation)	31
3.1.6	Penyebaran (Deployment).....	32
3.2	Tempat dan Waktu Penelitian	32
1.3	Alat dan Bahan.....	34
3.3.1	Perangkat Keras	34
3.3.2	Perangkat Lunak	34
BAB IV HASIL DAN PEMBAHASAN		35
4.1	Hasil.....	35
4.2	Pembahasan.....	36
4.2.1	Pemahaman Bisnis (Business Understanding).....	36
4.2.2	Pemahaman Data (Data Understanding)	37
4.2.3	Persiapan Data (Data Preparation).....	39

4.2.4	Pemodelan (Modelling)	54
4.2.5	Evaluasi (Evaluation)	64
4.2.6	Penyebaran (Deployment)	73
BAB V KESIMPULAN DAN SARAN		77
5.1	Kesimpulan	77
5.2	Saran	78
DAFTAR PUSTAKA		80
LAMPIRAN.....		85
	Lampiran 1 Surat Permohonan & Disposisi.....	85

DAFTAR GAMBAR

Gambar 2. 1 CRISP-DM: 6 Fase	15
Gambar 2. 2 Kerangka Berpikir	
Gambar 2. 3 CRISP-DM: 6 Fase	15
Gambar 2. 4 Kerangka Berpikir	25
Gambar 3. 1 Contoh Data Akademik Kadet Mahasiswa Cohort 2	28
Gambar 3. 2 Contoh Nilai Evaluasi Garjas Cohort 1 Putra	29
Gambar 3. 3 Contoh Nilai Evaluasi Garjas Cohort 1 Putri	29
Gambar 3. 4 Pemodelan Penelitian	31
Gambar 4. 1 Sebaran data berdasarkan jenis kelamin	38
Gambar 4. 2 Sebaran data berdasarkan program studi	39
Gambar 4. 3 Syntax tahapan pembersihan data	41
Gambar 4. 4 Syntax tahapan label encoding	43
Gambar 4. 5 Syntax tahapan menampilkan nilai Outlier	47
Gambar 4. 6 Hasil identifikasi nilai outlier tiap variabel	48
Gambar 4. 7 Syntax tahapan mengatasi nilai outlier	50
Gambar 4. 8 Hasil implementasi mean pada nilai outlier	51
Gambar 4. 9 Syntax tahapan standarisasi data	52
Gambar 4. 10 Syntax tahapan implementasi metode K-Means	54
Gambar 4. 11 Sebaran anggota cluster hasil metode K-Means	56
Gambar 4. 12 Syntax tahapan visualisasi sebaran cluster hasil K-Means	57
Gambar 4. 13 Visualisasi sebaran cluster pada metode K-Means	58
Gambar 4. 14 Syntax tahapan implementasi metode GMM	59
Gambar 4. 15 Jumlah anggota tiap cluster hasil metode Gaussian Mixture Models	61
Gambar 4. 16 Syntax tahapan visualisasi sebaran cluster hasil GMM	62
Gambar 4. 17 Visualisasi sebaran cluster Gaussian Mixture Models	63
Gambar 4. 18 Syntax tahapan pengukuran metode K-Means dan GMM menggunakan Silhouette score, Calinski-Harabasz Score, Davies-Bouldin Indeks	68
Gambar 4. 19 Syntax tahapan pengukuran inertia atau Elbow	71

Gambar 4. 20 Halaman Utama Situs.....	74
Gambar 4. 21 Halaman Situs yang Berisikan Hasil Clustering	74
Gambar 4. 22 Halaman Situs yang Berisikan Hasil Evaluasi Clustering...	75
Gambar 4. 23 Halaman Penjelasan Mengenai Isi Situs	75

DAFTAR TABEL

Tabel 2. 1 Hasil Penelitian Terdahulu	22
Tabel 3. 1 Jadwal Penelitian.....	33
Tabel 4. 1 Hasil evaluasi kinerja metode K-Means dan Gaussian Mixture Models	35
Tabel 4. 2 Data sampel	40
Tabel 4. 3 Hasil implementasi nilai mean dalam mengatasi missing value.	42
Tabel 4. 4 Hasil implementasi Label encoder.....	44
Tabel 4. 5 Perubahan tipe data hasil standarisasi data	53
Tabel 4. 6 Hasil dari perhitungan.....	64
Tabel 4. 7 Hasil dari perhitungan.....	66
Tabel 4. 8 Hasil evaluasi kinerja metode K-Means dan Gaussian Mixture Models	69

BAB I

PENDAHULUAN

1.1 Latar Belakang

Universitas Pertahanan Republik Indonesia memegang peran vital dalam pendidikan pertahanan nasional. Sebagai lembaga pendidikan tinggi, Unhan RI menyediakan program studi yang bernuansa pertahanan pada level Diploma (D-3), Sarjana (Strata-1), Magister (Strata-2), dan Doktorat (Strata-3). Institusi ini secara operasional pendidikan berada di bawah pengawasan akademik oleh Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, serta secara fungsi teknis pertahanan berada pengawasan fungsional dari Kementerian Pertahanan, sesuai dengan Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 53 Tahun 2022 tentang Organisasi dan Tata Kerja Unhan (Kementerian Pendidikan, 2022). Dengan ini kadet mahasiswa Unhan RI memiliki peran ganda, tidak hanya sebagai kader intelektual bela negara, tetapi juga sebagai komponen pertahanan, yang disiapkan untuk menjadi perwira melalui Sekolah Perwira Prajurit Karier Tentara Nasional Indonesia dengan ilmu pertahanan dari lembaga.

Universitas Pertahanan Republik Indonesia sebagai salah satu Lembaga yang mendapat tugas untuk membentuk Komponen Pertahanan dengan subjek Kadet Mahasiswa dibawah naungan Lembaga Pengembangan Pertahanan Negara (LPPN) dan dipimpin oleh perwira Tentara Nasional Indonesia berpangkat bintang dua dengan binaan dan asuhan langsung dari Komandan Resimen Kadet Mahasiswa atau Danmenkad dalam organisasi Resimen Kadet Mahasiswa (Universitas Pertahanan RI, 2023) Dalam pelaksanaan tersebut Resimen Kadet

Mahasiswa membentuk sikap dan kemampuan dasar keprajuritan dan melatih mental kejuangan. Pelatihan mental kejuangan dilakukan setiap hari mulai dari bangun tidur sampai selesai apel malam. Adapun bentuk pelatihan mental kejuangan diantaranya adalah kegiatan baris berbaris secara rapih dan tertib berangkat dan kembali dari mess menuju tempat belajar. Mematuhi segala aturan dan ketentuan yang diberlakukan dalam kehidupan berasrama di Unhan RI yang awalnya diatur dalam Peraturan Unhan RI Nomor 59 Tahun 2021 Tentang Peraturan Khusus Kadet Mahasiswa Universitas Pertahanan Republik Indonesia (Universitas Pertahanan RI, 2021) lalu mengalami penyesuaian kembali pada Peraturan Unhan RI Nomor 68 Tahun 2023 Tentang Pedoman Kehidupan Kadet Mahasiswa Program Diploma Tiga dan Program Sarjana/Strata Satu Universitas Pertahanan Republik Indonesia (Universitas Pertahanan RI, 2023). Namun, terdapat tantangan signifikan dalam proses pembinaan dan pengasuhan di Resimen Kadet Mahasiswa, terutama karena kurangnya sumber daya manusia (SDM) yang mumpuni. Hal ini berdampak pada kurang maksimalnya bina asuh terlebih pada rendahnya kesadaran kadet mahasiswa dalam mempertahankan kemampuan yang telah diperoleh.

Dalam konteks ini, terdapat kebutuhan mendesak untuk mengevaluasi dan meningkatkan efektivitas program pembentukan komponen pertahanan. Salah satu pendekatan yang dapat diterapkan untuk menganalisis kesiapan kadet mahasiswa menggunakan data yang relevan adalah melalui penerapan metode klusterisasi. Dalam penelitian ini, algoritma *K-Means* dan *Gaussian Mixture Models* dipilih sebagai alat klusterisasi karena keduanya efektif dalam mengelompokkan data dengan pola yang kompleks, seperti kesiapan personil. Penggunaan *K-Means* yang mengandalkan pembentukan pusat klaster dan *Gaussian Mixture Models* yang memodelkan distribusi probabilitas setiap klaster memberikan pendekatan yang holistik untuk memahami variasi dari data yang relevan mengenai kesiapan kadet. Komparasi kedua algoritma ini diharapkan dapat memberikan wawasan tentang metode yang lebih efektif dalam

mengidentifikasi tingkat kesiapan kadet, sesuai dengan standar prajurit karier yang ditetapkan oleh TNI. Hasil dari penelitian ini diharapkan dapat memberikan rekomendasi yang berguna bagi Unhan RI dalam mengoptimalkan program pembinaan kadet mahasiswa, sehingga secara langsung dapat meningkatkan kualitas sumber daya manusia dan membantu pengambilan keputusan strategis di Universitas Pertahanan Republik Indonesia.

1.2 Rumusan Masalah

Masalah yang ditemui sudah dibahas pada latar belakang. Di sana, penulis menemukan hal-hal yang terkait dengan topik yang diteliti. Dengan demikian peneliti merumuskan masalah sebagai berikut:

- a. Bagaimana efektivitas algoritma *K-Means* dan *Gaussian Mixture Models* dalam mengelompokkan kesiapan kadet mahasiswa S1 Unhan RI untuk SEPA PK TNI berdasarkan karakteristik dan kompetensi yang diukur?
- b. Manakah antara algoritma *K-Means* dan *Gaussian Mixture Models* yang menunjukkan performa klasterisasi yang lebih baik dalam hal akurasi dan relevansi dengan kriteria kesiapan prajurit karier TNI, dalam konteks analisis kesiapan SEPA PK TNI pada kadet mahasiswa S1 Unhan RI?
- c. Bagaimana analisis klasterisasi menggunakan *K-Means* dan *Gaussian Mixture Models* dapat memberikan insight untuk peningkatan program pembinaan di Unhan RI, khususnya dalam meningkatkan kesiapan kadet mahasiswa S1 dalam menghadapi SEPA PK TNI dan kontribusinya terhadap penguatan struktur pertahanan nasional?

1.3 Tujuan Penelitian

Dalam rumusan masalah, masalah yang akan dibahas dalam penelitian ini dijelaskan. Tujuan penelitian dapat ditentukan berdasarkan rumusan masalah. Tujuan dari penelitian ini adalah:

- a. Mengevaluasi dan membandingkan kinerja algoritma *K-Means* dan *Gaussian Mixture Models* dalam mengidentifikasi dan mengklasifikasikan kesiapan kadet mahasiswa S1 Unhan RI untuk SEPA PK TNI, dengan menggunakan variabel-variabel yang relevan terhadap kompetensi dan karakteristik prajurit karier TNI.
- b. Menentukan metode klasterisasi yang lebih efisien dan efektif antara *K-Means* dan *Gaussian Mixture Models* untuk menganalisis dan mengelompokkan data kesiapan SEPA PK TNI kadet mahasiswa S1 Unhan RI, berdasarkan akurasi, kepraktisan, dan relevansi dengan kebutuhan dan standar kesiapan prajurit karier TNI.
- c. Memberikan rekomendasi berdasarkan hasil analisis klasterisasi untuk meningkatkan program pembinaan di Unhan RI, yang secara spesifik bertujuan untuk memperkuat kesiapan kadet mahasiswa S1 dalam menghadapi SEPA PK TNI dan berkontribusi pada peningkatan efektivitas struktur pertahanan nasional Indonesia.

1.4 Manfaat Penelitian

Manfaat yang diharapkan didapat dari hasil penelitian dapat memberikan kontribusi bagi Resimen Korps Kadet Mahasiswa dan kadet mahasiswa S-1 Unhan RI dengan memberikan wawasan dalam hal kemampuan atau kapabilitas yang dimilikinya. Hasil penelitian ini dapat membantu dalam pengambilan keputusan mengenai perlunya pelaksanaan latihan khusus yang sesuai dengan kebutuhan dan potensi yang dimiliki oleh kadet. Penelitian ini bermanfaat bagi Universitas Pertahanan Republik Indonesia:

- a. Memberikan *insight* (nilai hasil dari penemuan pola dan tren yang diperoleh dari data yang terkumpul (analisis), sehingga dapat dijadikan sebuah tujuan, acuan, ataupun data yang dapat dipercaya) mengenai efektivitas pendekatan klasterisasi dalam memahami kesiapan dan distribusi kompetensi kadet, yang dapat diintegrasikan ke dalam strategi pembinaan dan pengembangan kurikulum di Unhan RI.
- b. Menambahkan literatur akademis melalui studi empiris tentang penerapan teknik *data mining* dalam analisis kesiapan militer, khususnya di lingkungan akademik pertahanan.

Penelitian ini bermanfaat pula bagi Lembaga Pertahanan/Kementrian Pertahanan:

- a. Menawarkan rekomendasi strategis berdasarkan temuan analisis klasterisasi untuk meningkatkan penyiapan kadet mahasiswa dalam menghadapi SEPA PK TNI, mendukung peningkatan kualitas sumber daya manusia dan membantu pengambilan keputusan strategis di Universitas Pertahanan Republik Indonesia.
- b. Meningkatkan pemahaman tentang potensi pemanfaatan data dan analitik dalam meningkatkan kesiapan kadet mahasiswa, khususnya melalui pengembangan SDM yang berorientasi pada pengambilan keputusan strategis di Universitas Pertahanan Republik Indonesia.

Secara keseluruhan, penelitian ini bertujuan untuk mendukung pengembangan strategi dan kebijakan yang lebih informatif bagi institusi pendidikan militer, pembuat kebijakan di sektor pertahanan, dan akademisi, dengan memberikan pemahaman yang lebih mendalam tentang bagaimana teknik analitik modern dapat diterapkan untuk meningkatkan kualitas sumber daya manusia.

1.5 Batasan Masalah

Penelitian ini akan fokus pada penggunaan dan perbandingan dua algoritma klasterisasi, yaitu *K-Means* dan *Gaussian Mixture Models*, untuk mengelompokkan dan mengevaluasi kesiapan kadet mahasiswa S1 di Universitas Pertahanan Republik Indonesia untuk SEPA PK TNI. Analisis ini akan dilakukan dengan menggunakan dataset yang mencakup variabel-variabel yang relevan terhadap kompetensi dan karakteristik prajurit karier TNI, seperti kemampuan akademik (misalnya, Indeks Prestasi Kumulatif) dan keterampilan fisik. Dataset yang digunakan akan mencakup informasi dari berbagai angkatan kadet, dengan fokus pada data yang terkumpul dari tahun terakhir masa pendidikan untuk memastikan relevansi dan keaktualan data dalam konteks saat ini.

Analisis akan dibatasi pada data yang dapat diakses dan telah disetujui oleh Universitas Pertahanan RI, dengan jumlah sampel yang ditentukan berdasarkan ketersediaan dan representasi yang memadai dari populasi kadet sebanyak kurang lebih 1204 kadet. Untuk menilai dan membandingkan efektivitas kedua algoritma dalam konteks ini, penelitian akan menggunakan indeks validitas kluster seperti indeks Davies-Bouldin dan Calinski-Harabasz, yang akan membantu dalam mengidentifikasi metode klasterisasi yang lebih akurat dan efektif untuk tujuan penelitian.

Pembatasan ini bertujuan untuk memfokuskan penelitian pada area yang spesifik dan mengelola sumber daya dengan efisien, sambil memastikan bahwa hasil analisis dapat memberikan wawasan yang signifikan untuk meningkatkan program pembinaan di Unhan RI dan mendukung program-program yang dapat meningkatkan kualitas sumber daya manusia serta membantu pengambilan keputusan strategis di Universitas Pertahanan Republik Indonesia.

BAB II

TINJAUAN PUSTAKA

2.1 Landasan Teori

Berikut adalah landasan teori terkait penelitian Komparatif Klasterisasi Algoritma *K-Means* dan *Gaussian Mixture Models* untuk Kesiapan Sekolah Perwira Prajurit Karier TNI di Universitas Pertahanan Republik Indonesia.

2.1.1 Komponen Utama

Komponen utama merujuk kepada bagian yang paling penting atau krusial dalam suatu sistem. Komponen utama dalam sistem pertahanan negara Indonesia, seperti diatur dalam Pasal 1 ayat 56 dan 7 UU No. 3 Tahun 2002 tentang Pertahanan Negara, adalah Tentara Nasional Indonesia/TNI (Sulistyo, 2020). Komponen utama mengacu pada peran sentral TNI dalam sistem pertahanan negara. TNI, yang terdiri dari TNI Angkatan Darat, TNI Angkatan Laut, dan TNI Angkatan Udara, dianggap sebagai komponen utama yang bertugas mempertahankan, melindungi, dan memelihara keutuhan serta kedaulatan negara. Sebagai komponen utama, TNI memiliki peran sentral dalam menjaga keamanan nasional, serta berperan dalam membantu pemerintah dalam menjaga ketertiban masyarakat dan menangani berbagai situasi darurat. Dengan demikian, TNI sebagai komponen utama memiliki tanggung jawab yang sangat besar dalam menjaga keamanan dan kedaulatan negara Indonesia. Melalui perannya sebagai komponen utama, TNI menjadi tulang punggung dalam menjaga stabilitas dan keamanan negara.

Tentara Nasional Indonesia memegang peranan krusial sebagai institusi militer yang bertugas menjaga kedaulatan dan keamanan negara (Rohmy et al., 2020). Sebagai lembaga yang beroperasi di bawah naungan

Presiden dan koordinasi Departemen Pertahanan. Menurut peraturan perundang-undangan, TNI dibentuk oleh warga negara yang berdedikasi dan terdiri dari berbagai elemen masyarakat, termasuk prajurit, baik yang secara sukarela maupun wajib (Yuli W et al., 2023), serta prajurit siswa yang menjalani pendidikan dasar militer untuk menjadi bagian dari kekuatan pertahanan.

2.1.1.1 Tentara Nasional Indonesia (TNI)

Identitas TNI, yang tertuang dalam Undang-Undang, menggarisbawahi empat aspek utama: sebagai Tentara Rakyat, Tentara Pejuang, Tentara Nasional, dan Tentara Profesional (Uksan, 2022). Aspek-aspek ini mencerminkan sifat TNI yang inklusif, perjuangan tanpa kenal menyerah, komitmen terhadap kepentingan nasional yang melampaui kepentingan sempit lainnya, serta profesionalisme yang diukur melalui pelatihan, pendidikan, kedisiplinan, netralitas politik, dan kepatuhan terhadap hukum nasional dan internasional. Kajian Samuel P. Huntington tahun 1957 menunjukkan bahwa TNI dianggap sebagai salah satu profesi yang memiliki standar kompetensi, keterampilan, dan tanggung jawab terhadap masyarakat yang menjadikan hal-hal tersebut sebagai indikator profesionalismenya. Profesionalisme TNI juga tercantum dalam UU No. 34 Tahun 2004 tentang TNI, dimana TNI harus dikembangkan berdasarkan kepentingan politik negara merujuk pada prinsip demokrasi, HAM, dan supremasi sipil sesuai dengan hukum yang telah diresmikan (Lubis, 2022).

Dalam hal pengerahan dan penggunaan kekuatan militer, kedudukan TNI yang strategis di bawah Presiden dan koordinasi dengan Departemen Pertahanan menegaskan peranannya sebagai pilar keamanan yang beroperasi dalam kerangka strategis nasional untuk menghadapi berbagai ancaman, baik itu ancaman militer maupun ancaman non-militer, yang dapat mengganggu kedaulatan dan keutuhan wilayah negara (Anugrah, 2023).

Peran, fungsi, dan tugas Tentara Nasional Indonesia sebagai alat negara dalam bidang pertahanan dijabarkan secara terperinci melalui Undang-Undang Republik Indonesia (Kamil, 2022). TNI memiliki peran kritis dalam menjaga kedaulatan negara, keutuhan wilayah, dan keselamatan bangsa dari ancaman dan gangguan, baik militer maupun bersenjata, dari dalam maupun luar negeri. Fungsi TNI mencakup penangkalan, penindakan, dan pemulihan kondisi keamanan negara (Amri, 2023). Tugas pokok TNI meliputi operasi militer untuk perang dan operasi militer selain perang (Putra & Lubis, 2022), yang mencakup berbagai aktivitas mulai dari mengatasi separatisme dan terorisme hingga mendukung tugas pemerintahan dan pengamanan objek vital nasional.

Dalam Undang-Undang Republik Indonesia Nomor 34 Tahun 2004 tentang Tentara Nasional Indonesia (Republik Indonesia, 2004), terdapat beberapa ketentuan yang mengatur proses pengangkatan prajurit. Pasal 28 menguraikan persyaratan umum yang harus dipenuhi oleh calon prajurit, termasuk kewarganegaraan Indonesia, keyakinan kepada Tuhan Yang Maha Esa, kesetiaan kepada Negara Kesatuan Republik Indonesia yang berlandaskan Pancasila dan Undang-Undang Dasar Negara Republik Indonesia Tahun 1945, usia minimum 18 tahun saat dilantik, tidak memiliki catatan kriminalitas tertulis dari Kepolisian Negara Republik Indonesia, kesehatan jasmani dan rohani yang baik, serta tidak memiliki putusan pengadilan yang mencabut hak menjadi prajurit. Selain itu, calon prajurit juga harus lulus dari pendidikan pertama yang membentuk prajurit siswa menjadi anggota TNI, serta memenuhi persyaratan lain yang sesuai dengan kebutuhan.

Pasal 29 menjelaskan bahwa pendidikan untuk pengangkatan prajurit terbagi menjadi tiga tingkatan, yaitu pendidikan perwira, bintara, dan tamtama. Pelaksanaan pendidikan ini diatur lebih lanjut oleh keputusan Panglima.

Pasal 33 mengatur bahwa Presiden mengangkat perwira atas usul dari Panglima, sedangkan bintara dan tamtama diangkat langsung oleh Panglima. Detail terkait ketentuan ini akan diatur lebih lanjut dalam Peraturan Pemerintah.

Pasal 34 menjelaskan bahwa pelantikan menjadi prajurit dilakukan dengan mengucapkan Sumpah Prajurit, dan bagi prajurit golongan perwira, juga mengucapkan Sumpah Perwira. Pelantikan ini merupakan momen penting dalam proses pengangkatan prajurit.

2.1.1.2 Sekolah Perwira Prajurit Karier TNI

Sebagaimana diatur dalam Undang-Undang Republik Indonesia Nomor 34 Tahun 2004 tentang Tentara Nasional Indonesia tepatnya pada pasal 30 menguraikan bahwa perwira dapat dibentuk melalui dua jalur, yaitu pendidikan pertama perwira bagi calon yang berasal langsung dari masyarakat, melalui Akademi TNI atau Sekolah Perwira, dengan masukan dari Sekolah Lanjutan Tingkat Atas atau Perguruan Tinggi. Selain itu, perwira juga dapat dibentuk melalui pendidikan pembentukan perwira yang berasal dari prajurit golongan bintara. Rincian mengenai pendidikan perwira diatur lebih lanjut oleh keputusan Panglima (Republik Indonesia, 2004).

Profesionalisme TNI dalam perihal seleksi SEPA PK (Sekolah Perwira Prajurit Karier) TNI mencakup aspek intelektual dan fisik yang sangat penting (Lubis, 2022). Tes akademik yang ketat memastikan bahwa calon perwira TNI memiliki kemampuan intelektual yang tinggi, selain tes akademik hal ini ditunjukkan dengan minimal indeks prestasi kumulatif sebesar 3,2 (bisa berubah suatu saat sesuai dengan keputusan tertinggi dan kebutuhan pada tahun itu) agar dengan mudah mampu memahami konsep-konsep militer, dan memiliki kemampuan analisis yang baik. Di samping itu, tes kesegaran jasmani dan postur juga memastikan bahwa calon perwira TNI memiliki kondisi fisik yang prima, memakai indikator yang sesuai dengan dengan Peraturan Panglima (Perpang) Tentara Nasional

Indonesia Nomor 1388 (Panglima TNI, 2018) dimana nilai tertinggi tes kesegaran jasmani kategori 1 (18-22 tahun) sebagai berikut: nilai maksimum untuk pria lari selama 12 menit adalah 3444 meter, diikuti pull up sebanyak 18 kali, sit up sebanyak 41 kali, push up sebanyak 43 kali, dan shuttle dengan hitungan 15 detik. Sedangkan untuk Wanita lari selama 12 menit ada pada 3000 meter, diikuti dengan nilai chinning, sit up, push up, berturut-turut pada angka 63 kali, 43 kali, 28 kali dan shuttle run pada hitungan 17 detik. Sementara itu nilai minimum untuk pria lari selama 12 menit sejauh 2367 meter, pull up 6 kali, sit up 26 kali, dan push up sebanyak 24 kali serta dalam hitungan 22,8 detik untuk shuttle run. Untuk wanita lari 12 menit sebanyak 1953 meter, chinning sebanyak 35 kali, sit up sebanyak 23 kali, push up sebanyak 16 kali, dan shuttle dihitung 23,1 detik. Indikator ini memastikan agar calon perwira mampu menjalani tugas-tugas militer dengan baik, serta mampu bertahan dalam kondisi yang menguras fisik. Dengan mengintegrasikan aspek intelektual dan fisik dalam seleksi SEPA PK TNI.

Profesionalisme TNI tidak hanya mencakup kemampuan akademik yang tinggi, tetapi juga kesiapan fisik yang optimal untuk menjalankan tugas-tugas militer dengan efektif dan efisien. Hal ini menegaskan komitmen TNI dalam mencari dan melatih para pemimpin militer yang profesional dan siap menghadapi berbagai tantangan dalam menjaga keamanan dan kedaulatan negara (Lubis, 2022).

2.1.2 Data Mining

Data mining merupakan bidang ilmu yang berfokus pada menemukan pengetahuan yang belum terungkap dari sejumlah besar data. Istilah *data mining* seringkali sinonim dengan '*knowledge discovery*' atau '*pattern recognition*', yang masing-masing memiliki nuansa yang tepat dalam konteks yang berbeda. *Knowledge discovery* mengacu pada tujuan utama *data mining* yaitu menemukan pengetahuan yang belum terungkap, sedangkan *pattern recognition* berkaitan dengan pengenalan pola yang ada

dalam data tersebut (Rahayu et al., 2021). Sedangkan menurut (Hutagalung, 2022) *data mining* adalah proses penting yang melibatkan pemilahan melalui repositori data yang sangat besar untuk menemukan korelasi, pola, dan tren baru yang bermakna. Proses ini menggunakan teknologi pengenalan pola serta teknik statistik dan matematika untuk mengekstrak informasi berharga dari kumpulan data yang besar. Pada intinya, *data mining* dapat dipahami sebagai pemeriksaan kumpulan data observasi yang besar untuk menemukan hubungan yang tidak terduga dan meringkas data dengan cara yang baru, dapat dimengerti, dan berguna bagi pemilik data. Ini adalah domain interdisipliner yang menggabungkan metode dari pembelajaran mesin, pengenalan pola, statistik, basis data, dan visualisasi untuk memecahkan tantangan ekstraksi informasi dari basis data yang luas.

Proses *data mining* bukanlah sesuatu yang dapat dilakukan secara mekanis dan hanya menunggu hasilnya. Sebaliknya, *data mining* adalah proses interaktif yang sesuai dengan strategi bisnis atau penelitian yang lebih luas, dengan metodologi khusus seperti CRISP-DM menjadi salah satu cara untuk menyusun proses ini. Data mining dapat menyelesaikan beberapa tugas utama, yang umumnya dicari di lapangan. Tugas-tugas ini meliputi:

- a. Deskripsi: Menemukan cara untuk mendeskripsikan pola dan tren dalam data.
- b. Estimasi: Memperkirakan nilai atau keadaan yang tidak diketahui dari suatu sistem berdasarkan data yang diketahui.
- c. Prediksi: Memperkirakan keadaan atau kejadian di masa depan berdasarkan data saat ini dan data historis.
- d. Klasifikasi: Menetapkan item ke dalam kategori atau kelas yang telah ditentukan.
- e. Pengelompokan: Mengelompokkan item ke dalam kelompok berdasarkan kesamaan atau karakteristik terkait lainnya.

- f. Asosiasi: Menemukan aturan yang menggambarkan bagaimana item atau peristiwa dikaitkan satu sama lain.

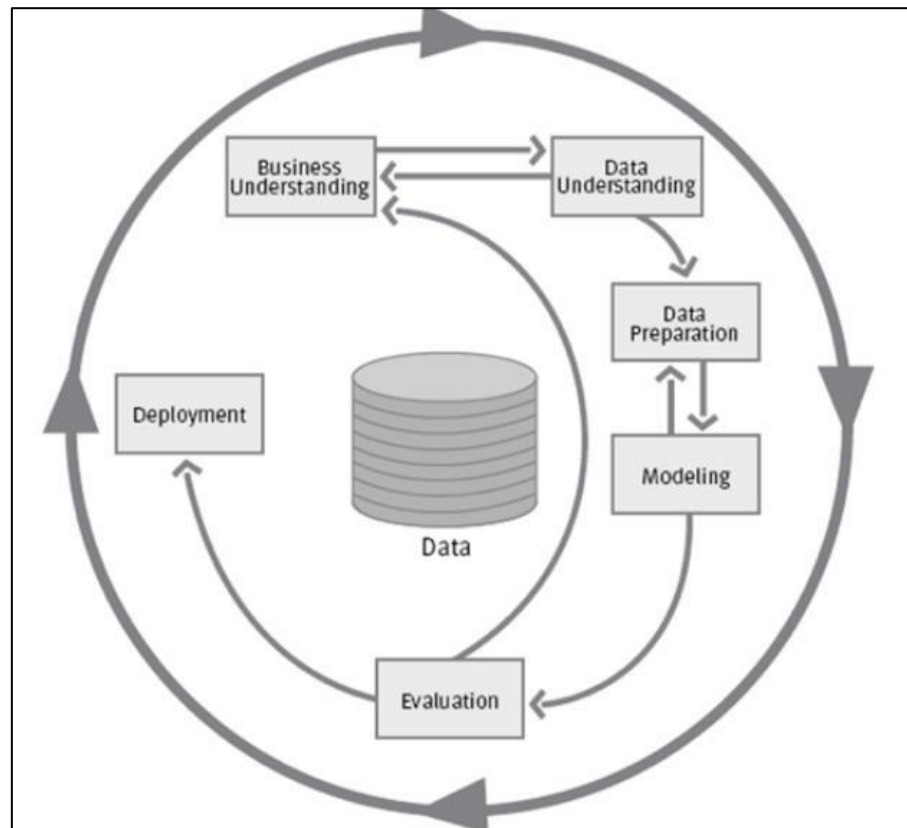
Tugas-tugas ini merupakan inti dari praktik data mining dan biasanya merupakan tujuan yang mendorong proses data mining (Umagapi et al., 2023).

2.1.2.1 CRISP-DM

Singkatan dari *Cross-Industry Standard Process for Data Mining*, merupakan proses standar terbuka yang menggambarkan pendekatan umum untuk menerapkan proyek *data mining*. Model ini ditandai dengan fleksibilitasnya, karena tidak spesifik untuk sektor industri atau teknik penggalian data apa pun, dan tidak bergantung pada perangkat lunak penggalian data komersial apa pun. Karena penilaian dan pengambilan keputusan manusia diperlukan untuk pengenalan data baru dan pemantauan kualitas model secara terus menerus, CRISP-DM melibatkan aspek manusia dalam pengolahan data. Menghapus beberapa kesalahpahaman tentang *data mining*, seperti bahwa itu dapat berjalan secara otonom tanpa bantuan manusia. Singkatnya, *data mining* membutuhkan pemahaman yang mendalam tentang konteks bisnis atau penelitian, keterampilan domain, dan kemampuan analisis. Selain itu, tidak secara otomatis mengidentifikasi penyebab utama masalah atau membersihkan basis data. Tugas-tugas ini merupakan bagian dari proses penemuan pengetahuan, yang bergantung pada analisis manusia untuk menginterpretasikan pola dan membersihkan data yang mungkin sudah tidak terpakai atau diperiksa selama bertahun-tahun. Proses CRISP-DM dibagi menjadi enam fase, yaitu:

- a. Pemahaman Bisnis: fase awal ini berfokus pada pemahaman tujuan dan persyaratan proyek dari perspektif bisnis. Fase ini melibatkan pemahaman terhadap tujuan proyek, latar belakang, dan sumber daya yang tersedia.

- b. Pemahaman Data: hal ini melibatkan pengumpulan atau mengakses data, memahami data yang dimiliki, dan menemukan wawasan pertama ke dalam data, atau mendeteksi subset yang menarik untuk membentuk hipotesis untuk informasi yang tersembunyi.
- c. Persiapan Data: fase ini mencakup semua kegiatan yang diperlukan untuk membangun kumpulan data akhir dari data mentah awal. Tugas persiapan data kemungkinan besar akan dilakukan beberapa kali, dan tidak dalam urutan tertentu. Tugas-tugas tersebut meliputi pemilihan tabel, catatan, dan atribut, serta transformasi dan pembersihan data untuk alat pemodelan.
- d. Pemodelan: pada tahap ini, berbagai teknik pemodelan dipilih dan diterapkan, serta parameternya dikalibrasi ke nilai yang optimal. Mungkin ada beberapa teknik untuk jenis masalah *data mining* yang sama. Beberapa teknik memiliki persyaratan khusus pada bentuk data. Oleh karena itu, kembali ke fase persiapan data sering kali diperlukan
- e. Evaluasi: pada tahap ini dalam proyek, telah dibangun sebuah model (atau beberapa model) yang tampaknya memiliki kualitas tinggi, dari perspektif analisis data. Sebelum melanjutkan ke penerapan akhir model, penting untuk mengevaluasi model secara lebih menyeluruh, dan meninjau langkah-langkah yang dilakukan untuk membangun model, untuk memastikan bahwa model tersebut telah mencapai tujuan bisnis dengan benar.
- f. Penyebaran: pembuatan model pada umumnya bukanlah akhir dari proyek. Pengetahuan yang diperoleh perlu diorganisir dan disajikan sedemikian rupa sehingga pelanggan dapat menggunakannya. Bergantung pada kebutuhan, fase penerapan dapat sesederhana membuat laporan atau serumit mengimplementasikan proses penggalian data yang berulang di seluruh perusahaan (Muttaqin et al., 2022).



Gambar 2. 1 CRISP-DM: 6 Fase

Sumber: (Rahayu et al., 2021)

Pentingnya dokumentasi yang baik dalam setiap fase CRISP-DM memastikan transparansi dan kemampuan replikasi untuk proyek masa depan. Proses siklus data mining memungkinkan penyempurnaan berkelanjutan terhadap pemahaman bisnis dengan setiap iterasi, menambah efisiensi dan fokus terhadap tujuan bisnis.

2.1.3 Klasterisasi

Pengorganisasian data, pengamatan, atau kasus ke dalam kelompok atau "klaster" berdasarkan tingkat kemiripan merupakan proses klastering. Data dalam setiap klaster sangat mirip satu sama lain dan sangat berbeda dari data dalam klaster lain. Tujuan utama dari proses ini adalah untuk memastikan bahwa data dalam satu klaster memiliki tingkat kemiripan yang tinggi, sementara tingkat kemiripan antar klaster dijaga serendah mungkin.

agar menghasilkan kelompok-kelompok yang terdefinisi dengan jelas dan berbeda. Tidak ada yang namanya variabel target ketika menggunakan *unsupervised method*. Alih-alih, mencari struktur dan pola dalam semua variabel. klustering adalah teknik *unsupervised data mining* yang paling banyak digunakan (Oyewole & Thopil, 2023).

Clustering adalah proses pengelompokan data menjadi beberapa kelompok atau klaster, dimana data dalam satu klaster memiliki karakteristik yang lebih mirip satu sama lain daripada dengan data di klaster lain. Proses ini melibatkan pengidentifikasian pola dalam kumpulan data dan pengelompokan data berdasarkan kesamaan atribut. *Clustering* digunakan dalam berbagai aplikasi seperti pengelolaan data, analisis statistik, *machine learning*, dan *data mining*. Teknik *clustering* memungkinkan pengambilan kesimpulan atau keputusan berdasarkan karakteristik data yang dikelompokkan, yang dapat memperkaya analisis data dan membantu dalam pengambilan keputusan (Herlinda & Darwis, 2021).

2.1.3.1 Kluster-Means (K-Means)

Menurut (Rahmati & Wijayanto, 2021) *K-Means* adalah algoritma *clustering* yang populer dalam analisis data dan pembelajaran mesin. Metode ini mengelompokkan data menjadi sejumlah klaster (k) yang telah ditentukan sebelumnya. Proses ini terdiri dari penentuan pusat klaster (*centroid*) secara acak, setiap titik data ke klaster terdekat berdasarkan jarak *Euclidean*, dan pembaruan posisi *centroid*. Proses ini diulang hingga posisi *centroid* stabil atau kelompok data tidak mengalami perubahan yang signifikan. *K-means* bekerja dengan baik untuk data dengan distribusi yang hampir seragam dan terpisah; namun, tidak bekerja dengan data dengan distribusi yang kompleks atau klaster non-bulat. Nilai awal *centroid* serta jumlah klaster yang ditentukan oleh algoritma ini juga penting.

K-Means pada dasarnya melakukan dua tugas: menemukan lokasi pusat kluster dan mencari anggota setiap kluster. Algoritmanya digambarkan sebagai berikut:

- a. Tentukan nilai k sebagai jumlah cluster yang ingin dibentuk.
- b. Bangkitkan titik pusat *cluster*, atau *centroid*, sebanyak k secara acak
- c. Gunakan rumus korelasi antara dua objek, yaitu jarak geometris, untuk menghitung jarak masing-masing data ke *centroid*.

$$d_u = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.1)$$

- d. Grupkan setiap data berdasarkan jarak terdekat antara data dan *centroid*-nya.
- e. Iterasinya, kemudian cari posisi *centroid* baru.
- f. Jika posisi *centroid* tidak sama, ulangi langkah ketiga.

Selain sangat populer pada industri data mining, algoritma *K-Means* bermanfaat untuk mengidentifikasi segmentasi yang muncul selama sebuah kejadian tanpa harus mengetahui kriteria khusus dari setiap segmentasi. Namun, diperlukan iterasi terus-menerus sampai hasil yang diperoleh tidak berubah secara signifikan, yang berarti bahwa parameter yang lebih banyak diperlukan untuk mendapatkan hasil yang lebih optimal (Kamila et al., 2021).

2.1.3.2 *Gaussian Mixture Models (GMM)*

Distribusi Normal ialah nama lain dari distribusi *Gaussian*, merupakan distribusi yang bersifat probabilitas kontinu dengan persamaan (Patel & Kushwaha, 2020):

$$N(X|\mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}} \sqrt{|\Sigma|}} \exp \left\{ -\frac{(X - \mu)^T \Sigma^{-1} (X - \mu)}{2} \right\} \quad (2.2)$$

Dimana μ adalah vektor rata-rata D-dimensi, Σ merupakan matriks D x D, yang menginterpretasikan *Gaussian* dan $|\Sigma|$ menunjukkan determinan dari Σ .

Standar deviasi dan rata-rata menunjukkan distribusi *Gaussian* yang simetris terhadap rata-rata. Namun, properti unimodal distribusi *Gaussian* tunggal tidak dapat memodelkan berbagai wilayah kepadatan dalam kumpulan data multimodal yang ditemukan dalam situasi praktis. Akibatnya, distribusi ini tidak dapat mempelajari perilaku penggunaan sumber daya yang berbeda dari beban kerja. Sebaliknya, *Gaussian Mixture Model* dapat digunakan untuk memodelkan distribusi yang kompleks. GMM adalah metode pengelompokan tanpa pengawasan (*unsupervised method*) yang menggunakan *Expectation-Maximization (EM)* untuk mengestimasi kepadatan probabilitas dan menghasilkan kelompok elips (Wahidah & Utari, 2023). *Gaussian Mixture Model* digunakan untuk setiap klaster. Kuantitas kebugaran per klaster dapat diukur lebih baik dengan GMM.

Mampu memodelkan lebih dari satu distribusi *Gaussian* untuk sebuah set data adalah keunggulan metode GMM dengan konsep menggunakan distribusi probabilitas kelas dengan kombinasi linier k distribusi *Gaussian* yang dikenal juga sebagai komponen GMM (Adhinata et al., 2021). Dengan menggunakan kombinasi linier distribusi dasar probabilitas *Gaussian*, GMM dapat digambarkan sebagai berikut:

$$p(X) = \sum_{k=1}^K \pi_k N(X|\mu_k, \Sigma_k) \quad (2.3)$$

Di mana k adalah jumlah komponen dalam model campuran dan π_k disebut koefisien pencampuran, yang memberikan estimasi densitas setiap komponen *Gaussian*. $N(X|\mu_k, \Sigma_k)$ yang menandakan Densitas *Gaussian* disebut komponen model campuran. Tiap komponen k dijelaskan oleh distribusi *Gaussian* dengan rata-rata μ_k , kovariansi Σ_k dan koefisien pencampuran π_k .

$$\ln p(X|\pi, \mu, \Sigma) = \sum_{n=1}^K \ln \left\{ \sum_{k=1}^K \pi_k N(x_n|\mu, \Sigma) \right\} \quad (2.4)$$

Namun terdapat kelemahan dalam metode ini seperti; estimator GMM adalah *asymptotically efficient* yang dalam ukuran besar tetapi kurang efisien dalam ukuran yang terbatas (*finite*) dimana estimator ini terkadang memerlukan implementasi pendukung seperti perangkat lunak (*software*) (Prasetyo, 2020).

2.1.3.3 Indeks Davies-Bouldin (Davies-Bouldin Index)

Dengan membandingkan *compactness* atau kepadatan klaster dengan jarak antar pusat klaster, indeks Davies-Bouldin mengukur kualitas pengelompokkan. Nilai indeks yang lebih rendah menunjukkan klaster yang lebih terpisah dan kohesif, yang merupakan indikator pengelompokan yang baik. Penilaian ini didasarkan pada rasio antara kepadatan internal klaster dan jarak antar klaster, dan nilai indeks yang lebih rendah menunjukkan konfigurasi klaster yang lebih optimal (Umagapi et al., 2023).

Persamaan *Davies-Bouldin Index* atau *DBI* menghitung validitas diantara dua klaster C_i dan C_j sebagai berikut:

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (2.5)$$

dalam indeks Davies-Bouldin, R_{ij} merepresentasikan nilai indeks antara dua klaster, C_i dan C_j . Ukuran dispersi dalam klaster C_i , yang diberi label S_i ,

dan dalam klaster C_j , yang diberi label S_j . Sementara itu, d_{ij} adalah jarak antara pusat kedua klaster tersebut, m_i dan m_j .

Menghitung Indeks Davies-Bouldin melibatkan beberapa langkah. Pertama, menentukan pusat klaster untuk setiap klaster. Kemudian, mengukur dispersi dalam setiap klaster dengan menghitung jarak rata-rata dari titik data ke pusat klaster. Selanjutnya, menghitung jarak antara pusat klaster. Setelah itu, menggunakan rumus Indeks Davies-Bouldin untuk menentukan nilai R_{ij} antara setiap pasangan klaster. Proses ini diulangi untuk semua kombinasi klaster. Akhirnya, nilai Indeks Davies-Bouldin secara keseluruhan dihitung menggunakan rumus:

$$DB = \frac{1}{N} \sum_{i=1}^n \max_{j \neq i} (R_{ij}) \quad (2.6)$$

DB menunjukkan *Davies-Bouldin Index* secara keseluruhan dan N adalah jumlah total klasternya.

2.1.3.4 Indeks Calinski-Harabasz (*Calinski-Harabasz Index*)

Indeks Calinski-Harabasz dikenal juga sebagai *variance ratio criterion*, merupakan metode yang efektif untuk mengevaluasi kinerja model pengelompokan data atau clustering (Sikana & Wijayanto, 2021) Jumlah klaster yang ideal sering dihitung dengan indeks ini (Rahmati & Wijayanto, 2021).

$$C = \frac{\frac{BGSS}{K-1}}{\frac{WGSS}{K-1}} = \frac{N-K}{K-1} \frac{BGSS}{WGSS} \quad (2.7)$$

Dimana $BGSS$ (*Between-Group Sum of Squares*) merupakan jumlah kuadrat antar-kelompok (*disperse* antar-kelompok), sedangkan $WGSS$ (*Within-Group Sum of Squares*) jumlah kuadrat dalam kelompok

(*disperse* dalam kelompok), N menunjukkan total dari data poin/observasi, dan K menunjukkan jumlah total klaster.

CH (Calinski-Harabasz) Index menghitung rasio dispersi antar-klaster dan intra-klaster. Nilai yang tinggi pada indeks ini menunjukkan dispersi antar-klaster yang besar dan dispersi intra-klaster yang rendah, yang menunjukkan kualitas pengelompokan yang baik, singkatnya jumlah klaster yang baik ditunjukkan dengan nilai *CH* yang meningkat (Luthfi & Wijayanto, 2021). Misalkan suatu himpunan data terdiri dari k klaster dan N titik data, di mana C_i adalah klaster ke- i , X_i adalah titik ke- i pada klaster ke- i , N_i adalah jumlah titik pada klaster ke- i , dan \bar{X}_i adalah titik pusat klaster ke- i .

$$BGSS = \sum_{i=1}^k N_i (X_i - \bar{X}_i)(X_i - \bar{X}_i)^T \quad (2.8)$$

$$WGSS = \sum_{i=1}^k \sum_{X_i \in C_i} (X_i - \bar{X}_i)(X_i - \bar{X}_i)^T \quad (2.9)$$

2.2 Hasil Penelitian Terdahulu

Ketika melakukan penelitian, penelitian terdahulu dapat membantu memperluas dan memahami teori yang akan digunakan dalam penelitian. Berdasarkan penelitian terdahulu ditemukan beberapa penelitian yang relevan dengan penelitian ini:

Tabel 2. 1 Hasil Penelitian Terdahulu

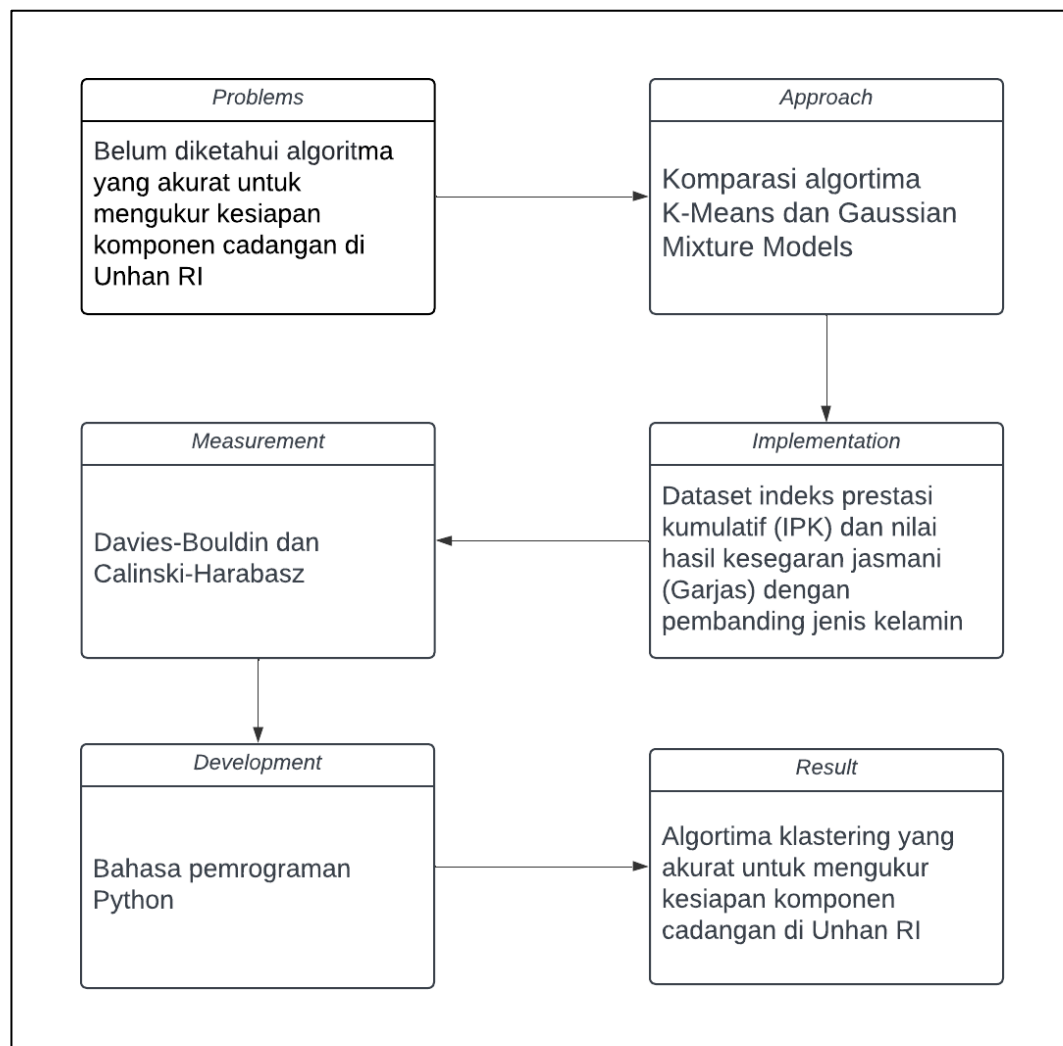
No.	Penulis (Tahun) dan Judul	Metode	Hasil	Gap Penelitian
1.	Hutagalung, J., Syahputra, Y. H., & Tanjung, Z. P. (2022). <i>Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K- Means Clustering</i>	<i>K-Means</i>	Menghasilkan 3 kelompok siswa dengan masing-masing kemampuan: - <i>Cluster 1</i> :10 siswa berkemampuan pemrograman <i>android</i> - <i>Cluster 2</i> :62 siswa berkemampuan pemrograman <i>web</i> - <i>Cluster 3</i> :48 siswa berkemampuan pemrograman <i>desktop</i>	Algoritma yang dipakai hanya <i>K-Means</i> dan tidak menerapkan perbandingan sehingga tidak mengetahui algoritma yang lebih optimal untuk mengelompokkan siswa, tidak menggunakan indeks validitas.
2.	Umagapi, I. T., Umaternate, B., Hazriani, H., & ... (2023). <i>Uji Kinerja K- Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompoka</i>	<i>K-Means</i> dan indeks validasi <i>Davies-Bouldin</i>	- Hasil <i>DBI</i> (<i>Calinski-Harabasz Index</i>) untuk 3 <i>cluster</i> adalah 0.9250. - Hasil <i>DBI</i> untuk 4 <i>cluster</i> adalah 1.1584.	Tidak ada hasil yang menunjukkan algoritma yang lebih optimal untuk mengelompokkan siswa karena algoritma yang digunakan hanya <i>K-Means</i> dan tidak menerapkan perbandingan. Evaluasi dilakukan tetapi hanya

	<i>n Data Prestasi Siswa</i>			menggunakan 1 indeks validitas yaitu <i>DBI</i> .
3.	Novkaniza, F., Nico, & Kafi, R. Al. (2023) <i>Pemodelan Jumlah Kasus Baru Harian Covid-19 di Indonesia Menggunakan Gaussian Mixture Model</i>	<i>Gaussian Mixture Model</i>	<ul style="list-style-type: none"> - Probabilitas kasus baru harian kurang dari kasus terendah adalah 0,009598. - Probabilitas kasus baru harian lebih dari kasus rata-rata adalah 0,299443. - Probabilitas kasus baru harian lebih dari kasus tertinggi adalah 0,017669. 	Algoritma yang digunakan tidak menggunakan indeks validitas dan hanya menggunakan <i>Gaussian Mixture Models</i> , sehingga tidak diketahui algoritma mana yang lebih baik untuk mengelompokkan.
4.	Windia Ambarsari, E., Dwitianti, N., Selvia, N., Nur Cholifah, W., & Dina Mardika, P. (2023). <i>Comparison Approaches of the Fuzzy C-Means and Gaussian Mixture Models in Clustering the Welfare of</i>	<i>Fuzzy C-Means (FCM) dan Gaussian Mixture Model</i>	<ul style="list-style-type: none"> - Jumlah optimal cluster yang ditemukan untuk 34 provinsi adalah 2 (<i>FCM</i>) dan 6 (<i>GMM</i>). - Indeks validitas yang digunakan untuk pengujian termasuk <i>Silhouette</i>, <i>Davies-Bouldin</i>, dan <i>Calinski-Harabasz</i>. - Metode <i>GMM</i> memiliki kualitas kelompok yang lebih baik daripada <i>FCM</i>. 	Pada penelitian berikut menggunakan komparasi algoritma <i>FCM</i> dan <i>GMM</i> , sedangkan untuk penelitian ini menggunakan komparasi algoritma <i>K-Means</i> dan <i>GMM</i> serta menggunakan 3 indeks validitas yaitu <i>Silhouette</i> , <i>Davies-Bouldin</i> , dan <i>Calinski-Harabasz</i> sedangkan penelitian ini menggunakan 2 indeks

	<i>the Indonesian People</i>			validitas yaitu <i>Davies-Bouldin</i> dan <i>Calinski-Harabasz</i> .
5.	<p>Wahidah, Z., & Utari, D. T. (2023)</p> <p><i>Comparison of K-Means and Gaussian Mixture Model in Profiling Areas by Poverty Indicators</i></p>	<i>K-Means dan Gaussian Mixture Model</i>	<p><i>-K-Means</i> <i>cluster 1: 10</i> kabupaten/kota <i>cluster 2: 17</i> kabupaten/kota <i>cluster 3: 8</i> kabupaten/kota <i>-GMM</i> <i>cluster 1: 10</i> kabupaten/kota <i>cluster 2: 19</i> kabupaten/kota <i>cluster 3: 6</i> kabupaten/kota -Memakai indeks validasi <i>Connectivity, Dunn, Silhoutte</i> - <i>GMM</i> memberikan hasil terbaik dengan 3 <i>cluster</i>. - Jumlah anggota untuk <i>cluster: 10, 19, dan 6</i>. - Algoritma <i>GMM</i> menghasilkan hasil analisis <i>cluster</i> yang lebih baik daripada algoritma <i>K-Means</i>.</p>	<p>Penelitian ini menggunakan 3 indeks validitas yaitu <i>Connectivity, Dunn, Silhoutte</i> sedangkan penelitian ini menggunakan 2 indeks validitas yaitu <i>Davies-Bouldin</i> dan <i>Calinski-Harabasz</i>. Metode penelitian tidak menggunakan metode CRISP-DM.</p>

2.3 Kerangka Pemikiran

Menggabungkan teori dengan bukti empiris dan penelitian terdahulu membentuk kerangka berpikir yang kuat. Membantu dalam mendefinisikan variabel penelitian dan memberikan penjelasan yang lengkap dan relevan dengan masalah yang diteliti. Untuk menjawab masalah penelitian, kerangka berpikir ini dapat digunakan seperti pada Tabel 1. Hal ini bermanfaat untuk mengubah pernyataan atau/dan membangun materi penjelasan (Syahputri et al., 2023)



Gambar 2. 4 Kerangka Berpikir

Sumber: diolah oleh peneliti

Gambar 2.7 menginterpretasikan beberapa langkah yang akan digunakan pada penelitian ini. Di Unhan RI, masalah utama yang diteliti adalah kurangnya pengetahuan tentang algoritma yang akurat untuk mengukur kesiapan komponen cadangan. Untuk menemukan metode yang akurat, solusi yang disarankan adalah membandingkan dua algoritma: *Gaussian Mixture Models* dan *K-Means*. Dua indeks validitas (*Davies-Bouldin* dan *Calinski-Harabasz*) digunakan untuk menguji kinerja kedua algoritma ini. Berdasarkan pengembangan dari model yang dibuat menggunakan Python dengan desain eksperimen CRISP-DM.

BAB III

METODOLOGI PENELITIAN

3.1 Metode dan Desain Penelitian

Penelitian ini menggunakan model CRISP-DM, yang terdiri dari langkah-langkah berikut:

3.1.1 Pemahaman Bisnis (*Business Understanding*)

Dalam fase Pemahaman Bisnis (*Business Understanding*) dari metodologi CRISP-DM yang diadaptasi untuk penelitian ini, tujuan utama adalah untuk mengidentifikasi dan memahami tantangan yang dihadapi dalam mengukur kesiapan komponen cadangan di Unhan RI. Peneliti akan mengevaluasi efektivitas algoritma klustering yang ada, dengan fokus pada *K-Means* dan *Gaussian Mixture Models*, untuk menemukan solusi yang lebih akurat. Kajian ini akan mengembangkan pendekatan melalui bahasa pemrograman Python dan mengimplementasikannya pada dataset yang relevan, dengan hasil diukur menggunakan penilaian validasi seperti *Davies-Bouldin* dan *Calinski-Harabasz*. Hasil akhirnya bertujuan untuk menghasilkan algoritma klustering yang dapat diandalkan untuk tujuan pengukuran tersebut.

3.1.2 Pemahaman Data (*Data Understanding*)

Penelitian ini akan mengumpulkan langsung data primer dari evaluasi akademik dan fisik terakhir yang diambil melalui masing-masing fakultas dan *spreadsheet* yang disebar langsung kepada kadet mahasiswa. Data ini termasuk Indeks Prestasi Kumulatif dan nilai kesegaran jasmani, yang mencerminkan prestasi akademik dan kondisi fisik mahasiswa kadet selama masa pendidikan mereka di Universitas Pertahanan RI. Semua data

ini akan diproses dengan memperhatikan variabel jenis kelamin untuk menjamin kesetaraan dan keakuratan analisis.

REKAPITULASI NILAI UTS COHORT-2 TA. 2022 / 2023 – 2												
No	Nama	NIM	Kripto UTS	Mikropro UTS	AI UTS	MFLA UTS	IMIONS UTS	C4ISR UTS	Forensic UTS	Mining UTS	Rata-Rata Nilai UTS	Rank
1	Abic	3:	01	100	95			100		92	97	2
2	Ahr	3:	02	98	98			77		72	86	20
3	Alfia	3:	03	100	95			100		90	96	4
4	Alvi	3:	04	92	98			88		70	87	18
5	And	3:	05	100	98			93		98	97	1
6	Ded	3:	06	100	98			100		85	96	6
7	Dick	3:	07	100	98			87		76	90	16
8	Hel	3:	08	100	98			90		80	92	14
9	Kad	3:	09	98	93			95		85	93	11
10	Luth	3:	10	100	95			100		77	93	9
11	Man	3:	11	100	95			100		89	96	5
12	M. A	3:	12	80	98			93		60	83	24
13	M. F	3:	13	100	98			89		81	92	14
14	M. F	3:	14	88	98			98		85	92	13
15	Reg	3:	15	100	98			93		81	93	9
16	Ren	3:	16	93	93			85		70	85	22
17	Rick	3:	17	100	98			98		77	93	8
18	Rizk	3:	18	84	98			74		77	83	23
19	Roh	3:	19	95	98			83		50	82	25
20	Rol	3:	20	100	93			80		70	86	21
21	Rya	3:	21	96	95			93		90	94	7
22	Seft	3:	22	89	95			89		98	93	11
23	Teg	3:	23	100	93			90		64	87	19
24	Yog	3:	24	87	98			84		82	88	17
25	Yus	3:	25	78	98			70		63	77	26
26	Zefe	3:	26	98	90			100		98	97	3

Gambar 3. 1 Contoh Data Akademik Kadet Mahasiswa Cohort 2

Sumber: Arsip Prodi Informatika FSTP

87	Yosl Huta	Sermada Kadet	3	15	2,420	10	41	43	19	83	62.25	B	10 M	Bebas	0.46	
KIMIA																
88	Agu	Sermada Kadet	3	11	2,230	10	41	40	19	80	55.88	C	20 M	Bebas	1.08	
89	Ang	Sermada Kadet	3	12	2,010	11	41	43	19	83	52	C	75 M	Dada	3.05	
90	Ardy	Sermada Kadet	3	13	2,670	11	41	43	19	85	69.88	B	100 M	Dada	3.50	
91	Felo	Sermada Kadet	3	19	2,810	20	41	43	18	96	79.38	B	100 M	Dada	2.38	
92	Fikri	Sermada Kadet	3	11	2,890	20	41	43	17	97	82.13	BS	TIM RESQUE	Dada	2.36	
93	Fua	Sermada Kadet	3	12	2,500	18	41	43	17	97	71.75	B	100 M	Dada	2.41	
94	Moh	Sermada Kadet	3	13	2,790	19	40	43	18	94	78	B	50 M	Bebas	2.44	
95	Ori	Sermada Kadet	3	15	2,900	18	4	43	18	70	69	B	100 M	Dada	2.46	
96	Syal	Sermada Kadet	3	13	2,500	18	41	43	18	95	71	B	25 M	Bebas	1.17	
FISIKA																
97	Agu	Sermada Kadet	3	12	2,560	18	41	43	19	93	71.5	B	100 M	Dada	2.50	
98	Ardi	Sermada Kadet	3	13	2,310	12	41	43	18	86	61.63	B	60 M	Bebas	3.20	
99	Dim	Sermada Kadet	3	15	2,470	1	41	43	20	70	57.38	C	50 M	Bebas	2.58	
100	Faki	Sermada Kadet	3	17	2,530	18	41	43	17	97	72.38	B	100 M	Dada	2.40	
101	Fido	Sermada Kadet	3	18	2,460	12	41	43	19	85	64.25	B	100 M	Dada	2.42	
102	Imar	Sermada Kadet	3	19	2,280	10	41	43	19	82	58.38	C	80 M	Dada	2.39	
103	Jofir	Sermada Kadet	3	11	2,610	14	41	43	19	87	69.38	B	100 M	Dada	2.45	
104	Moc	Sermada Kadet	3	13	2,450	11	41	43	18	86	65	B	100 M	Dada	2.47	
105	Muh	Sermada Kadet	3	14	2,820	14	41	43	19	88	75.38	B	100 M	Dada	2.48	
106	Muh Ram	Sermada Kadet	3	15	2,320	10	41	43	19	83	59.75	C	100 M	Dada	2.50	
107	Muh Sapu	Sermada Kadet	3	16	2,420	1	41	43	19	72	56.88	C	75 M	Dada	2.34	
108	Naw	Sermada Kadet	3	17	2,370	5	41	43	19	77	58.25	C	100 M	Dada	2.36	
109	Sarr	Sermada Kadet	3	11	2,830	18	41	43	17	98	81	BS	100 M	Dada	2.34	
TEKNIK INFORMATIKA																
110	Adri Panc	Sermada Kadet	3	11	2,670	6	41	43	18	81	67.88	B	100 M	Dada	2.50	
111	Akhi	Sermada Kadet	3	12	2,220	10	41	42	18	84	58	C	100 M	Dada	2.39	
112	Alfia	Sermada Kadet	3	13	2,080	5	37	43	18	73	48.38	C	25 M	Bebas	1.16	
113	Arys	Sermada Kadet	3	14	1,770	0	24	20	19	29	18.38	KS	50 M	Bebas	1.14	
114	Emil	Sermada Kadet	3	16	2,440	5	41	43	18	79	61.13	B	40 M	Bebas	1.35	
115	Gan	Sermada Kadet	3	18	2,330	5	41	43	19	77	57.5	C	100 M	Dada	2.49	
116	Guff	Sermada Kadet	3	19	2,440	10	41	43	18	86	64.5	B	60 M	Bebas	2.38	

Gambar 3. 2 Contoh Nilai Evaluasi Garjas Cohort 1 Putra
 Sumber: Arsip Resimen Korps Kadet Mahasiswa

KIMIA																
67	Artar	Sermada Kadet	3	4	1,960	56	41	28	24	80	60.5	C	5 M	Bebas	0.22	
68	Asnz	Sermada Kadet	3	5	1,700	56	42	28	24	81	50.88	C	5 M	Bebas	0.25	
69	Desi	Sermada Kadet	3	6	2,320	66	42	28	20	93	82.13	BS	80 M		3.10	
70	Dita	Sermada Kadet	3	7	1,970	63	39	28	20	90	65.75	B	5 M	Bebas	0.21	
71	Fani	Sermada Kadet	3	9	2,000	63	42	28	20	94	68.88	B	100 M	Dada	2.48	
72	Fide	Sermada Kadet	3	0	1,840	63	30	28	20	80	55.75	C	50 M	Bebas	2.50	
73	Muti	Sermada Kadet	3	4	1,660	63	42	28	22	88	53.5	C	50 M	Bebas	2.49	
74	Patri Bolia	Sermada Kadet	3	6	1,710	60	42	28	22	87	54.38	C	5 M	Bebas	0.23	
75	Ratli	Sermada Kadet	3	7	1,810	65	42	28	22	88	59.13	C	20 M	Bebas	0.45	
76	Rezi	Sermada Kadet	3	8	2,000	53	42	28	22	84	64	B	5 M	Bebas	0.19	
77	Riya	Sermada Kadet	3	9	2,150	63	42	28	20	93	74.25	B	75 M	Dada	2.41	
78	Shah Juliar	Sermada Kadet	3	0	2,390	62	42	28	19	95	86.38	BS	100 M	Dada	2.47	
79	Siti Lubis	Sermada Kadet	3	1	1,840	60	42	28	20	92	61.75	B	100 M	Dada	2.47	
80	Sop	Sermada Kadet	3	2	1,990	58	42	28	21	90	66.38	B	25 M	Bebas	0.56	
81	Tiar	Sermada Kadet	3	4	1,990	65	42	28	19	96	69.38	B	25 M	Bebas	0.56	
82	Winc	Sermada Kadet	3	5	2,350	60	42	28	22	87	80.5	B	100 M	Dada	2.50	
MATEMATIKA																
83	Anni	Sermada Kadet	3	2	1,980	62	42	28	23	86	64.25	B	100 M	Dada	2.46	
84	Devi	Sermada Kadet	3	6	1,650	47	42	28	22	82	50.25	C	5 M	Bebas	0.28	
85	Divi	Sermada Kadet	3	7	1,710	59	42	28	21	89	55.5	C	5 M	Bebas	0.20	
86	Filtri	Sermada Kadet	3	2	2,200	68	42	28	19	96	77.75	B	90 M	Dada	2.52	
87	Ghin	Sermada Kadet	3	3	1,790	54	42	28	22	86	56.88	C	100 M	Dada	2.48	
88	Kirar	Sermada Kadet	3	9	1,590	55	39	28	28	68	41.5	C	25 M	Bebas	0.33	
89	Muti	Sermada Kadet	3	2	2,110	55	37	28	20	83	67.63	B	25 M	Bebas	0.35	
TEKNIK ELEKTRO																
90	Andr	Sermada Kadet	3	6	1,830	650	42	28	21	91	60.88	C	100 M	Dada	2.45	
91	Ang	Sermada Kadet	3	7	2,080	63	42	28	20	94	71.75	B	100 M	Dada	2.47	
92	Pas	Sermada Kadet	3	2	2,320	61	38	28	21	85	77.75	B	100 M	Dada	2.33	
TEKNIK INFORMATIKA																
93	Ayur	Sermada Kadet	3	5	2,040	63	42	28	19	97	71.88	B	35 M	Bebas	1.35	
94	Envir	Sermada Kadet	3	7	2,190	63	42	28	19	97	78.38	B	100 M	Dada	2.44	
95	Mad Pang	Sermada Kadet	3	2	2,100	63	42	28	19	95	73	B	25 M	Bebas	0.48	
96	Nan	Sermada Kadet	3	8	2,100	63	42	28	20	94	72.38	B	100 M	Dada	2.48	

Gambar 3. 3 Contoh Nilai Evaluasi Garjas Cohort 1 Putri
 Sumber: Arsip Resimen Korps Kadet Mahasiswa

Sebelum melanjutkan ke tahap pemodelan, langkah penting selanjutnya adalah *pre-processing data*. Hal ini mencakup pembersihan data dari *noise* atau ketidakkonsistenan yang mungkin ada, normalisasi skala pengukuran, dan penanganan data yang hilang atau tidak lengkap. Setelah proses ini selesai, dataset yang telah dibersihkan siap untuk digunakan dalam pembangunan model klustering.

3.1.3 Persiapan Data (*Data Preparation*)

Karena banyaknya ketidakkonsistenan yang masih ada, data yang diperoleh tidak dapat digunakan secara langsung, yang dapat mengurangi akurasi pemodelan klustering yang akan digunakan dalam penelitian ini. Maka dari itu proses penyiapan data harus dilakukan menggunakan data *pre-processing* atau pra-proses data dengan cara standarisasi data:

Data tersebut akan mengalami perubahan skala yang mana kita merubah nilai dari data kedalam skala diantara 0-1. *Standar scaler* merupakan nama lainnya:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3.1)$$

Persamaan tersebut merupakan formula dari normalisasi dengan X_{max} adalah nilai maximum dari interval data yang dimiliki, dan X_{min} adalah minimumnya. Adapun atribut yang akan dipakai untuk klustering pada penelitian ini:

a. Indeks Prestasi Kumulatif:

Atribut ini dipilih karena sebagai indikator keberhasilan akademik dan dedikasi dalam pembelajaran.

b. Nilai Kesegaran Jasmani:

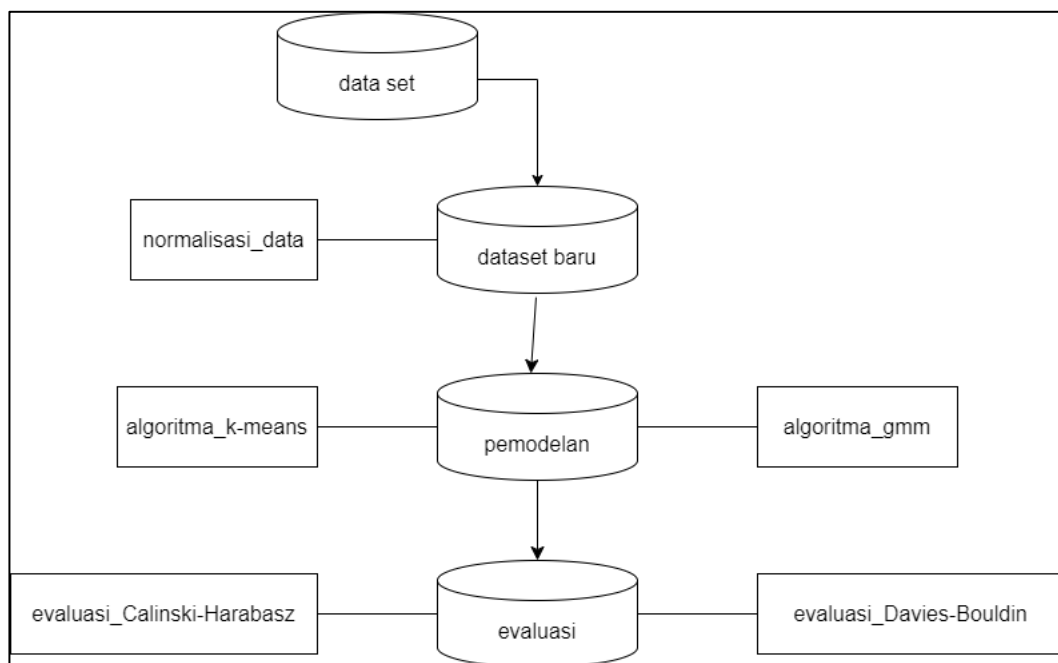
Atribut ini terpilih karena berkorelasi dengan kesiapan komponen cadangan dalam konteks mengukur kondisi fisik di mana memberikan gambaran objektif tentang tingkat kebugaran jasmani fisik. Terdiri dari tujuh aspek penilaian.

c. Jenis Kelamin:

Atribut ini dipilih karena akan menjadikan pembanding informasi yang terdiri dua variabel, yaitu lak-laki dan perempuan.

3.1.4 Pemodelan (*Modelling*)

Data yang telah melalui proses persiapan akan digunakan untuk melatih model klustering. Tahap ini dilakukan dengan menggunakan algoritma *K-Means* dan *Gaussian Mixture Models* sehingga dapat digambarkan sebagai berikut:



Gambar 3. 4 Pemodelan Penelitian

Sumber: diolah oleh peneliti

Tujuan pemodelan ini ialah mencari algoritma terbaik untuk melakukan klustering agar dapat mengoptimalkan kesiapan komponen cadangan.

3.1.5 Evaluasi (*Evaluation*)

Evaluasi terhadap pemodelan klustering akan menggunakan dua indeks validitas; *davies-bouldin* dan *calinski-harabasz* untuk menguji kinerja

kedua algoritma tersebut. Berdasarkan indeks *davies-bouldin* semakin kecil nilai evaluasinya maka itu merupakan indikator pengelompokan yang baik (Umagapi et al., 2023), berbanding terbalik dengan *calinski-harabasz* untuk hasil kluster yang lebih baik harus semakin besar (Doi et al., 2023). Sehingga akan diperoleh model klustering yang terbaik.

3.1.6 Penyebaran (*Deployment*)

Model klustering yang telah dievaluasi siap untuk tahap *deployment*, menjadi langkah akhir dalam metode penelitian CRISP-DM. Tahap ini menandakan kesiapan model untuk diimplementasikan dalam tugas pengelompokan kesiapan dan distribusi personil komponen cadangan di Unhan RI. Hasil evaluasi model, yang disajikan dalam bentuk chart, akan menjadi dasar untuk menentukan program-program yang dapat meningkatkan kualitas sumber daya manusia dan membantu pengambilan keputusan strategis di Universitas Pertahanan Republik Indonesia.

Selain *chart* evaluasi, tahap *deployment* juga mencakup data hasil prediksi dari model klustering. Model ini dapat diuji langsung dengan *input* berupa data yang sudah dinormalisasi, seperti data angka hasil evaluasi. Data tersebut kemudian digunakan untuk pengelompokan kesiapan dan distribusi personil.

3.2 Tempat dan Waktu Penelitian

3.2.1 Tempat Penelitian

Lokasi penelitian ini akan dilakukan di Universitas Pertahanan Republik Indonesia, Kawasan IPSC (*Indonesia Peace and Security Center*) Sentul, Sukahati, Kec. Citeureup, Kabupaten Bogor, Jawa Barat.

1.2.2 Waktu Penelitian

Rencana penelitian dimulai pada Desember 2023 dan diharapkan berakhir pada Mei 2024. Tabel 3.1 di bawah menunjukkan bagaimana jadwal ini disusun.

Tabel 3. 1 Jadwal Penelitian

No	Kegiatan	2023	2024				
		Des	Jan	Feb	Mar	Apr	Mei
1.	Pembuatan Draft Proposal Skripsi						
2.	Proses Pembimbingan						
3.	Seminar Proposal Skripsi						
4.	Perbaikan Proposal Skripsi						
5.	Pengumpulan Data						
6.	<i>Pre-processing</i> Data						
7.	Pelatihan Model						
8.	Evaluasi Model						
9.	<i>Deployment Model</i>						
10.	Penyusunan Laporan Penelitian/Skripsi						
11.	Seminar Skripsi						
12.	Perbaikan Skripsi						
13.	Penyerahan Skripsi kepada Program Studi						

1.3 Alat dan Bahan

Dalam melakukan penambangan data, penelitian ini membutuhkan perangkat keras dan perangkat lunak demi menunjang kelancaran proses penambangan data, di antaranya:

3.3.1 Perangkat Keras

Berikut perangkat keras yang akan digunakan selama proses penambangan data:

- a. Processor Intel(R) Core (TM) i7-1065G7 CPU @ 1.30GHz 1.50 GHz
- b. RAM (*Random Access Memory*) 16,0 GB
- c. VGA Intel® Iris® Graphics Family
- d. Memori 512 GB
- e. *Harddisk* dengan minimal memori kosong sebanyak 100 *gigabytes*

3.3.2 Perangkat Lunak

Berikut perangkat lunak yang akan digunakan selama proses penambangan data:

- a. Sistem operasi Windows 10
- b. Bahasa pemrograman Python
- c. Kode Editor VS (*Visual Studio*) Code
- d. Kode Editor Google Collab

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil

Pada sub bab hasil ini, akan dibahas hasil dari penelitian yang membandingkan klasterisasi menggunakan algoritma *K-Means* dan *Gaussian Mixture Models (GMM)* dalam konteks kesiapan Sekolah Perwira Prajurit Karier (SEPA PK) TNI. Hasil akan mencakup gambaran data, *pre-processing*, implementasi *K-Means*, dan *GMM*, serta evaluasi hasilnya yang dapat dilihat dari tabel 4.1.

Tabel 4. 1 Hasil evaluasi kinerja metode K-Means dan Gaussian Mixture Models

Model	<i>Silhouette_Score</i>	<i>Calinski_Harabasz_Score</i>	<i>Davies_Bouldin_Index</i>
<i>KMeans</i>	0.349682	598.6378	1.185072
<i>Gaussian Mixture Models</i>	0.283797	431.0565	1.633811

Sumber: Hasil pengolahan

Menurut hasil pengukuran metrik evaluasi *clustering*, kedua model *K-Means* dan *Gaussian Mixture Models (GMM)* menghasilkan nilai-nilai sebagai berikut: Model *K-Means* menghasilkan skor *silhouette* sebesar 0,349682, skor *Calinski-Harabasz* sebesar 598.63779, dan indeks *Davies-Bouldin* sebesar 1.185072. *Gaussian Mixture Models* menghasilkan skor *silhouette* sebesar 0.283797, skor *Calinski-Harabasz* sebesar 431.056536, dan indeks *Davies-Bouldin*.

Hasil pengamatan menunjukkan bahwa model *K-Means* unggul atas *GMM* dalam hal *clustering* ini. *Silhouette score* yang lebih tinggi pada *K-Means* menunjukkan bahwa *cluster* yang dihasilkan oleh *K-Means* memiliki separasi dan kepadatan *cluster* yang lebih tinggi. Data poin biasanya lebih

dekat dengan *cluster* mereka sendiri daripada *cluster* lainnya, seperti yang ditunjukkan oleh nilai yang mendekati 1.

Untuk *K-Means*, skor *Calinski-Harabasz*, yang juga lebih tinggi, menunjukkan bahwa *cluster* yang dihasilkan oleh *K-Means* memiliki distribusi yang lebih baik, dengan varian yang lebih kecil dalam *cluster* dan varian yang lebih besar antar *cluster*. Skor ini mengukur rasio antara jumlah dispersi antara *cluster* dan jumlah dispersi dalam *cluster*, dan nilai yang lebih tinggi menunjukkan *cluster* yang lebih kompak dan terpisah.

Skor ini menunjukkan bahwa dalam situasi ini, *K-Means* mempartisi data ke dalam *cluster* yang berbeda dengan lebih baik daripada *GMM*. Meskipun *GMM* memiliki kemampuan untuk menangani distribusi data yang lebih kompleks dan heterogen, struktur data dalam situasi ini mungkin lebih sesuai dengan asumsi bentuk dan ukuran *cluster K-Means*. Oleh karena itu, berdasarkan metrik evaluasi yang digunakan, *K-Means* memberikan cluster yang lebih jelas, terdefinisi, dan terpisah dengan baik. Secara keseluruhan, dari hasil pengukuran ini, disimpulkan bahwa *K-Means* adalah metode yang lebih tepat untuk *clustering* data ini dibandingkan dengan *GMM*. Evaluasi yang dilakukan dengan ketiga metrik tersebut *Silhouette score*, *Calinski-Harabasz score*, dan *Davies-Bouldin Index* konsisten menunjukkan bahwa *K-Means*

4.2 Pembahasan

Untuk lebih memahami hasil pada penelitian ini, selanjutnya akan dijelaskan kembali langkah-langkah yang dilakukan sesuai dengan CRISP-DM.

4.2.1 Pemahaman Bisnis (*Business Understanding*)

Pada fase Pemahaman Bisnis (*Business Understanding*) dalam metodologi CRISP-DM yang digunakan dalam penelitian ini, tujuan utama adalah untuk mengevaluasi efektivitas algoritma klasterisasi dalam mengukur kesiapan komponen cadangan di Universitas Pertahanan

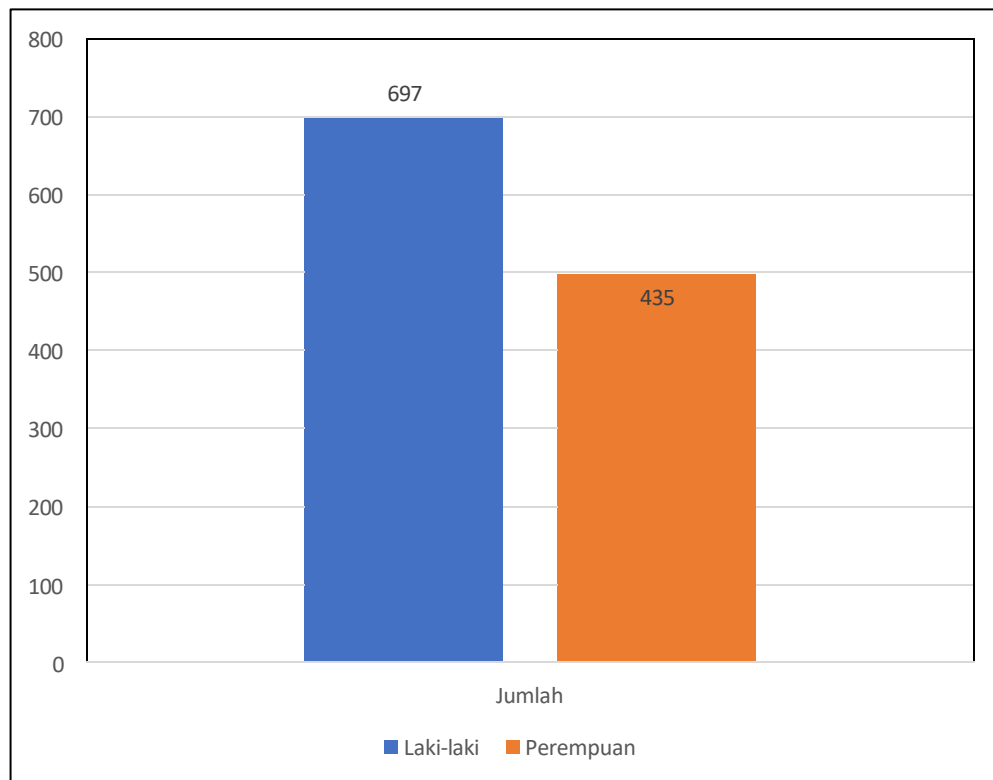
Republik Indonesia (Unhan RI). Penelitian ini berfokus pada dua algoritma utama yaitu *K-Means* dan *Gaussian Mixture Models (GMM)*.

Tujuan dari penelitian ini adalah untuk menemukan solusi yang lebih akurat dalam pengukuran kesiapan tersebut. Untuk mencapai tujuan ini, penelitian akan dilakukan melalui pengembangan dan implementasi algoritma klasterisasi dengan menggunakan bahasa pemrograman Python pada dataset yang relevan. Evaluasi hasil klasterisasi akan dilakukan menggunakan penilaian validasi seperti *Davies-Bouldin Index* dan *Calinski-Harabasz Index*.

Pendekatan ini bertujuan untuk menghasilkan algoritma klasterisasi yang handal dan dapat diandalkan untuk tujuan pengukuran kesiapan komponen cadangan di Unhan RI, yang secara langsung berkontribusi terhadap peningkatan efektivitas program pembinaan dan pelatihan kadet mahasiswa, serta mendukung keamanan nasional.

4.2.2 Pemahaman Data (*Data Understanding*)

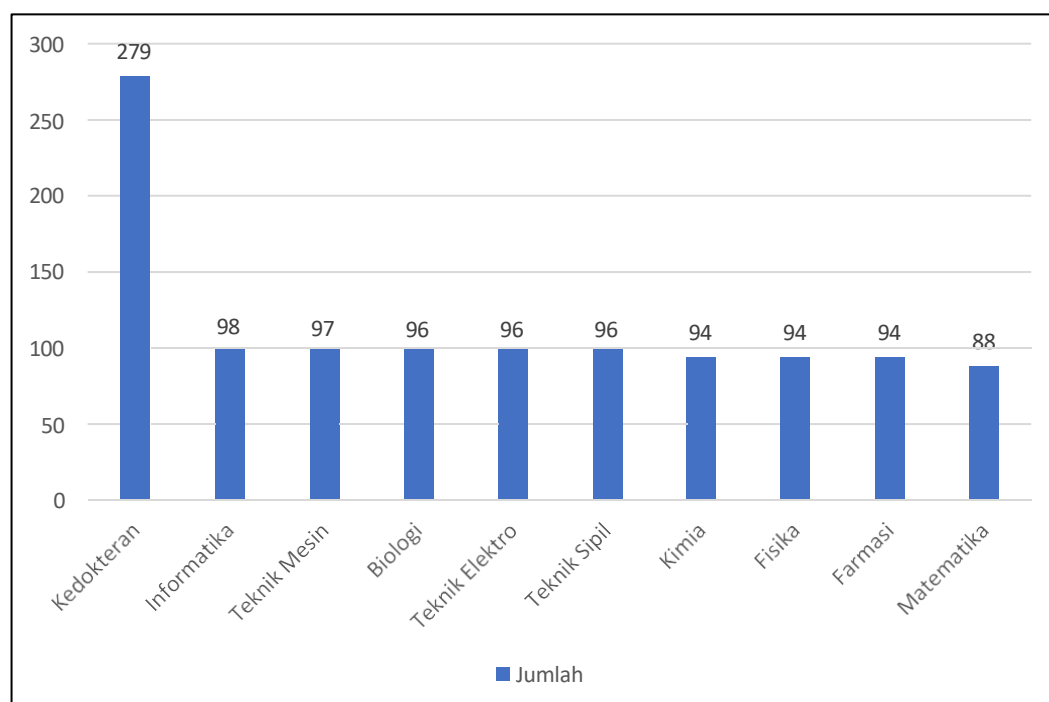
Data yang digunakan dalam penelitian ini diperoleh dari Universitas Pertahanan Republik Indonesia, dengan total 1132 data. Data ini mencakup beragam variabel yang relevan untuk mengukur kesiapan Sekolah Perwira Prajurit Karier (SEPA PK) TNI, termasuk pada aspek fisik, mental, dan kemampuan akademik. Data tersebut telah menjalani proses pengumpulan yang cermat dan memiliki representasi yang baik terhadap populasi yang diteliti, memungkinkan analisis yang mendalam untuk mendukung tujuan penelitian ini.



Gambar 4. 1 Sebaran data berdasarkan jenis kelamin

Sumber: Hasil analisis peneliti

Berdasarkan hasil analisis, data yang diperoleh dari Universitas Pertahanan Republik Indonesia memiliki dua aspek utama yang diperhatikan: jenis kelamin dan program studi. Dalam hal jenis kelamin, tertera pada Gambar 4.1 terdapat sebanyak 697 data laki-laki dan 435 data perempuan. Sementara itu, pada Gambar 4.2 dalam hal program studi, terdapat variasi jumlah data di setiap program studi, dengan Kedokteran memiliki jumlah terbanyak (279), diikuti oleh Informatika (98), Teknik Mesin (97), dan lainnya.



Gambar 4. 2 Sebaran data berdasarkan program studi

Dengan demikian, data ini memberikan gambaran yang cukup representatif tentang distribusi jenis kelamin dan program studi di lingkungan Universitas Pertahanan Republik Indonesia. Analisis lebih lanjut terhadap data ini diharapkan dapat memberikan wawasan yang lebih dalam mengenai kesiapan Sekolah Perwira Prajurit Karier (SEPA PK) TNI dalam konteks yang lebih spesifik.

4.2.3 Persiapan Data (*Data Preparation*)

Beberapa proses dilakukan untuk mempersiapkan data sebelum dilakukan klasterisasi terlihat pada Tabel 4.2. Proses-proses tersebut meliputi penanganan *missing value*, transformasi data, dan standarisasi data. Pertama, langkah penanganan *missing value* dilakukan untuk mengatasi nilai yang kosong dalam dataset yang dapat memengaruhi hasil analisis. Selanjutnya, transformasi data dilakukan untuk memastikan distribusi data sesuai dengan asumsi model yang akan digunakan. Terakhir, standarisasi data dilakukan untuk menormalkan rentang nilai dari setiap fitur sehingga memudahkan dalam perbandingan dan interpretasi hasil

klasterisasi. Proses ini merupakan langkah awal yang penting untuk memastikan kualitas dan keandalan data sebelum dilakukan analisis lebih lanjut menggunakan algoritma *K-Means* dan *Gaussian Mixture Models*.

Tabel 4. 2 Data sampel

No.	nim	jk	lari	.. .	ren_wa k	ipk	prodi	cohor t
1	32020010 1001	laki-laki	2150	.. .	2.4	3.5	Kedokteran	1
2	32020010 1005	laki-laki	2170	.. .	3.05	3.3	Kedokteran	1
3	32020010 1006	laki-laki	2520	.. .	3.4	3.51	Kedokteran	1
4	32020010 1007	laki-laki	2020	.. .	1.2	3.29	Kedokteran	1
5	32020010 1008	laki-laki	2170	.. .	2.42	3.82	Kedokteran	1
...
1128	32023040 4005	perempuan	2280	.. .	2.38	3.68	Teknik Sipil	4
1129	32023040 4009	perempuan	2133	.. .	0.4	3.66	Teknik Sipil	4
1130	32023040 4017	perempuan	2320	.. .	1.12	3.59	Teknik Sipil	4
1131	32023040 4020	perempuan	2150	.. .	1.3	3.65	Teknik Sipil	4
1132	32023040 4023	perempuan	3120	.. .	1	3.41	Teknik Sipil	4

4.2.3.1 Missing Values

Penanganan *missing value* dilakukan dengan menggunakan teknik *mean* atau nilai rata-rata. Dua data kosong yang terdeteksi terdapat pada variabel hasil tes *shuttle run* (*stl_run*) dan waktu tes renang (*ren_wak*). Langkah pertama adalah mengidentifikasi variabel yang memiliki nilai kosong, kemudian nilai-nilai tersebut diganti dengan nilai rata-rata dari variabel tersebut. Proses ini memastikan bahwa data yang kosong diisi dengan nilai yang mewakili karakteristik umum dari variabel yang bersangkutan, sehingga mempertahankan integritas dan kegunaan data untuk analisis selanjutnya. Bisa dilihat dari Gambar 4.3 *syntax* pada proses ini.

```
pd.DataFrame(data.isna().sum())  
  
data['stl_run'].fillna(data['stl_run'].mean(), inplace=True)  
data['ren_wak'].fillna(data['ren_wak'].mean(), inplace=True)  
data.drop(columns=['No. ', 'nim', 'cohort'], inplace=True)
```

Gambar 4. 3 Syntax tahapan pembersihan data

Sumber: Hasil pengolahan

Proses pengolahan data dimulai dengan mengidentifikasi jumlah nilai yang hilang (*missing values*) pada setiap kolom dalam dataset. Hal ini dilakukan dengan menggunakan fungsi `data.isna().sum()`, yang akan menghasilkan *DataFrame* berisi jumlah nilai hilang untuk setiap kolom. Setelah itu, langkah selanjutnya adalah menangani nilai-nilai hilang pada kolom tertentu. Kolom *stl_run* dan *ren_wak* memiliki nilai hilang yang kemudian diisi dengan menggunakan nilai rata-rata (*mean*) dari masing-masing kolom tersebut. Proses pengisian ini dilakukan dengan metode *fillna()*, yang menggantikan nilai hilang dengan nilai rata-rata dari kolom yang bersangkutan.

Selanjutnya, beberapa kolom yang dianggap tidak relevan untuk analisis lebih lanjut dihapus dari dataset. Kolom-kolom yang dihapus adalah *No.*, *nim*, dan *cohort*. Penghapusan ini dilakukan menggunakan metode *drop(columns=[...])*. Kolom *No.*, merupakan nomor urut yang tidak memiliki pengaruh terhadap analisis, *nim* adalah nomor induk mahasiswa yang merupakan identitas unik dan tidak berpengaruh dalam klasterisasi, dan *cohort* adalah penyebutan angkatan yang mengindikasikan kelompok mahasiswa tertentu namun tidak relevan untuk analisis ini. Dengan demikian, dataset yang dihasilkan menjadi lebih bersih dan siap untuk tahap analisis selanjutnya.

Tabel 4. 3 Hasil implementasi nilai *mean* dalam mengatasi *missing value*.

Variabel	Jumlah data null awal	Hasil implementasi nilai <i>mean</i>
No.	0	0
nim	0	0
jenis kelamin	0	0
lari	0	0
<i>pull_up</i>	0	0
<i>sit_up</i>	0	0
<i>push_up</i>	0	0
<i>stl_run</i>	1	0
ren_jar	0	0
ren_wak	1	0
ipk	0	0
prodi	0	0
cohort	0	0

Sumber: Hasil pengolahan

Tabel 4.3 menunjukkan hasil analisis dan perbaikan missing value pada variabel yang digunakan dalam penelitian. Variabel-variabel ini mencakup data akademik dan fisik mahasiswa meliputi No. (nomor urut data), nim (nomor induk mahasiswa) yang merupakan identifikasi unik untuk setiap mahasiswa, jenis kelamin atau *jk* (dikodekan menggunakan Label Encoder), nilai hasil tes lari, *pull_up*, *sit_up*, dan *push_up*. Selain itu, variabel *stl_run* mengukur ketangkasan melalui tes shuttle run, dengan satu data yang missing. Ren_jar dan ren_wak mencatat nilai hasil tes jarak jauh renang dan kecepatan renang, dengan satu data missing pada ren_wak. ipk mencerminkan prestasi akademik mahasiswa, sementara prodi dan cohort menunjukkan program studi yang diikuti serta angkatan atau kelompok mahasiswa berdasarkan tahun masuk. Implementasi nilai mean

digunakan untuk mengatasi missing value, memastikan dataset lengkap dan siap untuk analisis lebih lanjut.

4.2.3.2 Transformasi Data

Metode *label encoder* diterapkan pada variabel jenis kelamin dan program studi. *Label encoder* adalah teknik yang digunakan untuk mengubah nilai-nilai dalam suatu variabel kategorikal menjadi nilai numerik, sehingga memungkinkan penggunaan variabel tersebut dalam analisis yang melibatkan model-model matematika atau statistik. Dalam konteks ini, variabel jenis kelamin dan program studi diubah menjadi nilai numerik sesuai dengan *label encoder*, memudahkan dalam proses analisis dan klasterisasi data. Proses ini meningkatkan kualitas data dan mempersiapkannya untuk langkah-langkah analisis berikutnya. Berikut *syntax* pada proses ini bisa terlihat pada Gambar 4.4.

```
label_encoder = LabelEncoder()

# Encode 'jk' column
data['jk'] = label_encoder.fit_transform(data['jk']) + 1 # Start
encoding from 1

# Encode 'prodi' column
data['prodi'] = label_encoder.fit_transform(data['prodi']) + 1 #
Start encoding from 1
data.info()
```

Gambar 4. 4 Syntax tahapan *label encoding*

Sumber : Hasil pengolahan

Proses pengolahan data dilanjutkan dengan transformasi variabel kategorikal menjadi bentuk numerik agar dapat digunakan dalam analisis klasterisasi. Dalam hal ini, dua kolom yang memiliki data kategorikal, yaitu

jk (jenis kelamin) dan *prodi* (program studi), akan diubah menjadi bentuk numerik menggunakan metode *Label encoding*.

Pertama, sebuah objek *LabelEncoder* dari *library sklearn* dibuat. *LabelEncoder* adalah alat yang mengubah nilai kategorikal menjadi nilai numerik. Proses *encoding* diawali dengan kolom *jk*, di mana nilai-nilai kategorikal diubah menjadi angka dimulai dari 1. Hal ini dilakukan dengan metode *fit_transform()* yang diikuti dengan penambahan nilai 1 untuk memastikan bahwa *encoding* dimulai dari angka 1, bukan 0. Proses yang sama diterapkan pada kolom *prodi*.

Dengan melakukan *encoding* ini, variabel kategorikal *jk* dan *prodi* sekarang diubah menjadi variabel numerik yang dapat digunakan dalam model klasterisasi seperti *K-Means* dan *Gaussian Mixture Models*. Proses ini memastikan bahwa model dapat mengenali dan bekerja dengan data kategorikal yang telah diubah menjadi bentuk yang sesuai. Informasi tentang dataset yang telah diperbarui dapat dilihat dengan menggunakan metode *data.info()*, yang akan menampilkan ringkasan struktur dataset, termasuk tipe data dari setiap kolom dan jumlah nilai *non-null*. Hasil *label encoding* variabel jenis kelamin dan program studi dipaparkan pada Tabel 4.4 sebagai berikut.

Tabel 4. 4 Hasil implementasi *Label encoder*

Variabel	Parameter	Hasil <i>encoding</i>
prodi	Biologi	1
	Farmasi	2
	Fisika	3
	Informatika	4
	Kedokteran	5
	Kimia	6
	Matematika	7
	Teknik Elektro	8

Variabel	Parameter	Hasil <i>encoding</i>
	Teknik Mesin	9
	Teknik Sipil	10
jenis kelamin	Laki-laki	1
	Perempuan	2

Sumber : Hasil pengolahan

4.2.3.3 *Outlier*

Pada proses ini, data hasil *label encoding* menghasilkan variabel dengan tipe data numerik, yang artinya data siap diimplementasikan metode *K-Means* dan *Gaussian Mixture Models (GMM)*. Namun, sebelum itu, perlu dilakukan pengecekan untuk nilai *outlier*. *Outlier* adalah data yang memiliki nilai yang sangat berbeda dari sebagian besar data lainnya dalam satu set data. *Outlier* bisa sangat tinggi atau sangat rendah dibandingkan dengan nilai lainnya. Secara umum, *outlier* dapat terjadi karena berbagai alasan, termasuk kesalahan pengukuran, variasi percobaan, atau kondisi yang sangat berbeda dari pengamatan lain dalam dataset.

Outlier dapat memiliki pengaruh signifikan terhadap analisis data dan hasil model *machine learning*. Beberapa pengaruh negatif dari *outlier* termasuk distorsi statistik deskriptif, di mana *outlier* dapat mengubah *mean*, *standard deviation*, dan metrik lainnya, membuatnya kurang representatif untuk dataset secara keseluruhan. Selain itu, *outlier* dapat mempengaruhi model *machine learning*, terutama algoritma seperti *K-Means* yang sensitif terhadap *outlier* karena menggunakan jarak sebagai ukuran. *Outlier* dapat menarik *centroid* lebih dekat ke mereka, yang dapat mengakibatkan pembentukan *cluster* yang tidak wajar. Pengaruh lainnya adalah pada visualisasi data, di mana *outlier* dapat membuat visualisasi seperti *boxplot* dan *scatter plot* menjadi kurang informatif atau *misleading*. Berikut *syntax* dari proses ini.

```

import pandas as pd
import numpy as np
import Seaborn as sns
import Matplotlib.pyplot as plt
from scipy.stats import skew

# Kolom numerik untuk pemeriksaan
numerical_cols = ['jk', 'lari', 'pull_up', 'sit_up', 'push_up',
                  'stl_run', 'ren_jar', 'ren_wak', 'ipk', 'prodi']

Outlier_counts = {}

# Plotting Outliers using boxplot
plt.figure(figsize=(15, 10))
for i, col in enumerate(numerical_cols, 1):
    plt.subplot(3, 4, i)
    sns.boxplot(x=data[col])
    plt.title(f'Boxplot of {col}')

# Menghitung jumlah Outlier
Q1 = data[col].quantile(0.25)
Q3 = data[col].quantile(0.75)
IQR = Q3 - Q1
Outliers = data[(data[col] < (Q1 - 1.5 * IQR)) | (data[col] >
(Q3 + 1.5 * IQR))]
num_Outliers = len(Outliers)
Outlier_counts[col] = num_Outliers
plt.xlabel(f'Number of Outliers: {num_Outliers}')
plt.tight_layout()

```



```
plt.show()
```

Gambar 4. 5 Syntax tahapan menampilkan nilai *Outlier*

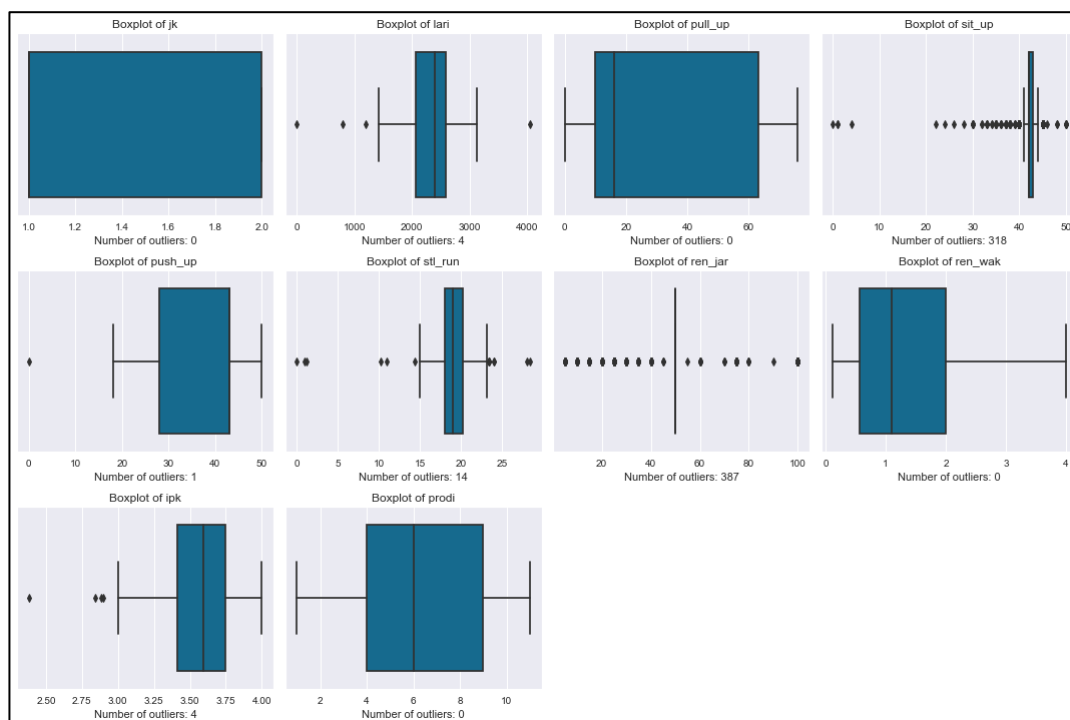
Sumber : Hasil pengolahan

Proses pada Gambar 4.5 berikut ini digunakan untuk mengidentifikasi dan memvisualisasikan nilai *outlier* pada dataset melalui penggunaan *boxplot* untuk setiap variabel numerik. Pertama-tama, kolom-kolom numerik yang akan diperiksa didefinisikan dalam variabel *numerical_cols*, yang mencakup kolom seperti 'jk', 'lari', 'pull_up', 'sit_up', 'push_up', 'stl_run', 'ren_jar', 'ren_wak', 'ipk', dan 'prodi'.

Selanjutnya, *dictionary outlier_counts* dibuat untuk menyimpan jumlah *outlier* yang terdeteksi di setiap kolom. Dengan menggunakan *Matplotlib* dan *Seaborn*, ukuran plot ditentukan untuk memastikan visualisasi yang jelas dan rapi. Dalam *loop for*, setiap kolom numerik diproses satu per satu.

Untuk setiap kolom, *boxplot* dibuat menggunakan *Seaborn* (*sns.boxplot*). Di dalam setiap *subplot*, nilai Q1 (kuartil pertama) dan Q3 (kuartil ketiga) dihitung. IQR (*interquartile range*) kemudian dihitung sebagai selisih antara Q3 dan Q1. Nilai *outlier* diidentifikasi sebagai nilai yang berada di bawah ($Q1 - 1.5 * IQR$) atau di atas ($Q3 + 1.5 * IQR$). Jumlah *outlier* dihitung dan disimpan dalam *dictionary outlier_counts* untuk setiap kolom.

Pada *boxplot*, jumlah *outlier* ditampilkan sebagai label pada sumbu x. Setelah *loop* selesai, *layout plot* ditata menggunakan *plt.tight_layout()* untuk memastikan bahwa *subplots* tidak saling tumpang tindih. Akhirnya, plot ditampilkan dengan *plt.show()*, memberikan visualisasi yang jelas dari distribusi data dan nilai *outlier* untuk setiap variabel. Berikut hasil identifikasi nilai *outlier* dijelaskan pada Gambar 4.6 di bawah.



Gambar 4. 6 Hasil identifikasi nilai *outlier* tiap variabel

Sumber : Hasil pengolahan

Dari hasil identifikasi, terdapat beberapa variabel yang memiliki *outlier* dalam jumlah signifikan. Misalnya, variabel *lari* memiliki 4 *outlier*, yang berarti ada empat nilai yang cukup ekstrem dibandingkan dengan distribusi umum data tersebut. Variabel *sit_up* menunjukkan jumlah *outlier* yang sangat besar, yaitu 318, yang dapat mengindikasikan adanya data yang tidak biasa atau kesalahan pencatatan data dalam jumlah besar. Variabel *push_up* hanya memiliki 1 *outlier*, menunjukkan distribusi data yang relatif lebih konsisten.

Untuk variabel *stl_run*, terdapat 14 *outlier*, yang menandakan adanya beberapa nilai yang berada jauh di luar rentang normal. Variabel *ren_jar* memiliki jumlah *outlier* yang signifikan, sebanyak 387, yang menunjukkan adanya variasi ekstrem dalam data tersebut. Variabel *ipk* juga memiliki 4 *outlier*, yang mungkin mencerminkan beberapa nilai yang sangat tinggi atau sangat rendah dalam nilai akademik. Beberapa variabel seperti

jk, *pull_up*, *ren_wak*, dan *prodi* tidak memiliki *outlier* sama sekali, menunjukkan distribusi data yang lebih normal dan konsisten tanpa adanya nilai yang ekstrem.

Untuk mengatasi *outlier*, ada beberapa pendekatan yang bisa diambil, salah satunya adalah dengan mengganti nilai *outlier* dengan nilai *mean*. Dengan mengganti nilai *outlier* dengan *mean*, dapat menjaga kestabilan statistik deskriptif dan meningkatkan performa model *machine learning*. Langkah-langkah untuk mengatasi *outlier* dalam dataset meliputi identifikasi *outlier* menggunakan metode seperti IQR (*Interquartile range*) atau *Z-score*, kemudian mengganti nilai *outlier* dengan *mean*. Berikut Gambar 4.7 yang berisikan *syntax* pada proses ini.

```
# Mengatasi Outlier Menggunakan Nilai Mean
def replace_Outliers_with_mean(df, column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    df[column] = np.where(df[column] < lower_bound,
                          df[column].mean(), df[column])
    df[column] = np.where(df[column] > upper_bound,
                          df[column].mean(), df[column])
    return df

# Penerapan Mengatasi Outlier dengan Nilai Mean
for col in data.columns:
    data = replace_Outliers_with_mean(data, col)
```

```
# Plotting Boxplot setelah Mengatasi Outlier dengan Nilai Mean
plt.figure(figsize=(15, 10)) # Mengatur ukuran figure
for i, col in enumerate(data.columns, 1):
    plt.subplot(3, 4, i)
    sns.boxplot(x=data[col], orient='v', palette='Set2')
    plt.title(f'Boxplot of {col}')
    plt.tight_layout()
    plt.show()
```

Gambar 4. 7 Syntax tahapan mengatasi nilai outlier

Sumber : Hasil pengolahan

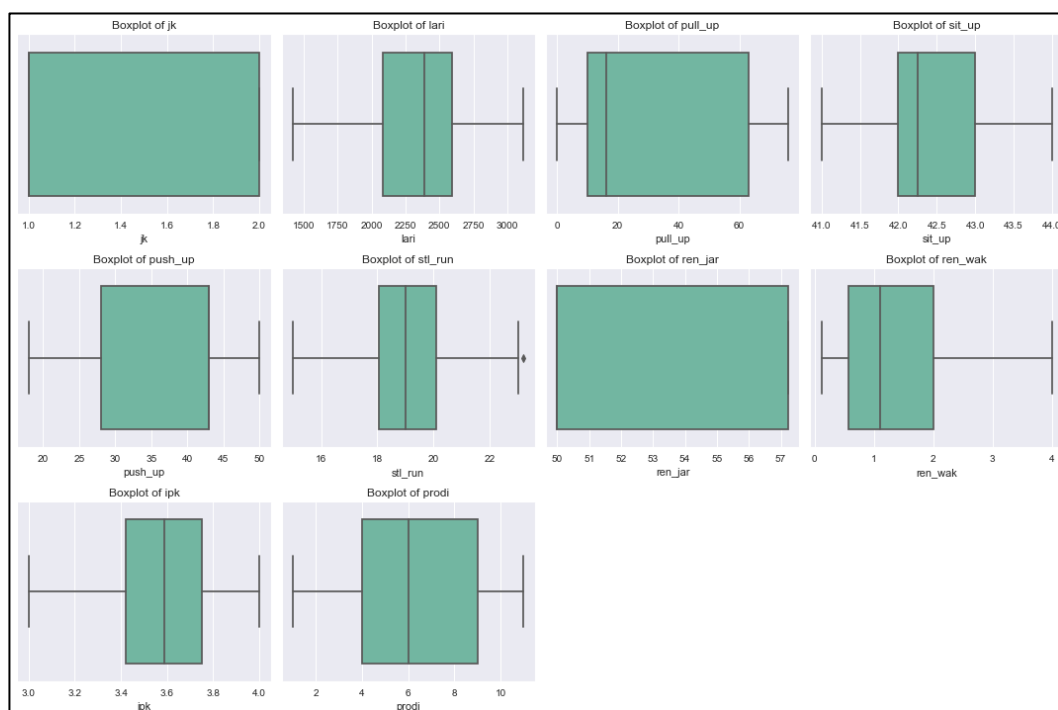
Proses berikut ini digunakan untuk mengatasi *outlier* dalam dataset dengan menggantinya menggunakan nilai *mean* dari masing-masing kolom. Pertama-tama, sebuah fungsi bernama *replace_outliers_with_mean* didefinisikan untuk menangani *outlier* pada kolom tertentu. Fungsi ini menghitung Q1 (kuartil pertama) dan Q3 (kuartil ketiga) dari kolom yang diberikan, kemudian menghitung rentang antar kuartil (IQR). Batas bawah dan atas untuk menentukan *outlier* dihitung sebagai $Q1 - 1.5 * IQR$ dan $Q3 + 1.5 * IQR$. Nilai-nilai yang berada di bawah batas bawah atau di atas batas atas diganti dengan nilai *mean* dari kolom tersebut.

Selanjutnya, fungsi *replace_outliers_with_mean* diterapkan pada setiap kolom dalam dataset menggunakan *loop for*. *Loop* ini iterasi melalui semua kolom dalam *DataFrame* data, dan setiap kolom diperbarui dengan mengganti nilai *outliernya* menggunakan nilai *mean*.

Setelah semua *outlier* ditangani, langkah selanjutnya adalah memvisualisasikan distribusi data setelah penanganan *outlier*. Ini dilakukan dengan membuat *boxplot* untuk setiap kolom dalam dataset. Ukuran *figure* diatur menggunakan `plt.figure(figsize=(15, 10))` untuk memastikan visualisasi yang rapi. Dalam *loop for* berikutnya, setiap kolom diplotkan sebagai *subplot* dalam *layout* 3x4. Untuk setiap *subplot*, *sns.boxplot*

digunakan untuk membuat *boxplot*, dengan orientasi vertikal dan palet warna 'Set2'. Judul *subplot* diberikan sesuai dengan nama kolom.

Setelah semua *subplot* dibuat, *layout plot* ditata menggunakan *plt.tight_layout()* untuk memastikan bahwa *subplot* tidak saling tumpang tindih. Akhirnya, plot ditampilkan dengan *plt.show()*, memberikan visualisasi yang jelas dari distribusi data setelah penanganan *outlier*. Berikut hasil implementasi nilai *mean* dalam menangani nilai *outlier* pada tiap variabel tertera pada Gambar 4.8 dibawah ini.



Gambar 4. 8 Hasil implementasi *mean* pada nilai *outlier*

Sumber : Hasil pengolahan

Setelah menerapkan metode penggantian nilai *outlier* dengan nilai *mean*, hasil analisis menunjukkan bahwa jumlah *outlier* pada tiap variabel telah berkurang secara signifikan atau bahkan sudah tidak ada lagi. Metode ini membantu menormalkan distribusi data dengan menggantikan nilai-nilai ekstrem yang berada jauh dari rentang normal dengan nilai rata-rata dari variabel tersebut. Sebagai contoh, variabel *lari*, yang sebelumnya memiliki

4 *outlier*, sekarang tidak lagi menunjukkan adanya nilai ekstrem setelah dilakukan penggantian dengan nilai *mean*. Demikian pula dengan variabel *sit_up*, yang memiliki 318 *outlier*, kini telah diatasi sehingga distribusinya menjadi lebih konsisten tanpa adanya nilai yang sangat ekstrem. Hal yang sama terjadi pada variabel *push_up*, *stl_run*, *ren_jar* (jarak renang), dan *ipk*, yang semuanya menunjukkan pengurangan atau hilangnya *outlier* setelah diterapkan metode ini.

4.2.3.4 Standarisasi Data

Teknik *standar scaler* digunakan untuk menormalkan rentang nilai dari setiap fitur dalam dataset. Proses ini dilakukan untuk memastikan bahwa semua fitur memiliki skala yang serupa, sehingga tidak ada fitur yang mendominasi perhitungan dalam algoritma klasterisasi. Dengan menggunakan *standar scaler*, setiap nilai fitur diubah sedemikian rupa sehingga memiliki rata-rata nol dan deviasi standar satu. Hal ini membuat data lebih mudah dibandingkan dan diinterpretasikan, serta meningkatkan kinerja algoritma klasterisasi. Proses standarisasi ini merupakan langkah penting dalam persiapan data sebelum dilakukan analisis lebih lanjut menggunakan algoritma *K-Means* dan *Gaussian Mixture Models*. Berikut *syntax* pada proses ini yang tertera pada Gambar 4.9.

```
from sklearn.preprocessing import StandardScaler
scaler= StandardScaler()
data_scaled= scaler.fit_transform(data)
```

Gambar 4. 9 Syntax tahapan standarisasi data

Sumber : Hasil pengolahan

Proses yang dilakukan pada kode berikut bertujuan untuk melakukan standarisasi data agar semua fitur dalam dataset berada dalam skala yang sama. Standarisasi ini penting karena banyak algoritma *machine learning* sensitif terhadap skala data, terutama algoritma yang

mengandalkan jarak antara data, seperti *K-Means* dan *Gaussian Mixture Models*.

Pertama, *StandardScaler* dari pustaka *sklearn.preprocessing* diimpor dan sebuah objek *scaler* dibuat dengan nama *scaler*. *StandardScaler* bekerja dengan menghapus *mean* dan menskalakan data ke unit *variance* (membuat *mean* = 0 dan standar deviasi = 1).

Selanjutnya, metode *fit_transform* dari objek *scaler* diterapkan pada dataset data. Metode ini pertama-tama menyesuaikan *StandardScaler* dengan data (menghitung *mean* dan standar deviasi untuk digunakan nanti dalam penskalaan) dan kemudian mengubah data sesuai dengan parameter yang telah dihitung. Hasilnya adalah *data_scaled*, yaitu versi terstandarisasi dari dataset asli, di mana setiap fitur memiliki *mean* = 0 dan standar deviasi = 1. Berikut hasil implementasi standarisasi data dijelaskan pada Tabel 4.5 berikut.

Tabel 4. 5 Perubahan tipe data hasil standarisasi data

Variabel	Data Type awal	Hasil Standarisasi
jenis kelamin	<i>object</i>	<i>float64</i>
lari	<i>int64</i>	<i>float64</i>
<i>pull_up</i>	<i>int64</i>	<i>float64</i>
<i>sit_up</i>	<i>int64</i>	<i>float64</i>
<i>push_up</i>	<i>int64</i>	<i>float64</i>
<i>stl_run</i>	<i>float64</i>	<i>float64</i>
ren_jar	<i>int64</i>	<i>float64</i>
<i>ren_wak</i>	<i>float64</i>	<i>float64</i>
ipk	<i>float64</i>	<i>float64</i>
prodi	<i>object</i>	<i>float64</i>
cohort	<i>int64</i>	<i>float64</i>

Sumber : Hasil pengolahan

4.2.4 Pemodelan (*Modelling*)

Tahap ini merupakan implementasi metode *K-Means* dan *Gaussian Mixture Models*, pemodelan dilakukan dengan menggunakan metode *Elbow* untuk menentukan nilai *k* terbaik. Metode *Elbow* digunakan untuk menemukan jumlah *cluster* yang optimal dalam algoritma klasterisasi, yaitu nilai *k* yang memberikan penurunan signifikan dalam varians intra-*cluster* tanpa peningkatan yang signifikan dalam varians antar-*cluster*. Nilai *k* yang dipilih berdasarkan titik “*Elbow*” dalam kurva kecepatan penurunan varians intra-*cluster* terhadap nilai *k*. Metode ini memungkinkan pemilihan nilai *k* yang optimal untuk kedua metode, *K-Means* dan *Gaussian Mixture Models*, sehingga hasil klasterisasi dapat mencerminkan struktur data dengan baik. Proses implementasi ini penting untuk memastikan bahwa klasterisasi dilakukan dengan efisien dan akurat sesuai dengan karakteristik data yang diamati.

4.2.4.1 Implementasi *K-Means*

Pada implementasi metode *K-Means* ini akan dibagi kelompok menjadi 3 *cluster*. Dengan demikian, proses klasterisasi menggunakan algoritma *K-Means* dilakukan dengan menggunakan nilai *k*=3. Pengaturan parameter yang digunakan dalam implementasi *K-Means* meliputi '*algorithm*' yang disetel ke '*lloyd*', '*init*' menggunakan metode '*K-Means++*' untuk inisialisasi *centroid*, '*max_iter*' sebanyak 300 iterasi, '*n_init*' yang disetel ke '*warn*' untuk jumlah inisialisasi *centroid*, '*random_state*' yang dibiarkan default, '*tol*' yang disetel ke 0.0001 untuk menentukan kriteria konvergensi, dan '*verbose*' yang diatur ke 0 untuk tidak menampilkan informasi proses secara detail. Berikut *syntax* pada proses ini.

```
from sklearn.cluster import KMeans
KMeans=KMeans(n_clusters=3)
KMeans_Predict=KMeans.fit_Predict(data_scaled)
```

Gambar 4. 10 Syntax tahapan implementasi metode *K-Means*

Sumber: Hasil pengolahan

Pada Gambar 4.10 di atas, kode tersebut mengimplementasikan algoritma *K-Means* untuk melakukan klasterisasi pada data yang telah distandarisasi (*data_scaled*). Tujuan dari proses ini adalah untuk mengelompokkan data ke dalam tiga *cluster* berdasarkan kesamaan karakteristik.

Pertama, mengimpor kelas *KMeans* dari pustaka *sklearn.cluster*. Kelas ini menyediakan implementasi dari algoritma *K-Means*, yang merupakan salah satu metode klasterisasi yang paling umum digunakan.

Selanjutnya, membuat sebuah *instance* dari *KMeans* dengan mengatur parameter *n_clusters* ke 3, yang berarti ingin membagi data ke dalam tiga *cluster*. Nilai *n_clusters*=3 dipilih berdasarkan hasil analisis sebelumnya menggunakan metode *Elbow*, yang menunjukkan bahwa tiga *cluster* adalah jumlah *cluster* yang optimal untuk data ini.

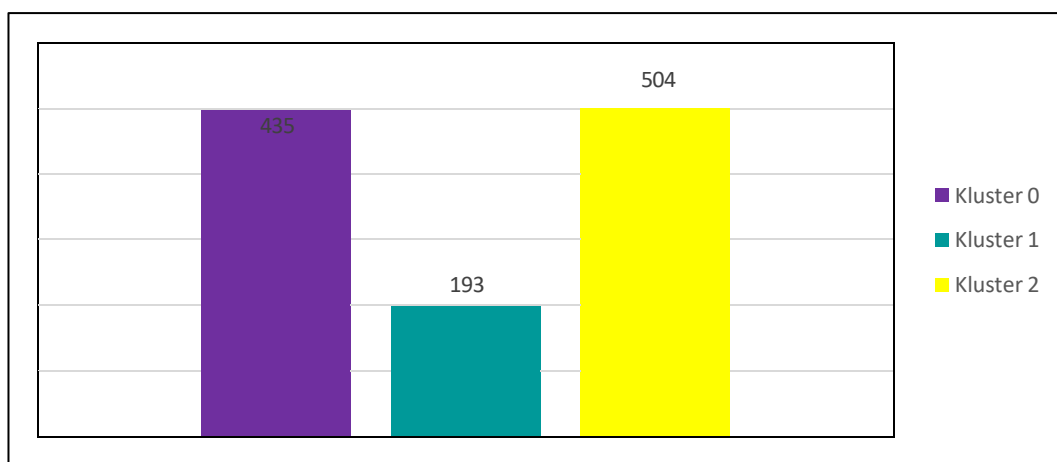
Kemudian, menggunakan metode *fit_Predict* dari objek *KMeans* untuk melatih model pada data yang telah distandarisasi (*data_scaled*) dan secara bersamaan membuat prediksi *cluster* untuk setiap titik data. Metode *fit_Predict* ini melakukan dua hal:

- a. *Fit*: Melatih model *K-Means* pada data yang diberikan, yang melibatkan penghitungan *centroid* untuk setiap *cluster*.
- b. *Predict*: Menetapkan setiap titik data ke *cluster* terdekat berdasarkan *centroid* yang dihitung.

Hasil dari *fit_Predict* disimpan dalam variabel *KMeans_Predict*, yang berisi label *cluster* untuk setiap titik data dalam dataset. Label-label ini menunjukkan *cluster* mana setiap titik data termasuk setelah proses klasterisasi.

Hasil pengamatan dari proses ini menunjukkan bahwa data berhasil dikelompokkan menjadi dua *cluster* berdasarkan fitur-fitur yang diamati.

Setiap *cluster* mewakili kelompok data yang memiliki kesamaan internal yang tinggi, sementara memiliki perbedaan yang signifikan antara satu *cluster* dengan *cluster* lainnya. Penggunaan nilai $k=3$ juga sejalan dengan hasil analisis *Elbow* yang menunjukkan penurunan yang signifikan pada nilai $k=3$. Proses klasterisasi ini memberikan pemahaman yang lebih baik tentang struktur data SEPA PK TNI, memungkinkan identifikasi pola-pola yang mendasari kesiapan para prajurit karier militer.



Gambar 4. 11 Sebaran anggota *cluster* hasil metode *K-Means*

Sumber: Hasil pengolahan

Berdasarkan Gambar 4.11 *cluster* yang terbentuk pada implementasi *K-Means*, terdapat tiga *cluster* yang dibentuk. Hasil implementasi *K-Means clustering* dengan tiga *cluster* ($k=3$) pada data menunjukkan bahwa data berhasil dibagi menjadi tiga kelompok dengan distribusi ukuran *cluster* yang bervariasi. *Cluster* pertama terdiri dari 435 data poin, *cluster* kedua terdiri dari 193 data poin, dan *cluster* ketiga terdiri dari 504 data poin.

Distribusi ini menunjukkan bahwa *cluster* ketiga adalah yang terbesar, diikuti oleh *cluster* pertama, sementara *cluster* kedua memiliki jumlah data poin paling sedikit. Perbedaan ukuran *cluster* ini mengindikasikan adanya variasi dalam karakteristik data yang berhasil ditangkap oleh algoritma *K-Means*.

Pada visualisasi *cluster* yang terbentuk digunakan *scatter plot* dengan data yang telah dikelompokkan berdasarkan hasil implementasi *K-Means*. Setiap titik dalam plot mewakili satu data observasi, yang diwarnai sesuai dengan *cluster* keanggotaannya, ditunjukkan oleh variabel '*k_means*'. Selain itu, *centroid* dari masing-masing *cluster* ditandai dengan titik merah yang lebih besar. Pemaparan visual ini memungkinkan untuk memperhatikan pola-pola yang mungkin ada di antara data dan *cluster*, serta memperoleh pemahaman visual tentang bagaimana data dikelompokkan oleh algoritma *K-Means* berdasarkan *syntax* dari proses ini.

```
plt.figure(figsize=(7, 4))
sns.scatterplot(data=data_vis, x=data_vis['lari'],
y=data_vis['pull_up'], hue='k_means', palette='viridis')
plt.scatter(centroid_KMeans['lari'], centroid_KMeans['pull_up'],
c="red", s=150, alpha=0.6)
plt.show()
```

Gambar 4. 12 Syntax tahapan visualisasi sebaran *cluster* hasil *K-Means*

Sumber: Hasil pengolahan

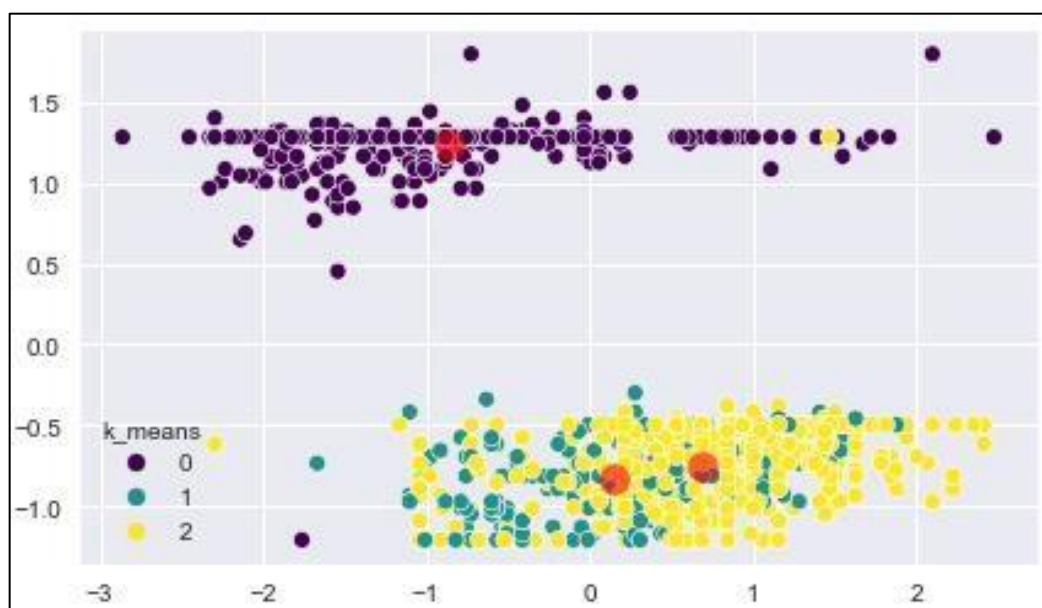
Pada Gambar 4.12 di atas, kode akan membuat visualisasi hasil klasterisasi menggunakan algoritma *K-Means* dengan memanfaatkan *scatter plot*. Tujuan dari visualisasi ini adalah untuk memberikan gambaran tentang distribusi data dalam *cluster* yang terbentuk serta posisi *centroid* dari masing-masing *cluster*.

Pertama, menggunakan *plt.figure(figsize=(7, 4))* untuk menentukan ukuran dari *figure* atau kanvas tempat plot akan ditampilkan. Ukuran yang dipilih adalah 7x4 inci, yang cukup besar untuk menampilkan data dengan jelas.

Selanjutnya, menggunakan *sns.scatterplot* dari pustaka *Seaborn* untuk membuat *scatter plot*. Parameter data diatur ke *data_vis*, yang

merupakan dataset yang ingin divisualisasikan. Parameter *hue* diatur ke '*k_means*', yang berarti titik-titik data akan diwarnai berdasarkan label *cluster* hasil dari algoritma *K-Means*. Palet warna yang digunakan adalah '*viridis*', yang memberikan gradasi warna yang jelas untuk membedakan *cluster*.

Kemudian, menambahkan *centroid* dari setiap *cluster* ke dalam plot menggunakan *plt.scatter*. *Centroid* adalah titik pusat dari setiap *cluster* yang dihitung oleh algoritma *K-Means*. Dalam kode ini, *centroid_KMeans* adalah *DataFrame* yang berisi koordinat *centroid* untuk *cluster*. Parameter *c= "red"* menentukan warna merah untuk *centroid*, *s=150* menentukan ukuran titik *centroid*, dan *alpha=0.6* menentukan transparansi titik *centroid* untuk memastikan mereka terlihat tetapi tidak terlalu mencolok. Berikut hasil *plotting* dijelaskan pada Gambar 4.13 di bawah ini.



Gambar 4. 13 Visualisasi sebaran *cluster* pada metode *K-Means*

Sumber: Hasil pengolahan

4.2.4.2 Implementasi *Gaussian Mixture Models*

Dalam penerapan *Gaussian Mixture Models (GMM)*, data akan dibagi menjadi tiga kelompok. Oleh karena itu, proses klasterisasi menggunakan algoritma *Gaussian Mixture Models* dilakukan dengan menggunakan nilai $k=3$. Pengaturan parameter yang digunakan dalam implementasi *Gaussian Mixture Models* meliputi '*covariance_type*' yang diatur ke '*full*' untuk menangani kovarian antara fitur, '*init_params*' menggunakan metode '*KMeans*' untuk inisialisasi parameter, '*max_iter*' sebanyak 100 iterasi, '*n_components*' disetel ke 3 untuk menunjukkan jumlah komponen *Gaussian*, '*n_init*' sebanyak 50 untuk jumlah inisialisasi parameter, '*random_state*' diatur ke 2 untuk mereproduksi hasil yang konsisten, '*reg_covar*' yang disetel ke $1e-06$ untuk regularisasi *covariance*, '*tol*' yang disetel ke 0.001 untuk menentukan kriteria konvergensi, dan parameter lainnya yang diatur sesuai dengan kebutuhan.

Hasil pengamatan dari proses ini menunjukkan bahwa data berhasil dikelompokkan menjadi tiga *cluster* berdasarkan fitur-fitur yang diamati. Setiap *cluster* mewakili kelompok data yang memiliki kesamaan internal yang tinggi, sementara memiliki perbedaan yang signifikan antara satu *cluster* dengan *cluster* lainnya. Penggunaan nilai $k=3$ juga sejalan dengan hasil analisis *Elbow* yang menunjukkan penurunan yang signifikan pada nilai $k=3$. Berikut *syntax* dari proses ini dapat dilihat dari Gambar 4.14.

```
from sklearn.mixture import GaussianMixture

gm= GaussianMixture(n_components=3, covariance_type = 'full',
n_init=50, random_state=2)

gm_Predict = gm.fit_Predict(data_scaled)
```

Gambar 4. 14 Syntax tahapan implementasi metode GMM

Sumber: Hasil pengolahan

Pada kode di atas, mengimplementasikan algoritma *Gaussian Mixture Models (GMM)* untuk melakukan klasterisasi pada data yang telah

di-scaling. *GMM* adalah metode probabilistik yang mengasumsikan bahwa data dapat dimodelkan sebagai campuran dari beberapa distribusi *Gaussian* (*normal distribution*).

Pertama, mengimpor kelas *GaussianMixture* dari pustaka *scikit-learn*. Kelas ini menyediakan implementasi dari algoritma *GMM* yang dapat digunakan untuk klasterisasi.

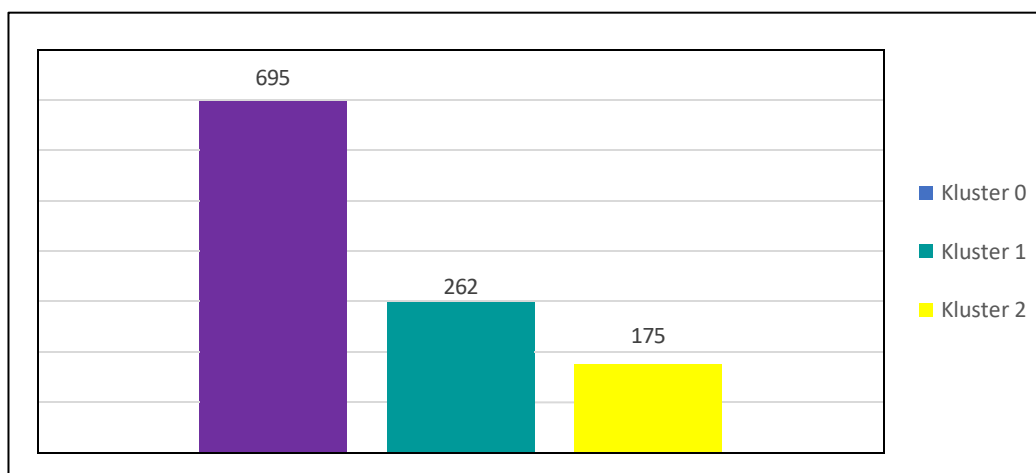
Selanjutnya, membuat sebuah *instance* dari *GaussianMixture* dengan beberapa parameter yang telah diatur:

- a. *n_components=3*: Menentukan bahwa ingin membagi data menjadi 3 *cluster*.
- b. *covariance_type='full'*: Menentukan jenis matriks *covariance* yang akan digunakan. Dalam hal ini, '*full*' berarti setiap *cluster* memiliki matriks *covariance* penuh yang berbeda.
- c. *n_init=50*: Menentukan jumlah inisialisasi yang berbeda untuk algoritma. *GMM* akan menjalankan algoritma sebanyak 50 kali dengan inisialisasi yang berbeda-beda dan memilih hasil terbaik berdasarkan *likelihood*.
- d. *random_state=2*: Menetapkan seed untuk generator angka acak, memastikan hasil yang dapat direproduksi.

Setelah itu, menerapkan model *GMM* pada data yang telah di-scaling dan mendapatkan prediksi *cluster* untuk setiap titik data menggunakan metode *fit_Predict*. Metode ini akan menyesuaikan model dengan data dan mengembalikan label *cluster* untuk setiap titik data.

Dengan langkah-langkah ini, berhasil mengelompokkan data ke dalam tiga *cluster* menggunakan *Gaussian Mixture Models*. *GMM* mempertimbangkan probabilitas dari setiap titik data untuk menjadi bagian dari setiap *cluster*, memberikan fleksibilitas lebih dalam menangani data yang mungkin tidak berbentuk bulat atau elips sempurna, berbeda dengan *K-Means* yang membagi data berdasarkan jarak *Euclidean* ke *centroid*.

Penggunaan $n_init=50$ memastikan bahwa mendapatkan hasil yang lebih stabil dan optimal dengan menjalankan inisialisasi yang berbeda beberapa kali. Setting $random_state=2$ memungkinkan reproduksi hasil yang sama jika kode dijalankan kembali. Berikut hasil sebaran hasil *cluster* dijelaskan pada Gambar 4.15 di bawah ini.



Gambar 4. 15 Jumlah anggota tiap *cluster* hasil metode *Gaussian Mixture Models*

Sumber: Hasil pengolahan

Hasil implementasi *Gaussian Mixture Models (GMM)* pada data menghasilkan tiga *cluster* dengan distribusi sebagai berikut: *cluster 0* terdiri dari 695 data poin, *cluster 1* terdiri dari 262 data poin, dan *cluster 2* terdiri dari 175 data poin. Analisis dari hasil klasterisasi GMM ini memberikan beberapa wawasan penting. Pertama, dari distribusi *cluster* terlihat bahwa *cluster 0* adalah yang terbesar, dengan 695 data poin. Ini menunjukkan bahwa sebagian besar data termasuk dalam *cluster* ini, yang bisa mengindikasikan adanya keseragaman atau pola umum yang mendominasi data tersebut. Sementara itu, *cluster 1* dan *cluster 2* masing-masing memiliki 262 dan 175 data poin, yang lebih kecil dibandingkan dengan *cluster 0*. Hal ini menunjukkan adanya subgrup atau segmentasi yang lebih kecil dalam data dengan karakteristik berbeda dari *cluster 0*. Berikut *syntax* pada Gambar 4.16 merupakan proses visualisasi sebaran *cluster*.

```
plt.figure(figsize=(7, 4))
sns.scatterplot(data=data_vis, x=data_vis['lari'],
y=data_vis['pull_up'], hue='GMM', palette='viridis')
plt.scatter(centroid_KMeans['lari'], centroid_KMeans['pull_up'],
c="red", s=150, alpha=0.6)
plt.show()
```

Gambar 4. 16 Syntax tahapan visualisasi sebaran *cluster* hasil GMM

Sumber: Hasil pengolahan

Pada kode di atas, memvisualisasikan hasil klasterisasi yang diperoleh dari algoritma *Gaussian Mixture Models (GMM)* menggunakan *scatter plot*. Pertama, membuat sebuah *figure* baru dengan ukuran 7x4 inci untuk menampilkan plot scatter menggunakan *plt.figure()* dari pustaka *Matplotlib*. Ukuran ini memastikan bahwa plot yang dihasilkan memiliki resolusi yang cukup untuk melihat perbedaan antara *cluster* dengan jelas.

Selanjutnya, menggunakan *sns.scatterplot()* dari pustaka *Seaborn* untuk membuat *scatter plot*. Titik-titik data diberi warna berdasarkan label *cluster* yang diperoleh dari algoritma GMM, yang tersimpan dalam kolom 'GMM' pada *DataFrame data_vis*. Palet warna 'viridis' digunakan untuk memberikan variasi warna yang jelas antar *cluster*.

Untuk memberikan konteks visual lebih lanjut, juga menambahkan titik-titik merah pada plot untuk menunjukkan posisi *centroid* dari *cluster K-Means*. Posisi *centroid* ini diambil dari *DataFrame centroid_KMeans*, yang berisi koordinat *centroid*. Titik-titik *centroid* ini digambar menggunakan *plt.scatter()* dengan parameter *c="red"* untuk warna merah, *s=150* untuk ukuran titik yang lebih besar, dan *alpha=0.6* untuk membuat titik-titik sedikit transparan sehingga tidak menutupi data asli.

Melalui proses ini, dapat mengamati bagaimana titik-titik data dikelompokkan oleh algoritma GMM. Warna berbeda mewakili *cluster*

berbeda yang dihasilkan oleh GMM, sedangkan titik-titik merah menunjukkan posisi *centroid cluster* yang dihasilkan oleh algoritma *K-Means*. Visualisasi ini memberikan wawasan yang berguna tentang distribusi dan pemisahan *cluster* dalam data, memungkinkan analisis lebih lanjut terhadap hasil klasterisasi. Berdasarkan hasil *cluster*, sebaran *cluster* yang didapatkan dijelaskan pada Gambar 4.17 berikut ini.



Gambar 4. 17 Visualisasi sebaran *cluster Gaussian Mixture Models*

Sumber: Hasil pengolahan

Pada visualisasi *cluster* yang terbentuk digunakan *scatter plot* dengan data yang telah dikelompokkan berdasarkan hasil implementasi *Gaussian Mixture Models*. Setiap titik dalam plot mewakili satu data observasi, yang diwarnai sesuai dengan *cluster* keanggotaannya, ditunjukkan oleh variabel '*GMM*'. Selain itu, *centroid* dari masing-masing *cluster* pada hasil klasterisasi *K-Means* ditandai dengan titik merah yang lebih besar. Pemaparan visual ini memungkinkan untuk memperhatikan pola-pola yang mungkin ada di antara data dan *cluster*, serta memperoleh pemahaman visual tentang bagaimana data dikelompokkan oleh algoritma *Gaussian Mixture Models*.

4.2.5 Evaluasi (*Evaluation*)

Pembahasan Pembahasan hasil evaluasi menggunakan metode *Davies-Bouldin* dan *Calinski-Harabasz Index* adalah bagian penting dari analisis klasterisasi untuk memahami kualitas dan performa dari model-model yang telah diimplementasikan. Metode *Davies-Bouldin* digunakan untuk mengukur kehomogenan *cluster* dan pemisahan antar *cluster*, di mana nilai yang lebih rendah menunjukkan pembentukan *cluster* yang lebih baik. Sementara itu, metode *Calinski-Harabasz* digunakan untuk mengevaluasi kepadatan dan pemisahan *cluster*, di mana nilai yang lebih tinggi menunjukkan bahwa *cluster*-klasternya lebih terpisah secara jelas.

Rumus *Calinski-Harabasz (CH)* Score digunakan untuk mengukur kualitas *clustering* berdasarkan rasio antara jumlah varians total dalam *cluster* dan jumlah varians antar *cluster*. Semakin tinggi nilai CH, semakin baik kualitas *cluster*. Rumus CH Score adalah sebagai berikut:

$$\text{Calinski} - \text{Harabasz} = \frac{\text{tr}(B_k)/(k-1)}{\text{tr}(W_k)/(n-1)}$$

Dimana $\text{tr}(B_k)$ adalah trace dari matriks antara *cluster scatter*, yang mengukur jarak antara *centroid cluster*. $\text{tr}(W_k)$ adalah trace dari matriks *within-cluster scatter*, yang mengukur penyebaran data dalam setiap *cluster*. k adalah jumlah *cluster*. n adalah jumlah total titik data. Berikut langkah-langkah dalam proses *Calinski-Harabasz*:

- a. Proses pertama adalah menghitung nilai *centroid* tiap *cluster* semua variabel dengan cara menghitung nilai mean tiap *cluster* semua variabel dapat juga menggunakan persamaan $\text{Centroid} = (\frac{\sum x_i}{n}, \frac{\sum y_i}{n})$. Dari perhitungan nilai mean diperoleh nilai *centroid* berikut.

Tabel 4. 6 Hasil dari perhitungan

k_m eans	jk	lari	pull_ up	sit_u p	push_ up	stl_r un	ren_j ar	ren_w ak	ipk	prod i
-------------	----	------	-------------	------------	-------------	-------------	-------------	-------------	-----	-----------

0	1.25	-0.86	1.24	-0.13	-1.21	0.90	0.13	-0.07	0.08	-0.30
1	-0.79	0.15	-0.82	-1.61	0.69	-0.46	1.09	0.96	0.18	0.07
2	-0.78	0.69	-0.75	0.73	0.78	-0.60	-0.53	-0.31	-0.14	0.23
gmm	jk	lari	pull_ up	sit_u p	push_ up	stl_r un	ren_j ar	ren_w ak	ipk	prod i
0	-0.79	0.53	-0.78	0.08	0.75	-0.57	-0.08	0.05	-0.05	0.19
1	1.25	-0.69	1.25	-0.02	-1.18	0.83	-0.72	-0.12	0.04	-0.29
2	1.27	-1.08	1.22	-0.28	-1.21	1.00	1.39	0.00	0.13	-0.32

- b. Total *scatter* dalam *cluster* ($tr(W_k)$) diperoleh dengan cara menjumlahkan *scatter* dalam setiap *cluster* atau dengan persamaan $W_k = \sum_{i=1}^k \sum_{x \in c_i} \|x - x_i\|^2$. Dari proses ini diperoleh nilai 5493.88 pada *cluster K-Means* dan 6418.66 untuk *cluster Gaussian Mixture Models*.
- c. Total *scatter* antar *cluster* ($tr(B_k)$) diperoleh dengan cara menjumlahkan *scatter* antar *cluster* dengan mempertimbangkan *centroid* dari setiap *cluster* dapat menggunakan persamaan $B_k = \sum_{i=1}^k n_i \|\mu_i - \mu\|^2$. Dari proses ini diperoleh nilai 5826.11 untuk *cluster K-Means* dan 4901.33 untuk *cluster Gaussian Mixture Models*.
- d. Menggunakan persamaan di atas dapat didistribusikan ke dalam persamaan *Calinski – Harabasz* = $\frac{tr(B_k)/(k-1)}{tr(W_k)/(n-1)}$ yaitu, nilai $tr(B_k) = 5826.11$, $tr(W_k) = 5493.88$, $k = 3$, $n = 1132$ untuk *cluster K-Means* dan nilai $tr(B_k) = 4901.33$, $tr(W_k) = 6418.66$, $k = 3$, $n = 1132$ untuk *cluster Gaussian Mixture Models*. Dari nilai-nilai tersebut, maka diperoleh bentuk $(K - means) \text{ Calinski – Harabasz} = \frac{5826.11/(3-1)}{5493.88/(1132-1)}$ dan $(Gmm) \text{ Calinski – Harabasz} = \frac{4901.33/(3-1)}{6418.66/(1132-1)}$
- e. Dari perhitungan di atas, diperoleh nilai *Calinski-Harabasz* untuk metode *K-Means* sebesar 598.63 dan untuk metode *Gaussian Mixture Models* sebesar 431.05.

Rumus *Davies-Bouldin Index (DBI)* digunakan untuk mengukur kualitas *clustering* dengan mengevaluasi seberapa baik *cluster* terpisah dan

seberapa kompak mereka. Semakin rendah nilai DBI, semakin baik kualitas *cluster*. Rumus DBI adalah sebagai berikut:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \left(\frac{s_i + s_j}{d_{ij}} \right)$$

Dimana, k adalah jumlah *cluster*, s_i adalah rata-rata jarak antara setiap titik dalam *cluster* i dan *centroid cluster* i , d_{ij} adalah jarak antara *centroid cluster* i dan *centroid cluster* j . Tahapan proses perhitungan *Davies-Bouldin Index (DBI)* dijelaskan pada poin-poin berikut :

- a. Hitung *centroid* masing-masing variabel menggunakan persamaan $Centroid = \frac{1}{n} \sum_{i=1}^n x_i$. Dari persamaan ini diperoleh titik *centroid* tiap variabel sebagai berikut.

Tabel 4. 7 Hasil dari perhitungan

model	jk	lari	pull_up	sit_up	push_up	stl_run	ren_jar	ren_wak	ipk	prod_i
gmm	1.27	-1.08	1.22	-0.28	-1.21	1.00	1.39	0.00	0.13	-0.32
<i>K-Means</i>	-0.78	0.69	-0.75	0.73	0.78	-0.60	-0.53	-0.31	-0.14	0.23

- b. Hitung rata-rata jarak antara setiap titik dalam *cluster* ke *centroid*-nya (s_i) menggunakan persamaan $S_i = \frac{1}{n_i} \sum_{x \in C_i} \|x - x_i\|$. Dari persamaan ini diperoleh nilai jarak 2.30, 2.11, dan 1.94 pada metode *K-Means*. Pada metode *Gaussian Mixture Models* diperoleh nilai jarak 2.44, 1.96, dan 2.17.
- c. Proses selanjutnya adalah hitung rasio (R_{ij}) untuk masing-masing pasangan *cluster* menggunakan persamaan $R = \frac{s_i + s_j}{d_{ij}}$. Dari persamaan ini diperoleh nilai 1.00, 0.99, 1.00, 1.27, 0.99, dan 1.27 untuk metode *K-Means* dan diperoleh nilai 1.09, 1.04, 1.09, 1.90, 1.04, dan 1.90 untuk metode *Gaussian Mixture Models*.
- d. Proses selanjutnya menghitung nilai *Davies-Bouldin Index* menggunakan persamaan di atas. Dari proses ini diperoleh nilai

Davies-Bouldin Index 1.185 pada metode *K-Means* dan 1.633 untuk metode *Gaussian Mixture Models*.

Dalam pembahasan ini, akan dianalisis hasil evaluasi dari kedua metode tersebut untuk model-model klasterisasi yang telah diimplementasikan, yaitu *K-Means* dan *Gaussian Mixture Models*. Evaluasi ini akan memberikan wawasan yang mendalam tentang kualitas klasterisasi yang telah dilakukan, serta membantu dalam memahami kelebihan dan kekurangan dari masing-masing model. Dengan demikian, pembahasan hasil evaluasi ini menjadi langkah penting dalam menentukan model yang paling sesuai untuk kebutuhan analisis data yang spesifik. Berikut proses *syntax* yang tertera pada Gambar 4.18.

```
from sklearn import metrics

KMeans_silhouette = metrics.silhouette_score(data_scaled,
KMeans_Predict, metric='Euclidean')

Gaussian_silhouette = metrics.silhouette_score(data_scaled,
gm_Predict, metric='Euclidean')

KMeans_db = metrics.davies_bouldin_score(data_scaled,
KMeans_Predict )

Gaussian_db = metrics.davies_bouldin_score(data_scaled,
gm_Predict)

KMeans_ch = metrics.calinski_harabasz_score(data_scaled,
KMeans_Predict )

Gaussian_ch = metrics.calinski_harabasz_score(data_scaled,
gm_Predict)

data_hasil = [
    ['KMeans', KMeans_silhouette, KMeans_ch, KMeans_db],
    ['Gaussian Mixture Model', Gaussian_silhouette, Gaussian_ch,
Gaussian_db]
```

```

    ]

result = pd.DataFrame(data_hasil, columns=['Model',
'Silhouette_Score', 'Calinski_Harabasz_Score',
'Davies_Bouldin_Index'])

result

```

Gambar 4. 18 Syntax tahapan pengukuran metode *K-Means* dan *Gaussian Mixture Models* menggunakan *Silhouette score*, *Calinski-Harabasz Score*, *Davies-Bouldin Index*

Sumber: Hasil pengolahan

Pada potongan kode di atas, melakukan evaluasi performa dua model klasterisasi, yaitu *K-Means* dan *Gaussian Mixture Models*, menggunakan beberapa metrik evaluasi. Pertama, menghitung nilai *silhouette score* untuk masing-masing model. *Silhouette score* mengukur seberapa baik objek berada dalam *cluster* mereka sendiri dibandingkan dengan *cluster* lain, dengan nilai yang lebih tinggi menunjukkan klasterisasi yang lebih baik. Kedua, menghitung nilai *Calinski-Harabasz Score*, yang merupakan metrik statistik untuk mengukur kepadatan dan sebaran *cluster*. Nilai yang lebih tinggi menunjukkan *cluster* yang lebih padat dan terpisah dengan baik. Terakhir, menghitung nilai *Davies-Bouldin Index*, yang mengevaluasi kemiripan antara *cluster* berdasarkan rata-rata jarak antara *cluster*. Nilai yang lebih rendah menunjukkan klasterisasi yang lebih baik.

Analisis metrik-metrik ini memberikan pemahaman yang lebih dalam tentang performa masing-masing model dalam melakukan klasterisasi terhadap data yang telah di-scaling sebelumnya.

Tabel 4. 8 Hasil evaluasi kinerja metode *K-Means* dan *Gaussian Mixture Models*

Model	<i>Silhouette_Score</i>	<i>Calinski_Harabasz_Score</i>	<i>Davies_Bouldin_Index</i>
<i>KMeans</i>	0.349682	598.6378	1.185072
<i>Gaussian Mixture Models</i>	0.283797	431.0565	1.633811

Sumber: Hasil pengolahan

Berdasarkan hasil pengukuran metrik evaluasi *clustering* yang tertera pada Tabel 4.8, diperoleh nilai-nilai sebagai berikut untuk model *K-Means* dan *Gaussian Mixture Models (GMM)*. Model *K-Means* menghasilkan *Silhouette score* sebesar 0.349682, *Calinski-Harabasz Score* sebesar 598.63779, dan *Davies-Bouldin Index* sebesar 1.185072. Sementara itu, *Gaussian Mixture Models* menghasilkan *Silhouette score* sebesar 0.283797, *Calinski-Harabasz Score* sebesar 431.056536, dan *Davies-Bouldin Index* sebesar 1.633811.

Dari hasil pengamatan, dapat dilihat bahwa model *K-Means* memiliki performa yang lebih baik dibandingkan dengan GMM dalam konteks *clustering* ini. *Silhouette score* yang lebih tinggi pada *K-Means* menunjukkan bahwa *cluster* yang dihasilkan oleh *K-Means* lebih baik dalam hal separasi antar *cluster* dan kepadatan dalam *cluster*. Nilai ini mendekati 1, menunjukkan bahwa data poin umumnya lebih dekat dengan *cluster* mereka sendiri dibandingkan dengan *cluster* lainnya.

Calinski-Harabasz Score, yang juga lebih tinggi untuk *K-Means*, mengindikasikan bahwa *cluster* yang dihasilkan oleh *K-Means* memiliki distribusi yang lebih baik, dengan varian dalam *cluster* yang lebih kecil dan varian antar *cluster* yang lebih besar. Skor ini mengukur rasio antara *sum of between-cluster dispersion* dan *sum of within-cluster dispersion*, sehingga nilai yang lebih tinggi menunjukkan *cluster* yang lebih kompak dan terpisah.

Sebaliknya, *Davies-Bouldin Index* yang lebih rendah pada *K-Means* menunjukkan bahwa *cluster* yang dihasilkan lebih baik karena nilai indeks yang lebih rendah menandakan *cluster* yang lebih jauh terpisah dan lebih kompak. *Davies-Bouldin Index* merupakan rata-rata dari rasio kesamaan antar *cluster* terhadap ukuran *cluster*, sehingga nilai yang lebih rendah mengindikasikan *cluster* yang lebih baik.

Analisis ini menunjukkan bahwa *K-Means* lebih efektif dalam mempartisi data ke dalam *cluster* yang berbeda dibandingkan dengan GMM dalam skenario ini. Meskipun GMM mampu menangani distribusi data yang lebih kompleks dan heterogen, dalam hal ini, struktur data mungkin lebih sesuai dengan asumsi yang dipegang oleh *K-Means* mengenai bentuk dan ukuran *cluster*. Oleh karena itu, *K-Means* menunjukkan performa yang superior berdasarkan metrik evaluasi yang digunakan, memberikan *cluster* yang lebih jelas terdefinisi dan terpisah dengan baik.

Setelah itu akan diimplementasikan evaluasi *elbow*. *Elbow* merupakan proses penting dilakukan untuk menentukan nilai *k* yang optimal dalam algoritma klasterisasi. Rentang nilai *k* yang digunakan adalah dari 1 hingga 9. Metode *Elbow* menjadi salah satu pendekatan yang efektif dalam menentukan jumlah *cluster* yang terbaik dalam dataset. Dengan menggunakan teknik ini, dilakukan analisis terhadap variasi nilai *k* dan varians intra-*cluster* yang terkait. Pada akhirnya, titik “*Elbow*” dalam kurva akan diidentifikasi, yang menandakan titik di mana penambahan *cluster* baru tidak memberikan penurunan varians intra-*cluster* yang signifikan lagi. Proses ini penting untuk memilih nilai *k* yang optimal sehingga hasil klasterisasi mencerminkan struktur data secara akurat dan efisien. Berikut *syntax* pada proses ini.

```
# Instantiate the clustering model and visualizer
model = KMeans()
# model = GaussianMixture()
```



```

visualizer = KElbowVisualizer(model, k=(1, 10))
visualizer.fit(data_scaled) # Fit the data to the visualizer
visualizer.show() # Finalize and render the figure
plt.show()

```

Gambar 4. 19 Syntax tahapan pengukuran *inertia* atau *Elbow*

Sumber : Hasil pengolahan

Pada Gambar 4.19 di atas, proses kode yang dilakukan bertujuan untuk menentukan jumlah *cluster* optimal untuk algoritma *K-Means* menggunakan metode *Elbow*. Metode *Elbow* adalah teknik yang digunakan untuk menemukan jumlah *cluster* yang tepat dengan memplot nilai *inertia* (*total within-kl sum of square*) terhadap jumlah *cluster* dan mencari “*Elbow point*” di mana penurunan *inertia* mulai melambat.

Pertama, objek model diinisialisasi sebagai model *K-Means* dari pustaka *sklearn.cluster*. Komentar di sebelahnya menunjukkan bahwa *Gaussian Mixture Models* juga dapat digunakan, tetapi dalam hal ini, fokus pada *K-Means*.

Selanjutnya, objek *visualizer* diinisialisasi sebagai *KElbowVisualizer* dari pustaka *yellowbrick.cluster*. *KElbowVisualizer* adalah alat visualisasi yang dirancang untuk membantu memilih jumlah *cluster* optimal dengan memplot skor *Elbow* untuk berbagai nilai *k*. Parameter *k*=(1, 10) menunjukkan bahwa jumlah *cluster* yang akan diuji adalah dari 1 hingga 10.

Kemudian, metode *fit* dari *visualizer* diterapkan pada data yang sudah distandarisasi (*data_scaled*). Proses ini melibatkan pelatihan model *K-Means* pada data untuk setiap nilai *k* dalam rentang yang ditentukan (1 hingga 10) dan menghitung nilai *inertia* untuk setiap *cluster*.

Setelah itu, metode *show* dipanggil untuk menampilkan plot *Elbow* yang dihasilkan. Plot ini menunjukkan nilai *inertia* untuk setiap jumlah

cluster dan membantu dalam mengidentifikasi titik di mana penurunan *inertia* mulai melambat secara signifikan (*Elbow point*).

Dari hasil perhitungan nilai *inertia* menggunakan metode *Elbow*, terlihat adanya perubahan signifikan pada nilai *inertia* ketika jumlah *cluster* ($n_clusters$) bertambah dari 1 hingga 9. Nilai *inertia* menunjukkan seberapa baik data telah dikelompokkan oleh model *clustering*, dengan nilai yang lebih rendah mengindikasikan bahwa data dalam *cluster* lebih kompak.

Pada $n_clusters = 1$, nilai *inertia* adalah 11320.0, yang sangat tinggi karena semua data berada dalam satu *cluster* besar. Ketika jumlah *cluster* ditingkatkan menjadi 2, terjadi penurunan drastis pada nilai *inertia* menjadi 6910.888. Penurunan ini menunjukkan bahwa pembagian data menjadi dua *cluster* menghasilkan pengelompokan yang lebih baik dibandingkan satu *cluster*. Penurunan *inertia* berlanjut saat jumlah *cluster* ditambah, dengan nilai *inertia* menjadi 5493.880 untuk 3 *cluster*, 5000.192 untuk 4 *cluster*, dan seterusnya.

Namun, penurunan terbesar terjadi antara $n_clusters = 2$ dan $n_clusters = 3$, dimana *inertia* berkurang dari 6910.888 menjadi 5493.880. Setelah itu, meskipun *inertia* terus berkurang, laju penurunannya menjadi lebih kecil dan stabil, seperti terlihat dari perubahan nilai *inertia* dari 5493.880 pada 3 *cluster* menjadi 5000.192 pada 4 *cluster*, dan seterusnya.

Berdasarkan pola perubahan nilai *inertia* ini, jumlah *cluster* yang optimal, sesuai dengan metode *Elbow*, adalah pada $n_clusters = 3$. Pada titik ini, tambahan *cluster* tidak lagi menghasilkan pengurangan signifikan pada nilai *inertia*, menunjukkan bahwa tiga *cluster* memberikan keseimbangan terbaik antara kompleksitas model dan kualitas pengelompokan. Dengan demikian, memilih $k = 3$ untuk model *clustering* dianggap sebagai keputusan terbaik karena dapat mengelompokkan data secara efektif tanpa menambah kompleksitas yang tidak perlu.

Secara keseluruhan, dari hasil pengukuran ini dapat disimpulkan bahwa *K-Means* merupakan metode yang lebih tepat untuk *clustering* data ini dibandingkan dengan GMM. Evaluasi dengan menggunakan ketiga metrik (*Silhouette score*, *Calinski-Harabasz Score*, dan *Davies-Bouldin Index*) konsisten menunjukkan bahwa *K-Means* menghasilkan *cluster* yang lebih kompak dan terpisah dengan baik, ditambah dengan pembuktian *elbow* untuk menunjukkan bahwa nilai *k* yang digunakan oleh *k-means* adalah nilai *k* yang optimal.

4.2.6 Penyebaran (*Deployment*)

Model klastering yang telah dievaluasi siap untuk tahap *deployment*, menjadi langkah akhir dalam metode penelitian CRISP-DM. Tahap ini menandakan kesiapan model untuk diimplementasikan dalam tugas pengelompokan kesiapan dan distribusi personil komponen cadangan di Unhan RI. Hasil evaluasi model yang disajikan dalam diagram menjadi dasar untuk pertimbangan peningkatan kualitas aset pertahanan, mendukung keamanan nasional, dan membantu pengambilan keputusan. *Deployment* memastikan model yang dikembangkan dapat digunakan secara fungsional untuk memperkuat pengembangan dan pengelolaan sumber daya manusia Universitas Pertahanan Republik Indonesia, dengan tujuan akhir memperkuat pertahanan nasional.

Memberikan gambaran komprehensif mengenai halaman web "Kesiapan SEPA PK TNI" yang dirancang untuk Universitas Pertahanan RI. Halaman ini menyajikan data kesiapan mahasiswa SEPA PK TNI melalui visualisasi yang informatif dengan struktur utama halaman web sebagai berikut:

a. Home

Gambar 4. 20 menunjukkan halaman utama situs dengan judul utama "Visualisasi Data Kesiapan SEPA PK TNI" dan subjudul yang menunjukkan afiliasi dengan Universitas

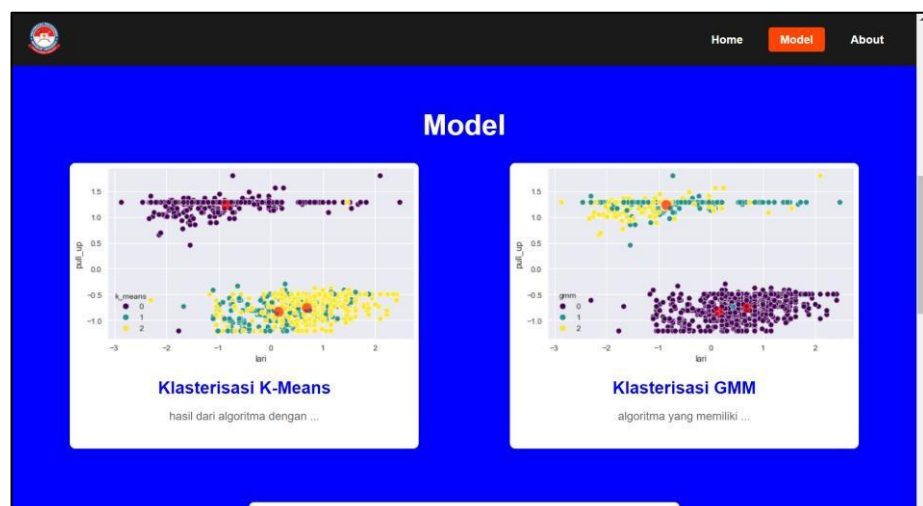
Pertahanan RI, bagian visual ini menampilkan latar belakang penuh dari bangunan Universitas Pertahanan RI.



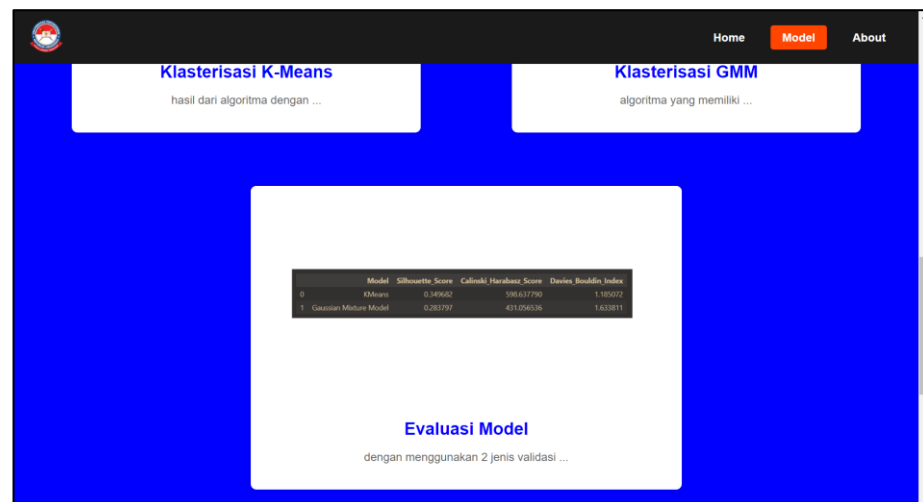
Gambar 4. 20 Halaman Utama Situs

b. Model

Bagian ini menampilkan tiga kartu utama dalam tata letak grid, masing-masing merepresentasikan aspek berbeda seperti yang bisa dilihat dari Gambar 4. 21 dan Gambar 4.22 dari studi—Klasterisasi *K-Means*, Klasterisasi *GMM*, dan Evaluasi Model.



Gambar 4. 21 Halaman Situs yang Berisikan Hasil *Clustering*

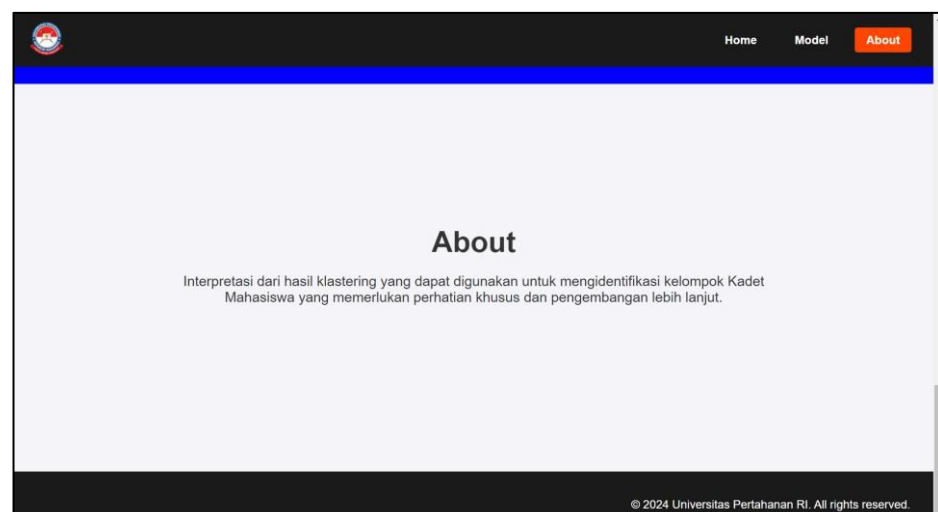


Gambar 4. 22 Halaman Situs yang Berisikan Hasil Evaluasi *Clustering*

Setiap kartu dilengkapi dengan gambar, judul, dan deskripsi singkat yang memberikan gambaran hasil yang diperoleh dari algoritma klasterisasi dan evaluasinya.

c. About

Ini memberikan penjelasan singkat tentang tujuan penelitian dan implikasinya, serta tujuan web ini yang bisa dilihat dari Gambar 4.23.



Gambar 4. 23 Halaman Penjelasan Mengenai Isi Situs

Bagian ini juga menekankan betapa pentingnya interpretasi hasil klasterisasi untuk mengidentifikasi kelompok siswa kadet yang membutuhkan pengembangan dan perhatian khusus. Melalui analisis yang mendalam dan strategis, penelitian ini bertujuan untuk meningkatkan kualitas pengelolaan sumber daya manusia di Universitas Pertahanan RI.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan penelitian dan analisis yang telah dilakukan diperolehnya hasil yang menjawab permasalahan sebagai landasan penelitian ini, dapat disimpulkan sebagai berikut:

1. Algoritma *K-Means* dan *Gaussian Mixture Models* telah berhasil mengklasifikasikan kesiapan kadet mahasiswa S1 Universitas Pertahanan RI untuk SEPA PK TNI berdasarkan karakteristik dan kompetensi yang diukur. *K-Means* menunjukkan hasil klasterisasi dengan pemisahan klaster yang lebih jelas dibandingkan *Gaussian Mixture Models*, berdasarkan nilai *Calinski-Harabasz Score* yang lebih tinggi. Namun, *Gaussian Mixture Models* menghasilkan klaster yang lebih homogen, ditunjukkan oleh nilai *Davies-Bouldin Index* yang lebih rendah.
2. Dalam hal akurasi dan relevansi dengan kriteria kesiapan prajurit karier TNI, algoritma *K-Means* menunjukkan performa klasterisasi yang lebih baik dibandingkan *Gaussian Mixture Models*. Hal ini ditunjukkan oleh nilai *Calinski-Harabasz Score* yang lebih tinggi pada *K-Means* (598.6378) dibandingkan *Gaussian Mixture Models* (431.0565), yang menandakan bahwa *K-Means* memiliki pemisahan klaster yang lebih baik. Dalam pengukuran nilai *Davies-Bouldin Indeks*, algoritma *K-Means* juga memiliki performa lebih baik dengan nilai lebih rendah yaitu (1.185072) dibandingkan *Gaussian Mixture Models* (1.633811), yang menunjukkan klaster yang lebih kompak.
3. Analisis klasterisasi menggunakan *K-Means* dan *Gaussian Mixture Models* memberikan *insight* yang berharga untuk peningkatan program pembinaan di Unhan RI. Hasil klasterisasi dapat digunakan untuk mengidentifikasi kelompok kadet yang

memerlukan perhatian khusus dan pengembangan lebih lanjut. Misalnya, klaster dengan performa fisik dan mental yang lebih rendah dapat diberikan program pembinaan tambahan untuk meningkatkan kesiapan mereka dalam menghadapi SEPA PK TNI. Dengan demikian, analisis ini dapat berkontribusi dalam mendukung program-program yang dapat meningkatkan kualitas sumber daya manusia dan dapat membantu mengambil keputusan strategis dengan cara dilaksanakan *deployment* di Universitas Pertahanan RI.

5.2 Saran

Terdapat beberapa perbaikan yang dapat dilakukan di masa depan untuk mendorong pengembangan penelitian yang relevan. Untuk pengembangan penelitian selanjutnya, beberapa rekomendasi yang diberikan oleh penulis adalah:

1. Penelitian selanjutnya dapat mencakup lebih banyak variabel yang relevan untuk menggambarkan kesiapan kadet mahasiswa dengan lebih komprehensif, termasuk faktor-faktor psikologis dan sosial.
2. Selain *K-Means* dan *Gaussian Mixture Models*, dapat dieksplorasi algoritma klasterisasi lain seperti *DBSCAN* atau *hierarchical clustering* untuk melihat perbandingan hasil yang mungkin memberikan *insight* tambahan.
3. Validasi hasil klasterisasi dengan data eksternal atau data dari institusi militer lain dapat memberikan gambaran yang lebih jelas mengenai generalisasi dan keakuratan model yang digunakan.
4. Melakukan studi longitudinal untuk memantau perkembangan kesiapan kadet dari tahun ke tahun akan memberikan pemahaman yang lebih baik tentang efektivitas program pembinaan dan intervensi yang dilakukan.

5. Berdasarkan hasil klasterisasi, implementasi program pembinaan yang khusus disesuaikan untuk setiap klaster dapat dievaluasi untuk mengukur dampaknya terhadap peningkatan kesiapan kadet secara keseluruhan.

DAFTAR PUSTAKA

- Adhinata, D. F., Rakhmadani, D. P., & Segara, J. T. A. (2021). Pengenalan Jenis Kelamin Manusia Berbasis Suara Menggunakan MFCC dan GMM. *Data Institut Teknologi Telkom Purwokerto*, 1(1). <https://research.google.com/audioset>
- Amri, P. (2023). Urgensi Peraturan Presiden Tentang Pelibatan Militer Dalam Pengangguhan Terorisme di Indonesia. *Universitas Atma Jaya Yogyakarta JL. Mrican Baru*, 52(2), 55281.
- Anugrah, F. N. (2023). Analisis Relevansi Penunjukan Anggota TNI/POLRI Sebagai Penjabat (PJ) Kepala Daerah. *Jurnal Kebijakan Pembangunan*, 18(1), 121–134. <https://doi.org/10.47441/jkp.v18i1.294>
- Astarini, D. R. S., & Rofii, M. S. (2021). Siber Intelijen Untuk Keamanan Nasional. *Jurnal Renaissance*, 6(1), 703. <https://doi.org/10.53878/jr.v6i1.143>
- Azizah, R. Z. H. (2021). Mendefinisikan Kembali Konsep Keamanan dalam Agenda Kebijakan Negara-Bangsa. *Jurnal Diplomasi Pertahanan*, 6(3), 94–104. <https://doi.org/10.33172/jdp.v6i3.666>
- Doi, M. D., Rusgiyono, A., & Wuryandari, T. (2023). ANALISIS k-MEDOIDS DENGAN VALIDASI INDEKS PADA IPM DAERAH 3T DI INDONESIA. *Jurnal Gaussian*, 12(2), 178–188. <https://doi.org/10.14710/j.gauss.12.2.178-188>
- Herlinda, V., & Darwis, D. (2021). Analisis Clustering Untuk Recredesialing Fasilitas Kesehatan Menggunakan Metode Fuzzy C-Means. *Darwis, Dartono*, 2(2), 94–99.
- Hidayat, F. (2022). Keikutsertaan Membangun Sistem Pertahanan Dan Keamanan Negara Diera Globalisasi. *Researchgate.Net, March*.
- Hutagalung, J. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *JATISI (Jurnal Teknik Informatika Dan*

Sistem Informasi), 9(1), 606–620.
<https://doi.org/10.35957/jatisi.v9i1.1516>

Kamil, M. A. (2022). *Legalitas Anggota TNI Aktif Dalam Rangkap Jabatan Sipil (Studi Terhadap Undang-Undang Nomor 34 Tahun 2004 Tentang Tentara Nasional Indonesia)*. 42.

Kamila, C., Adiyatma, M., Namang, G. R., & Syah, R. R. F. (2021). Systematic Literature Review: Penggunaan Algoritma K-Means Untuk Clustering di Indonesia dalam Bidang Pendidikan. *JURNAL INTECH*, 2(1), 1.

Kementerian Pendidikan, K. R. dan T. (2022). *Peraturan Menteri Pendidikan, Kebudayaan, Riset, dan Teknologi Nomor 53 Tahun 2022 tentang Organisasi dan Tata Kerja Universitas Pertahanan*.

Kementerian Pertahanan Republik Indonesia. (2014). Buku Putih Pertahanan Indonesia. In <https://www.kemhan.go.id/> (Issue 8).

Lubis, A. F. (2022). HAM dan Profesionalisme Tentara Nasional Indonesia. *Ideas: Jurnal Pendidikan, Sosial, Dan Budaya*, 8(1), 253.
<https://doi.org/10.32884/ideas.v8i1.544>

Luthfi, E., & Wijayanto, A. W. (2021). Analisis perbandingan metode hirearchical, k-means, dan k-medoids clustering dalam pengelompokkan indeks pembangunan manusia Indonesia. *Inovasi*, 17(4), 761–773. <https://doi.org/10.30872/jinv.v17i4.10106>

Manurung, Y., Saragih, H., & Sarjito, A. (2022). Martabat Bangsa Dan Negara Di Atas Segala-Galanya: Tinjauan Aksiologis Filsafat Ilmu Pertahanan Dalam Upaya Meningkatkan Kemampuan Bela Negara. *Jurnal Civic Hukum*, 7(1), 21–30.

Mardhani, D. (2020). Security And Defence Dalam Studi Ketahanan Nasional Guna Mewujudkan Sistem Keamanan Nasional. *Jurnal*

- Pertahanan & Bela Negara*, 10(3), 279.
<https://doi.org/10.33172/jpbh.v10i3.862>
- Muttaqin, M. R., Hermanto, I. T., & Sunandar, A. M. (2022). *Penerapan K-Means Clustering dan Cross-Industry Standard Process for Data Mining (CRISP-DM) Untuk Mengelompokan Penjualan Kue*. 19(1), 38–53. <https://journal.unpak.ac.id/index.php/komputasi>
- Oyewole, G. J., & Thopil, G. A. (2023). Data clustering: application and trends. In *Artificial Intelligence Review* (Vol. 56, Issue 7). Springer Netherlands. <https://doi.org/10.1007/s10462-022-10325-y>
- Patel, E., & Kushwaha, D. S. (2020). Clustering Cloud Workloads: K-Means vs Gaussian Mixture Model. *Procedia Computer Science*, 171(2019), 158–167. <https://doi.org/10.1016/j.procs.2020.04.017>
- Prasetyo, A. S. (2020). Dampak Makroekonomi Terhadap Nonperforming Loans: Studi Kasus Bank Konvensional di Asean. *Competence: Journal of Management Studies*, 14(2).
- Putra, I., & Lubis, A. F. (2022). Operasi Militer Selain Perang (OMSP) TNI Untuk Mengatasi Aksi Terorisme Dalam Perspektif Ham Dan Penegakan Hukum. *Jurnal Ilmu Sosial Dan Pendidikan (JISIP)*, 6(2).
- Rahayu, N. L. W. S., Arifah, F. N., Wijaya, A. H., Septarini, R. S., Ahmad, N., & Putu Yudhi Ardiana, D. (2021). *Data Mining dan Penerapan Algoritma*. Yayasan Kita Menulis.
- Rahmati, R., & Wijayanto, A. W. (2021). Analisis Cluster Dengan Algoritma K-Means, Fuzzy C-Means Dan Hierarchical Clustering. *JIKO (Jurnal Informatika Dan Komputer)*, 5(2), 73–80.
- Republik Indonesia. (2002). *Undang-Undang Republik Indonesia Nomor 3 Tahun 2002 Tentang Pertahanan Negara*.

- Republik Indonesia. (2004). *Undang-Undang Republik Indonesia Nomor 34 Tahun 2004 Tentang Tentara Nasional Indonesia*. www.hukumonline.com
- Rohmy, A. M., Suratman, T., & Nihayaty, A. I. (2020). Peranan Tentara Nasional Indonesia dalam Penindakan Terorisme Berbasis Agama. *AT-TURAS: Jurnal Studi Keislaman*, 7(1), 86–112. <https://doi.org/10.33650/at-turas.v7i1.1012>
- Sikana, A. M., & Wijayanto, A. W. (2021). Analisis Perbandingan Pengelompokan Indeks Pembangunan Manusia Indonesia Tahun 2019 dengan Metode Partitioning dan Hierarchical Clustering. *Jurnal Ilmu Komputer*, 14(2), 66. <https://doi.org/10.24843/jik.2021.v14.i02.p01>
- Sulistyo, H. D. (2020). Tinjauan Yuridis Pertahanan Negara Yang Dilakukan Tentara Nasional Indonesia. *YUSTISIA MERDEKA: Jurnal Ilmiah Hukum*, 6(2). <http://yustisia.unmermadiun.ac.id/index.php/yustisia>
- Syahputri, A. Z., Fallenia, F. Della, & Syafitri, R. (2023). Kerangka berfikir penelitian kuantitatif. *Tarbiyah: Jurnal Ilmu Pendidikan Dan Pengajaran*, 2(1), 160–166.
- Trihartono, A., Indriastuti, S., & Nisya, C. (2020). Keamanan dan Sekuritisasi dalam Hubungan Internasional. In *Melvana Publishing* (Issue 8).
- Uksan, A. (2022). Manajemen Pendidikan Karakter TNI dalam Meningkatkan Mental Kejuangan Prajurit Kodam/XIV Hasanuddin. *Jurnal Kewarganegaraan*, 6(2).
- Umagapi, I. T., Umaternate, B., Hazriani, H., & ... (2023). Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa. *Prosiding ...*, 303–308.
- Universitas Pertahanan RI. (2021). *Peraturan Rektor Universitas Pertahanan Republik Indonesia Nomor 59 Tahun 2021*

Tentang Peraturan Khusus Kadet Mahasiswa Universitas Pertahanan Republik Indonesia.

Universitas Pertahanan RI. (2023a). *Peraturan Rektor Universitas Pertahanan Republik Indonesia Nomor 68 Tahun 2023 Tentang Pedoman Kehidupan Kadet Mahasiswa Program Diploma Tiga dan Program Sarjana/Strata Satu Universitas Pertahanan Republik Indonesia.*

Universitas Pertahanan RI. (2023b). *Sprin/3328/XI/2023 Tentang Pengasuh Resimen Kadet.*

Wahidah, Z., & Utari, D. T. (2023). Comparison of K-Means and Gaussian Mixture Model in Profiling Areas By Poverty Indicators. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 17(2), 0717–0726. <https://doi.org/10.30598/barekengvol17iss2pp0717-0726>

Yuli W, Y., Suherman, Satino, Yayi, D. D., & Hermina. (2023). Patriotisme Bela Negara dalam Memperkuat Karakter Rasa Nasionalisme. *National Conference on Law Studies (NCOLS)*, 5(1).

LAMPIRAN

Lampiran 1 Surat Permohonan & Disposisi

**UNIVERSITAS PERTAHANAN RI
FAKULTAS SAINS DAN TEKNOLOGI PERTAHANAN**

NOTA DINAS

NOMOR : B/ND/ /II/2024/FSTP

Kepada : Yth. Dekan FKIK Unhan RI
Dari : Dekan FSTP Unhan RI
Perihal : Permohonan Izin Meminta Seluruh Data Nilai Transkrip IPK Terakhir Seluruh Cohort untuk Keperluan Skripsi Kadet Mahasiswa Informatika FSTP Unhan RI

1. Dasar:

- a. Keputusan Rektor Universitas Pertahanan RI Nomor: KEP/212.a/VIII/2021 tanggal 29 Agustus 2022 tentang Perubahan 1 Kalender Akademik Program Sarjana (S1) Cohort-1 Universitas Pertahanan RI TA. 2020 s.d 2024;
- b. Keputusan Rektor Universitas Pertahanan RI Nomor: KEP/260/XII/2023 tanggal 11 Desember 2023 tentang Program Kerja dan Anggaran Universitas Pertahanan RI TA. 2024;
- c. Surat Perintah Rektor Universitas Pertahanan RI Nomor: SPRIN/173.a/I/2024 tanggal 16 Januari 2024 tentang Perubahan 1 Pembimbing Skripsi Kadet Mahasiswa Cohort-1 Program Studi Informatika Fakultas Sains dan Teknologi Pertahanan Universitas Pertahanan RI TA.2023/2024.

2. Sehubungan dengan dasar di atas, dengan hormat disampaikan Permohonan Izin Meminta Seluruh Data Nilai Transkrip IPK Seluruh Cohort untuk Keperluan Skripsi Kadet Mahasiswa s.p. Made Hayu Aretha Adira Pangastuti NIM. 320200401012, dengan Judul Skripsi Komparatif Klasterisasi Algoritma K-Means dan Gaussian Mixture Models untuk Kesiapan Komponen Cadangan di Universitas Pertahanan Republik Indonesia Guna Mendukung Keamanan Nasional. Dokumen Pendukung Terlampir.

3. Demikian untuk menjadikan periksa.

Bogor, Januari 2024

Dekan
Fakultas Sains dan Teknologi Pertahanan,


Prof. Dr. Ir. Muhammad Asvial, M.Eng
Pembina Tk. I IV/b

Tembusan:

1. Rektor Unhan RI
2. Warek I, Warek III, dan Warek IV Unhan RI
3. Kasatras Unhan RI
4. Ka. LPMPP, Ka. LPPM, dan Ka. LPPN Unhan RI
5. Kabag AK Unhan RI
6. Danmenkad Unhan RI.

**UNIVERSITAS PERTAHANAN RI
FAKULTAS FARMASI MILITER**

NOTA DINAS
NOMOR : B/ND/138 /II/2024/FFM

Kepada : Yth. Dekan Fakultas Sains dan Teknologi Pertahanan
Unhan RI

Dari : Plt. Dekan Fakultas Farmasi Militer Unhan RI

Hal : Pengiriman Data Transkrip Nilai IPK Prodi Farmasi
Fakultas Farmasi Militer Unhan RI.

1. Dasar :
 - a. Keputusan Rektor Unhan RI Nomor: KEP/260/XII/2023 tanggal 11 Desember 2023 tentang Program Kerja dan Anggaran Universitas Pertahanan RI TA. 2024;
 - b. Nota Dinas Dekan fakultas Sains dan teknologi Pertahanan Nomor B/Nd/184/II/2024/FSTP tanggal 1 Februari 2024 tentang Permohonan Izin meminta data Transkrip Nilai IPK Terakhir Seluruh *Cohort* untuk Keperluan Skripsi Kadet Mahasiswa Informatika FSTP Unhan RI.
2. Sehubungan dengan dasar di atas mohon dengan hormat dikirimkan Data Transkrip Nilai IPK Prodi Farmasi Fakultas Farmasi Militer Unhan RI Semester Ganjil TA 2023/2024 untuk keperluan Skripsi Kadet Mahasiswa a.n. Made Hayu Aretha Adira Pangastuti NIM 320200401012, dengan judul Skripsi Komparatif Klasterisasi Alogaritma *K-Means* dan *Gaussian Mixture Models* untuk kesiapan Komponen Cadangan di Universitas Pertahanan Republik Indonesia Guna Mendukung Keamanan Nasional.
3. Demikian mohon menjadikan periksa .

Bogor, 5 Februari 2024

Plt Dekan
Fakultas Farmasi Militer,



Prof. Dr. apt. Yahdiana Harahap, MS
Pembina Utama IV/e

Tembusan:

1. Rektor Unhan RI
2. Warek II Unhan RI
3. Kasatwas Unhan RI
4. Karo AK Unhan RI
5. Danmenkad Unhan RI



UNIVERSITAS PERTAHANAN RI
FAKULTAS FMIPA MILITER

LEMBAR DISPOSISI DEKAN FMIPA MILITER

SURAT DARI : Dekan FSTP Unhan RI

NOMOR :
TANGGAL : 02/02/2024

NOMOR : B/ND/184/II/2024/FSTP

TANGGAL : 01 Februari 2024

PERIHAL : Permohonan Izin Meminta Data Transkrip Nilai IPK
Terakhir Seluruh Cohort untuk Keperluan Skripsi
Kadet Mahasiswa Informatika FSTP Unhan RI

KLASIFIKASI
☐ SANGAT RAHASIA
☐ RAHASIA
☒ BIASA
DERAJAT
☐ KILAT
☒ SEGERA

KEPADA YTH.

<input checked="" type="checkbox"/> WAKIL DEKAN BID. AK			<input type="checkbox"/> WAKIL DEKAN BID. KUUM		
<input checked="" type="checkbox"/> KAPRODI MATEMATIKA FMIPA MILITER	<input checked="" type="checkbox"/> KAPRODI BIOLOGI FMIPA MILITER	<input checked="" type="checkbox"/> KAPRODI FISIKA FMIPA MILITER	<input checked="" type="checkbox"/> KAPRODI KIMIA FMIPA MILITER	<input checked="" type="checkbox"/> KALAB FMIPA MILITER	<input checked="" type="checkbox"/> KASUBBAG UMUM FMIPA MILITER
PETUNJUK DEKAN					
<input type="checkbox"/> Acc / Selesaikan/ Tindak Lanjuti <input type="checkbox"/> Untuk dilaksanakan sesuai ketentuan <input type="checkbox"/> Untuk Diketahui / Sebagai Info <input type="checkbox"/> Persiapkan dengan baik dan Koordinasikan <input type="checkbox"/> Tanggapan / Saran <input type="checkbox"/> Laporkan Perkembangannya <input type="checkbox"/> Simpan File <input type="checkbox"/> Spri Dekan ingatkan / Catat <input type="checkbox"/> Infokan Ke Dosen <input type="checkbox"/> Infokan Ke Mahasiswa <input type="checkbox"/> Agar dirapatkan		<input type="checkbox"/> Buatkan / Ajukan : <input type="checkbox"/> Slide Paparannya <input type="checkbox"/> Jawaban <input type="checkbox"/> Amanat / Sambutan <input type="checkbox"/> Makalah <input type="checkbox"/> SPRIN Ke Rektor <input type="checkbox"/> SPRIN Ke Warek <input type="checkbox"/> Nota Dinas Ke Rektor <input type="checkbox"/> Nota Dinas Ke <input type="checkbox"/> Surat Jalan		<input type="checkbox"/> Undangan/Pertemuan : <input type="checkbox"/> Dekan Hadir <input type="checkbox"/> Wakil <input type="checkbox"/> Dampingi <input type="checkbox"/> Acarakan <input type="checkbox"/> Tidak Bisa Hadir	
Paraf Kasubbag Um FMIPA MILITER :					

CATATAN DEKAN :

*Moam dapat difasilitasi' data' mengenai kepada
ketuhan dan peraturan yang berlaku*

DEKAN FMIPA MILITER	
Paraf:	Tgl:
<i>[Signature]</i>	12.24

CATATAN WADEK I :

WADEK I FMIPA MILITER	
Paraf:	Tgl:

CATATAN WADEK II :

WADEK II FMIPA MILITER	
Paraf:	Tgl: