**Jurnal Komputer Multidisipliner**

Vol.7 No.10, Oktober 2024
ISSN: 24559633

# ANALISIS PENGELOMPOKAN JENIS PEKERJAAN ALUMNI ILMU KOMPUTER MENGGUNAKAN ALGORITMA MEAN SHIFT CLUSTERING

**Muhammad Afif Nashi Ulwan[1], Siti Ananda Budiana Nasution[2], Lowis Tambunan[3], Bryan Ahmad Fahrezi[4], Arnita[5]**
**Universitas Negeri Medan**
E-mail: naashii.ulwan@gmail.com[1], nannst24@gmail.com[2], lowis11tambunan@gmail.com[3], bryanahmadf@gmail.com[4], arnita@unimed.ac.id[5]

**Abstrak**

Purpose: Penelitian ini bertujuan untuk mengelompokkan mahasiswa ilmu komputer berdasarkan karakteristik akademik dan keterampilan teknis mereka menggunakan algoritma Mean Shift Clustering. Analisis ini dilakukan untuk memahami variasi dalam profil mahasiswa, seperti usia, IPK (GPA), serta keterampilan teknis (Python, SQL, dan Java). Tujuan dari pengelompokan ini adalah untuk membantu institusi pendidikan dalam menyusun strategi pembelajaran yang lebih efektif, dengan mempertimbangkan perbedaan karakteristik mahasiswa. Method/Study design/approach: Penelitian ini menggunakan algoritma Mean Shift, yang tidak memerlukan penentuan jumlah cluster secara eksplisit. Algoritma ini memungkinkan identifikasi struktur alami dalam data, membuatnya lebih sesuai untuk dataset yang kompleks dan non-linear. Visualisasi hasil clustering menggunakan Principal Component Analysis (PCA) dan t-Distributed Stochastic Neighbor Embedding (t-SNE) memberikan gambaran yang jelas tentang distribusi data dalam cluster. Result/Finding: Kesimpulan dari penelitian ini adalah bahwa algoritma Mean Shift mampu mengelompokkan mahasiswa berdasarkan profil akademik dan keterampilan, yang dapat digunakan untuk mendukung program pendidikan yang lebih adaptif dan tepat sasaran, seperti program akselerasi bagi mahasiswa dengan prestasi tinggi dan dukungan tambahan bagi yang memerlukan bantuan akademik lebih lanjut. NoveltyOriginality/Value:: belum ada penelitian yang secara spesifik melakukan analisis pengelompokan jenis pekerjaan alumni Ilmu Komputer menggunakan algoritma Mean Shift Clustering. Kebanyakan studi sebelumnya menggunakan metode clustering seperti K-Means atau Hierarchical Clustering.

**Kata Kunci** — Clustering, K-Means, PCA, t-SNE, Profil Mahasiswa, Pendidikan, Mean Shift.

## 1. PENDAHULUAN

### Data Mining

Data mining adalah proses penemuan pola-pola menarik dan berguna dalam kumpulan data yang besar. Salah satu metode yang digunakan dalam data mining adalah clustering, yang bertujuan untuk mengelompokkan data menjadi beberapa kelompok (cluster) berdasarkan kemiripan atau kesamaan karakteristik [1].

### Clustering

Clustering atau klasterisasi adalah sebuah metode untuk mengelompokan data. Clustering sebuah proses untuk mengelompokan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum [2].

### Mean Shift Clustering

Mean Shift adalah metode clustering non-parametrik yang tidak memerlukan penentuan jumlah cluster secara eksplisit seperti algoritma K-Means. Algoritma ini

bekerja dengan mengidentifikasi mode (puncak) dalam fungsi kepadatan probabilitas dari data. Mean Shift memindahkan setiap titik data menuju pusat massa data dalam lingkungannya yang ditentukan oleh bandwidth. Proses ini berlanjut hingga konvergensi, menghasilkan cluster sebagai area kepadatan tinggi yang terpisah. Mean Shift cocok digunakan dalam pengelompokan pekerjaan karena dapat secara adaptif menemukan struktur alami dalam data [3]. Algoritma Mean Shift merupakan metode pengelompokan dan analisis data yang digunakan untuk menemukan pusat klaster dalam data [4].

**Algoritma Mean Shift dalam Pengelompokan Pekerjaan**

Algoritma Mean Shift telah digunakan secara luas dalam berbagai aplikasi, termasuk segmentasi citra, analisis data, dan pengelompokan objek. Dalam konteks pengelompokan pekerjaan alumni, Mean Shift dapat menangani distribusi data yang kompleks dan non-linear, karena ia tidak mengasumsikan bentuk atau jumlah cluster secara a priori [5]. Dengan demikian, algoritma ini memungkinkan identifikasi cluster yang lebih alami berdasarkan pola pekerjaan yang sebenarnya ada dalam data.

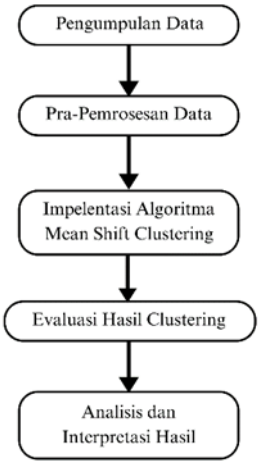
**Keunggulan Mean Shift**

Salah satu keunggulan Mean Shift dibandingkan algoritma clustering lain seperti K-Means adalah kemampuannya untuk mengidentifikasi jumlah cluster secara otomatis. K-Means mengharuskan pengguna menentukan jumlah cluster sebelumnya, yang bisa menjadi tantangan ketika distribusi data tidak diketahui [6]. Mean Shift, di sisi lain, dapat menemukan cluster dengan jumlah dan bentuk yang bervariasi, memberikan fleksibilitas yang lebih besar dalam pengelompokan data dengan distribusi yang tidak diketahui, seperti data pekerjaan alumni.

**Implementasi Mean Shift**

Implementasi Mean Shift dalam analisis data membutuhkan pemilihan bandwidth yang tepat, karena parameter ini menentukan ukuran lingkungan di sekitar setiap titik data. Bandwidth yang terlalu kecil dapat menghasilkan terlalu banyak cluster, sedangkan bandwidth yang terlalu besar dapat menggabungkan cluster yang seharusnya terpisah. Oleh karena itu, pemilihan bandwidth harus dilakukan dengan cermat berdasarkan distribusi data pekerjaan alumni yang dianalisis .

## 2. METODE PENELITIAN




*Gambar 1 Metode*

1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diambil dari Kaggle dengan judul Computer Science Students Dataset. Dataset ini mencakup informasi tentang mahasiswa

ilmu komputer dari sebuah universitas fiksi. Atribut yang terdapat dalam dataset ini meliputi:

- Student ID: Pengenal unik untuk setiap mahasiswa.
- Name: Nama mahasiswa.
- Gender: Jenis kelamin mahasiswa.
- Age: Usia mahasiswa.
- GPA: Nilai rata-rata akademik mahasiswa (Grade Point Average).
- Major: Bidang studi dalam ilmu komputer.
- Interested Domain: Area minat dalam bidang ilmu komputer.
- Projects: Proyek yang telah diselesaikan oleh mahasiswa.
- Python: Tingkat keterampilan mahasiswa dalam pemrograman Python.
- SQL: Tingkat keterampilan mahasiswa dalam pemrograman SQL.
- Java: Tingkat keterampilan mahasiswa dalam pemrograman Java.
- Future Career: Jalur karir atau aspirasi pekerjaan yang diinginkan oleh mahasiswa (variabel target).

Dataset ini dirancang untuk memberikan wawasan tentang prestasi akademik, aspirasi karir, dan keterampilan teknis mahasiswa dalam bidang ilmu komputer. Dataset ini sangat berguna untuk tugas-tugas seperti pemodelan prediktif untuk memahami faktor-faktor yang mempengaruhi pilihan karir mahasiswa ilmu komputer.

2. Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk membersihkan dan mempersiapkan data sebelum analisis. Langkah-langkah yang dilakukan meliputi:

- Pembersihan Data, yaitu Menghilangkan data duplikat, mengisi nilai yang hilang (missing values), dan menghapus data yang tidak relevan.
- Normalisasi, untuk memastikan bahwa variabel memiliki skala yang sama, sehingga tidak ada variabel yang dominan mempengaruhi hasil clustering.
- Pencatatan Kategori, yaitu Mengubah data kategorikal (misalnya, jenis pekerjaan atau sektor industri) menjadi bentuk numerik menggunakan teknik seperti one-hot encoding atau label encoding.

3. Implementasi Algoritma Mean Shift Clustering

Pada tahap ini, algoritma Mean Shift diterapkan pada data yang telah diproses. Langkah-langkah yang dilakukan meliputi:

- Penentuan Bandwidth

Menentukan parameter bandwidth yang optimal untuk algoritma Mean Shift. Bandwidth ini menentukan radius pencarian dalam ruang data dan mempengaruhi jumlah dan ukuran cluster yang dihasilkan.

- Pelaksanaan Mean Shift

Algoritma Mean Shift diimplementasikan dengan menghitung pusat massa (mean) dari titik data dalam lingkup bandwidth, dan secara iteratif memindahkan titik data menuju pusat massa tersebut hingga konvergensi tercapai.

- Pembentukan Cluster

Setelah konvergensi, titik data yang berkumpul di sekitar pusat massa yang sama dianggap sebagai satu cluster. Konvergensi dalam algoritma Mean Shift merujuk pada kondisi di mana titik data tidak lagi berpindah secara signifikan menuju pusat massa. Proses ini terjadi ketika setiap titik data telah mencapai pusat kepadatan tertinggi dalam lingkup bandwidth-nya. Dengan kata lain, setelah proses iteratif selesai dan tidak ada lagi perubahan posisi yang berarti, cluster terbentuk secara stabil. Setiap cluster yang dihasilkan akan mewakili kelompok dengan karakteristik tertentu, seperti keterampilan

atau minat dalam pekerjaan yang dipilih oleh alumni.

mungkin bisa tambahkan penjelasan apa itu konvergensi

4. Evaluasi Hasil Clustering

Hasil clustering dievaluasi untuk menilai kualitas dan validitas pengelompokan. Beberapa metrik yang digunakan meliputi:

- Silhouette Score yang digunakan untuk mengukur seberapa baik titik data berada di dalam cluster yang benar dan seberapa terpisah cluster yang berbeda. Silhoutte Coefficient merupakan metode evaluasi atau validasi algoritma clustering yang mengukur kualitas cluster yang terbentuk. [ ]

Hasil Perhitungan Silhouette Score adalah = 0.3607039021887175 mungkin bisa tambahkan perhitungan dari silhoutte score kemudian masukkan sitasinya.

- Visualisasi seperti scatter plot atau PCA (Principal Component Analysis) digunakan untuk memvisualisasikan hasil clustering dalam dua atau tiga dimensi, sehingga pola pengelompokan dapat diamati.

5. Analisis dan Interpretasi Hasil

Setelah clustering berhasil dilakukan, analisis lebih lanjut dilakukan untuk menginterpretasikan hasil. Setiap cluster dianalisis untuk memahami karakteristik umum dari jenis pekerjaan yang termasuk di dalamnya. Hasil ini dapat memberikan wawasan tentang pola karir alumni Ilmu Komputer dan hubungan antara pendidikan yang ditempuh dengan jenis pekerjaan yang dipilih.

Bagian metode sudah cukup bagus, namun belum ada pengenalan dari data seperti dari mana data diambil, berapa jumlah baris data, dan sebagainya. Umumnya struktru dari metode seperti berikut :

- Pengenalan data
- Preprocessing data
- implementasi algoritma clustering
- evaluasi

Kemudian masih banyak yang bisa ditambahkan pada setiap bagian tahapannya dan bisa ambil perhitungan-perhitungan dan penjelasan dari jurnal.

## 3. HASIL DAN PEMBAHASAN

1. Hasil Clustering Menggunakan Means Shift

| | Student ID | Name | Gender | Age | GPA | Major | Interested Domain |
|---|---|---|---|---|---|---|---|
| 0 | 1 | John Smith | 1 | 21 | 3.5 | 0 | 0 |
| 1 | 2 | Alice Johnson | 0 | 20 | 3.2 | 0 | 10 |
| 2 | 3 | Robert Davis | 1 | 22 | 3.8 | 0 | 24 |
| 3 | 4 | Emily Wilson | 0 | 21 | 3.7 | 0 | 26 |
| 4 | 5 | Michael Brown | 1 | 23 | 3.4 | 0 | 7 |

| | Projects | Future Career | Python | SQL | Java | Cluster |
|---|---|---|---|---|---|---|
| 0 | 8 | 21 | 0.5 | 0.5 | 1.0 | 1 |
| 1 | 15 | 7 | 0.0 | 0.5 | 1.0 | 1 |
| 2 | 21 | 29 | 0.5 | 0.5 | 0.0 | 0 |
| 3 | 26 | 32 | 1.0 | 0.5 | 0.5 | 0 |
| 4 | 40 | 18 | 0.0 | 1.0 | 0.5 | 0 |

Clustering dilakukan menggunakan algoritma Mean Shift dengan memanfaatkan pipeline untuk preprocessing dan clustering. Pada gambar ini, terlihat contoh hasil clustering dengan menampilkan beberapa mahasiswa yang telah dikelompokkan ke dalam Cluster 0 dan Cluster 1. Hasil clustering ini didasarkan pada usia, GPA, serta keterampilan

teknis seperti Python, SQL, dan Java.

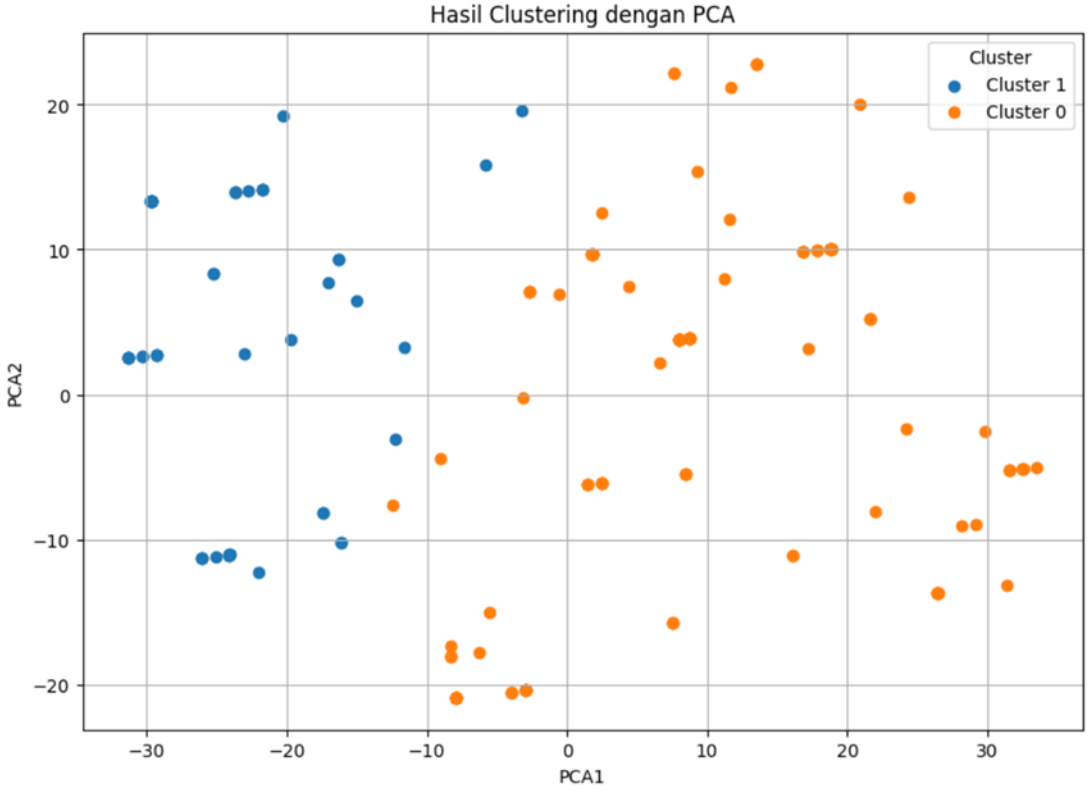
- Cluster 0 terdiri dari mahasiswa seperti John Smith, yang memiliki usia 21 tahun dengan GPA 3.5, dan menunjukkan keterampilan yang lebih kuat dalam Java (nilai 1.0), sedangkan keterampilan di Python dan SQL sedikit lebih lemah (nilai 0.5).
- Cluster 1 diwakili oleh mahasiswa seperti Alice Johnson, yang berusia 20 tahun dengan GPA 3.2 dan lebih fokus pada keterampilan di SQL (nilai 0.5), tetapi menunjukkan keterampilan yang lebih lemah di Python dan Java.

Penggunaan Mean Shift memungkinkan deteksi cluster secara otomatis, tanpa perlu menentukan jumlah cluster di awal. Algoritma ini mengidentifikasi dua cluster yang berbeda, di mana masing-masing cluster memiliki karakteristik unik, seperti variasi dalam usia dan keterampilan teknis.

2. Visualisasi dengan PCA dan t-SNE

Setelah melakukan clustering dengan Mean Shift, dilakukan visualisasi hasil clustering menggunakan Principal Component Analysis (PCA) dan t-SNE (t-Distributed Stochastic Neighbor Embedding). Kedua teknik ini membantu dalam memahami distribusi dan karakteristik data dalam setiap cluster yang terbentuk.

**Visualisasi dengan PCA**

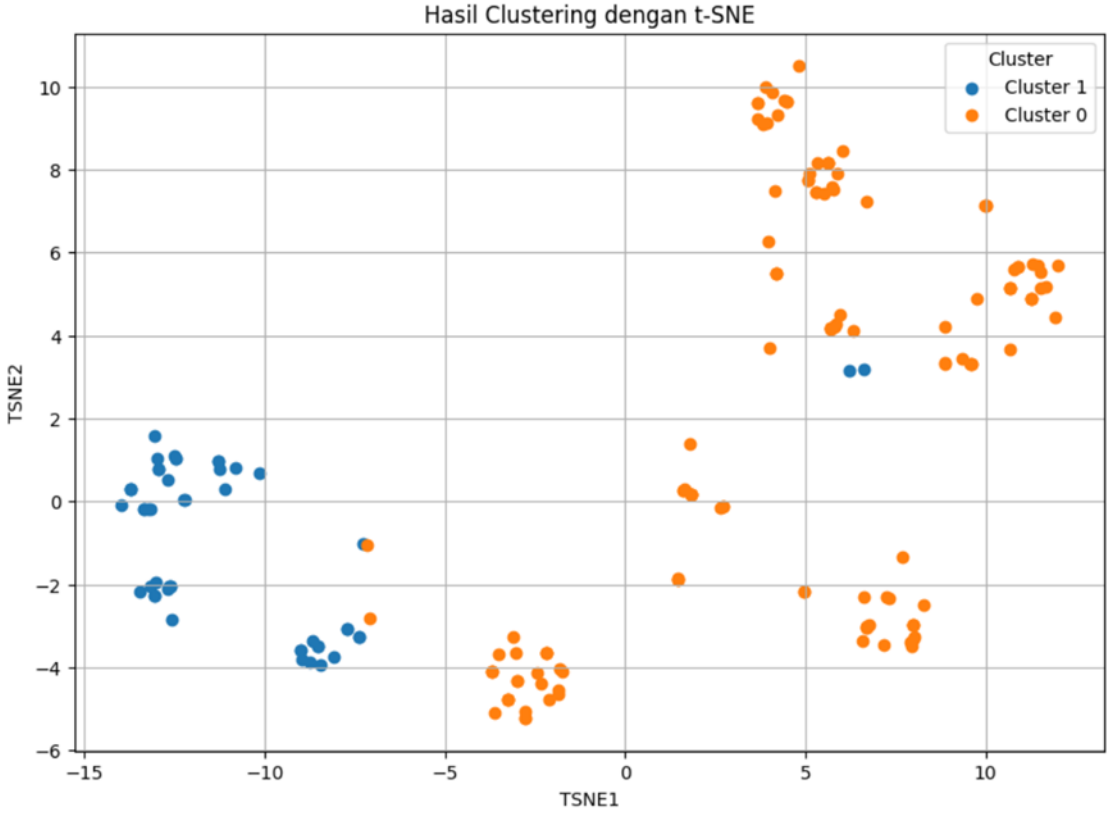



Gambar 2. Visualisasi Clustering dengan PCA

- PCA digunakan untuk mereduksi dimensi data menjadi dua komponen utama (PCA1 dan PCA2), seperti yang terlihat pada Gambar 2. Pada visualisasi ini, Cluster 0 (warna oranye) memiliki distribusi yang lebih tersebar di sepanjang sumbu PCA1, menandakan adanya variasi besar dalam variabel seperti usia dan GPA.
- Cluster 1 (warna biru), di sisi lain, terlihat lebih terpusat dan terfokus. Ini menandakan bahwa anggota Cluster 1 memiliki lebih banyak kesamaan dalam profil akademik mereka, seperti GPA dan keterampilan teknis (Python, SQL, Java).
- Penyebaran luas Cluster 0 menunjukkan adanya heterogenitas yang lebih tinggi, yang

mengindikasikan bahwa mahasiswa dalam cluster ini memiliki latar belakang akademik dan keterampilan yang lebih bervariasi.

**Visualisasi dengan t-SNE**




Gambar 3. Visualisasi Clustering dengan t-SNE

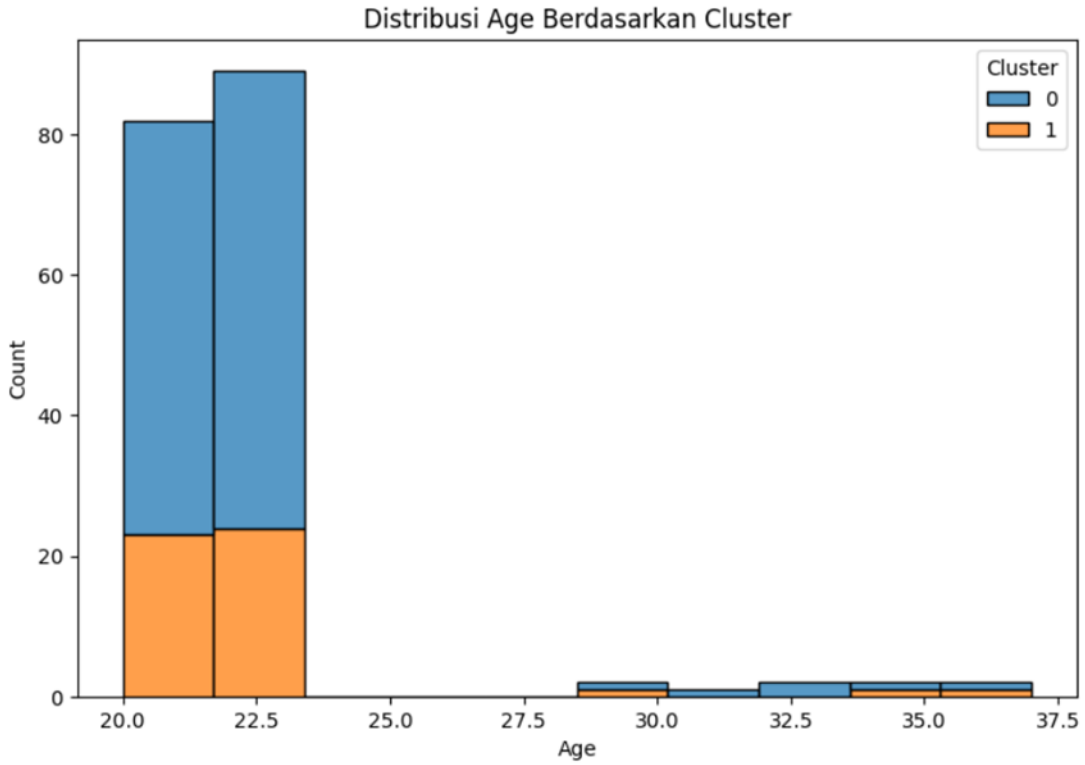
- t-SNE memberikan visualisasi yang lebih mendalam dan detail dari distribusi data dalam dua dimensi (TSNE1 dan TSNE2), seperti yang terlihat pada Gambar 3.
- Hasil visualisasi ini menunjukkan bahwa Cluster 0 tetap menunjukkan penyebaran yang luas dengan anggota yang bervariasi dalam hal usia dan GPA, serta keterampilan teknis.
- Sebaliknya, Cluster 1 sangat terfokus dengan anggota yang dikelompokkan lebih erat, memperjelas bahwa mahasiswa dalam cluster ini memiliki kemiripan kuat dalam hal tujuan karir, minat dalam domain tertentu, atau keterampilan teknis.

t-SNE bekerja dengan menjaga hubungan lokal di dalam data, sehingga distribusi cluster yang lebih padat dalam visualisasi ini mencerminkan bahwa mahasiswa dalam Cluster 1 lebih mirip satu sama lain dibandingkan dengan anggota Cluster 0 yang lebih bervariasi.

Dengan demikian, visualisasi melalui PCA dan t-SNE memberikan wawasan penting mengenai perbedaan antar cluster. Cluster 0 menunjukkan variabilitas yang lebih besar, sedangkan Cluster 1 lebih homogen dalam hal akademik dan keterampilan teknis. Kedua metode visualisasi ini memberikan bukti kuat bahwa algoritma Mean Shift mampu mengelompokkan mahasiswa berdasarkan kemiripan profil akademik dan keterampilan mereka.

3. Distribusi Fitur dalam Setiap Cluster dengan Histogram

Histogram digunakan untuk memvisualisasikan distribusi fitur-fitur utama seperti usia dan GPA dalam setiap cluster yang terbentuk dari proses clustering dengan Mean Shift. Dengan menggunakan histogram, kita dapat melihat bagaimana fitur-fitur ini tersebar di antara mahasiswa yang dikelompokkan ke dalam Cluster 0 dan Cluster 1.

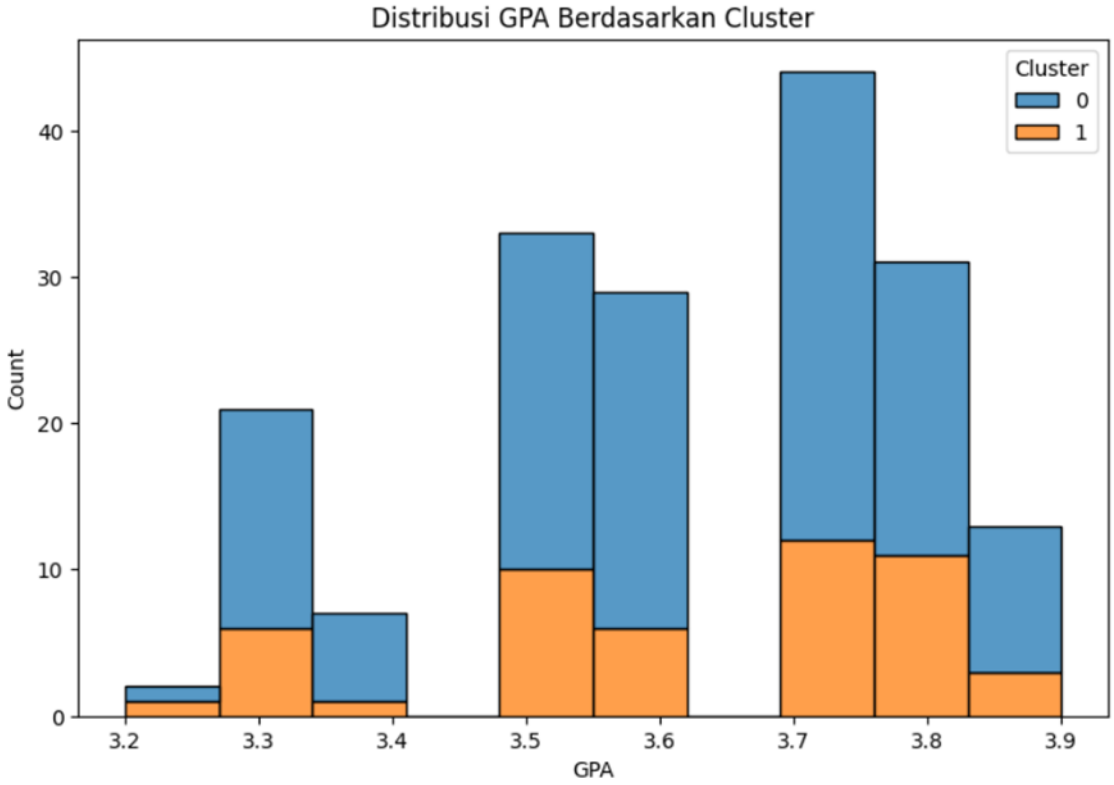



Gambar 4. Distribusi Usia dan GPA Berdasarkan Cluster

Pada histogram Distribusi Usia, kita bisa melihat bahwa:

- Cluster 0 (ditandai dengan warna biru) mendominasi mahasiswa dengan usia antara 20 hingga 22 tahun, yang menunjukkan bahwa sebagian besar mahasiswa di cluster ini berada dalam rentang usia yang relatif muda dan sesuai dengan usia rata-rata mahasiswa sarjana.
- Cluster 1 (ditandai dengan warna oranye) juga mencakup mahasiswa dalam rentang usia yang sama, yaitu 20 hingga 22 tahun, tetapi jumlahnya lebih sedikit dibandingkan Cluster 0.

Selain itu, terdapat outlier pada kedua cluster yang berusia lebih dari 25 tahun. Jumlah mahasiswa dalam kategori usia di atas 25 tahun sangat kecil, menunjukkan bahwa mahasiswa yang lebih tua jarang ditemukan di dalam dataset ini, dan mereka mungkin merupakan mahasiswa yang kembali melanjutkan studi setelah beberapa tahun bekerja atau melakukan jeda dalam pendidikan.




Gambar 5 Distribusi GPA Berdasarkan Cluster

Pada histogram Distribusi GPA, kita dapat melihat bagaimana GPA tersebar di setiap cluster:

- Cluster 0 memiliki proporsi yang lebih besar dari mahasiswa dengan GPA tinggi di rentang 3.7 hingga 3.8, menunjukkan bahwa mayoritas mahasiswa dalam cluster ini memiliki performa akademik yang baik. Ini menunjukkan bahwa Cluster 0 berisi mahasiswa yang lebih konsisten dalam prestasi akademik.
- Cluster 1 menunjukkan distribusi yang lebih bervariasi, dengan mahasiswa yang memiliki GPA lebih rendah mulai dari 3.2 hingga 3.4. Namun, meskipun Cluster 1 mencakup mahasiswa dengan GPA lebih rendah, sebagian dari mereka juga memiliki GPA yang cukup tinggi hingga 3.7.

Ini menunjukkan adanya heterogenitas dalam Cluster 1, di mana mahasiswa dalam cluster ini bervariasi dalam hal prestasi akademik. Sebaliknya, Cluster 0 lebih homogen dengan mayoritas mahasiswa memiliki GPA yang cukup tinggi, mengindikasikan stabilitas dalam performa akademik di antara anggota cluster ini.

Interpretasi Distribusi Fitur dalam Setiap Cluster

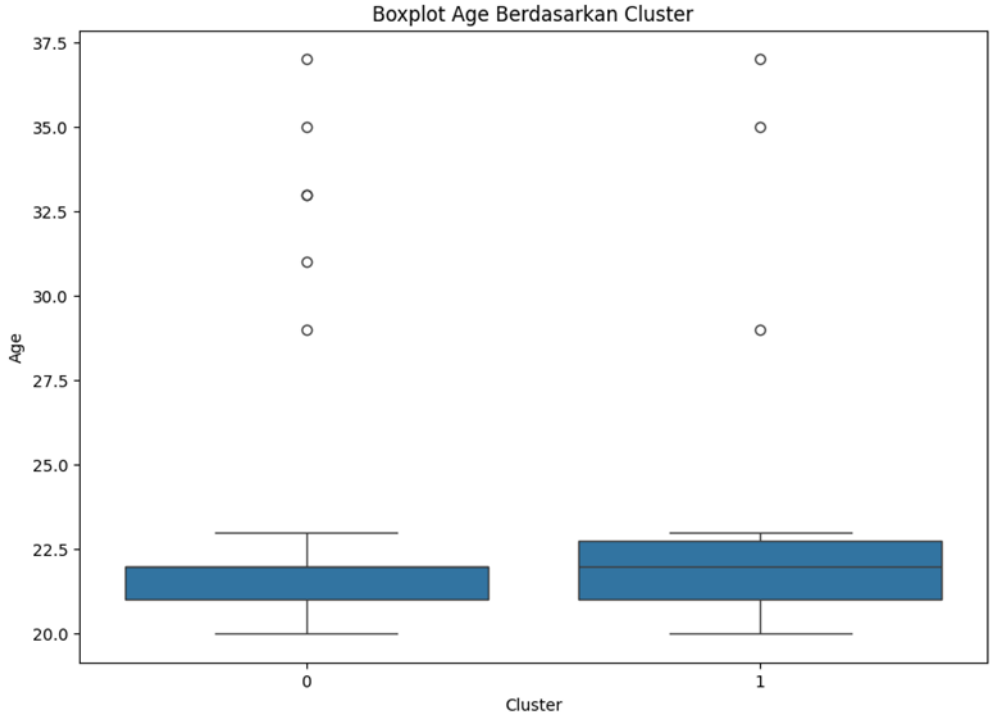
Dengan menggunakan histogram, kita dapat menarik beberapa kesimpulan penting terkait karakteristik setiap cluster:

1. Cluster 0 didominasi oleh mahasiswa dengan usia muda dan GPA yang tinggi, menunjukkan bahwa mereka mungkin lebih siap secara akademik dan membutuhkan tantangan tambahan atau program pembelajaran yang lebih intensif untuk menjaga motivasi dan keterlibatan mereka.
2. Cluster 1, meskipun juga terdiri dari mahasiswa dengan usia yang serupa dengan Cluster 0, memiliki GPA yang lebih beragam, yang mungkin mencerminkan perbedaan dalam tingkat kemampuan dan latar belakang akademik. Mahasiswa dalam Cluster 1 mungkin memerlukan dukungan akademik tambahan untuk meningkatkan performa akademik mereka, terutama bagi yang memiliki GPA lebih rendah.

Dengan analisis distribusi fitur ini, institusi pendidikan dapat lebih mudah dalam merancang program-program pembelajaran yang lebih adaptif, menyesuaikan kurikulum dengan kebutuhan dan kemampuan masing-masing cluster. Mahasiswa di Cluster 0 dapat diuntungkan dari program akselerasi, sementara mahasiswa di Cluster 1 mungkin membutuhkan pendekatan pembelajaran yang lebih personal untuk mendukung peningkatan akademik mereka.

3. Analisis Distribusi Fitur dengan Boxplot

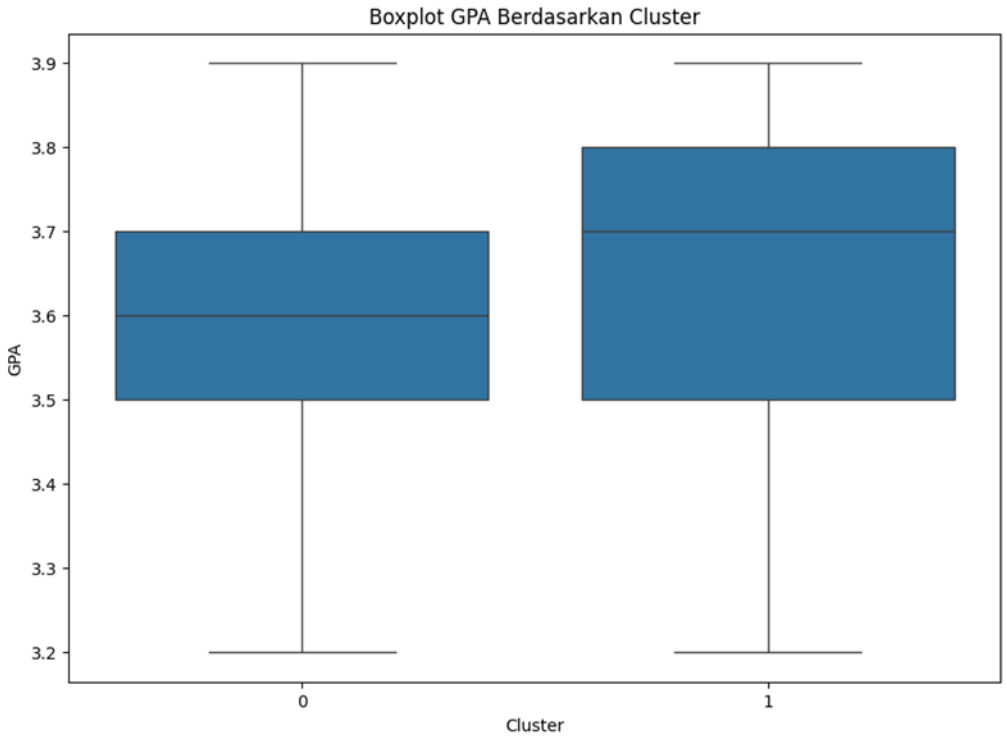
Boxplot digunakan untuk memvisualisasikan sebaran dan variasi dua fitur utama, yaitu usia dan GPA, dalam setiap cluster yang dihasilkan oleh algoritma Mean Shift. Analisis ini membantu mengidentifikasi distribusi data serta outlier dalam masing-masing cluster. Dengan visualisasi boxplot, kita dapat melihat median, kuartil pertama dan ketiga, serta rentang variabilitas data di dalam cluster.




Gambar 6 Boxplot Age Berdasarkan Cluster

Pada boxplot distribusi usia, terdapat beberapa poin yang dapat diambil dari hasil visualisasi:

- Median usia pada kedua cluster berada di sekitar 22 tahun, menunjukkan bahwa mayoritas mahasiswa dalam kedua cluster memiliki usia yang serupa. Ini mengindikasikan bahwa usia bukanlah faktor utama yang membedakan Cluster 0 dan Cluster 1.
- Rentang usia pada kedua cluster sangat sempit, yaitu antara 20 hingga 22 tahun, yang menunjukkan bahwa sebagian besar mahasiswa dalam dataset ini adalah mahasiswa sarjana dengan usia yang masih muda.
- Outlier terlihat jelas pada kedua cluster, terutama pada usia lebih dari 25 tahun hingga 37.5 tahun. Ini mengindikasikan bahwa ada sejumlah kecil mahasiswa yang lebih tua yang mungkin kembali ke pendidikan setelah beberapa tahun bekerja atau mengambil jeda dalam studi mereka. Outlier tersebut berada pada jarak yang cukup jauh dari rentang usia mayoritas di kedua cluster.




Gambar 7 Boxplot GPA Berdasarkan Cluster

Boxplot untuk distribusi GPA memperlihatkan perbedaan yang lebih signifikan dibandingkan dengan distribusi usia:

- Cluster 0 menunjukkan median GPA sekitar 3.6, dengan rentang GPA yang cukup lebar mulai dari 3.2 hingga 3.9. Ini menunjukkan bahwa meskipun mayoritas mahasiswa dalam cluster ini memiliki GPA yang relatif baik, ada beberapa variasi dalam prestasi akademik mereka.
- Cluster 1, di sisi lain, memiliki median GPA yang sedikit lebih tinggi yaitu sekitar 3.7. Rentang GPA di cluster ini lebih bervariasi dibandingkan Cluster 0, menunjukkan bahwa mahasiswa dalam Cluster 1 lebih heterogen dalam hal prestasi akademik.
- Tidak ada outlier yang mencolok dalam distribusi GPA, menunjukkan bahwa performa akademik mahasiswa di kedua cluster berada dalam batas yang dapat diterima.

Interpretasi Distribusi Fitur dengan Boxplot

Beberapa kesimpulan penting dapat diambil dari visualisasi boxplot ini:

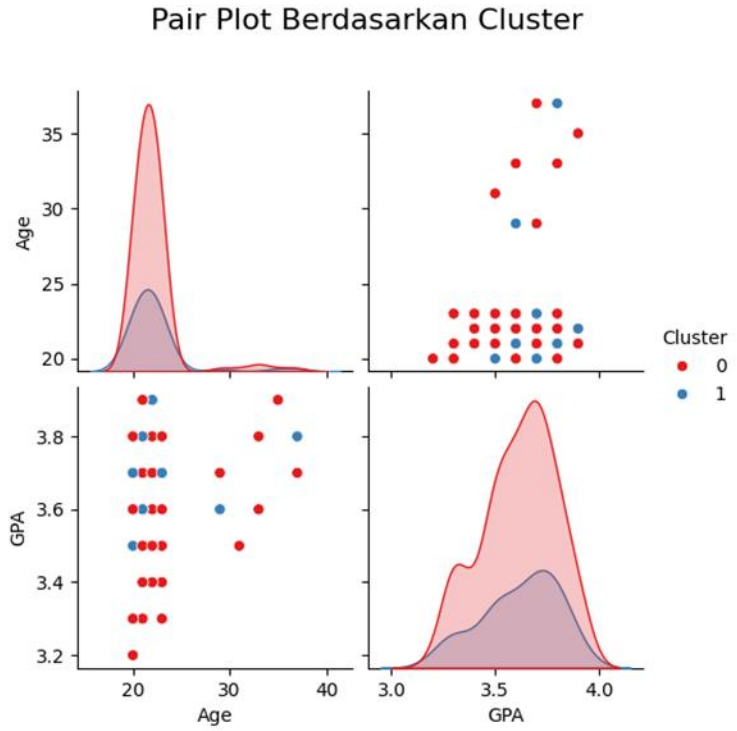
1. Usia: Meskipun usia median pada kedua cluster sangat mirip, keberadaan outlier pada rentang usia yang lebih tua menunjukkan bahwa ada sejumlah kecil mahasiswa dewasa yang kembali ke pendidikan setelah jeda karier atau karena faktor lain. Sebagian besar mahasiswa dalam dataset ini masih berada pada rentang usia tradisional mahasiswa sarjana.
2. GPA: Cluster 1 cenderung memiliki GPA yang lebih tinggi dan lebih bervariasi dibandingkan dengan Cluster 0, yang menunjukkan bahwa anggota Cluster 1 mungkin lebih beragam dalam kemampuan akademik mereka. Mahasiswa dalam cluster ini mungkin berasal dari latar belakang yang lebih beragam, baik dalam hal prestasi akademik maupun pengalaman pendidikan.
3. Outlier pada usia dapat menjadi indikasi bahwa sebagian kecil mahasiswa mungkin membutuhkan program pembelajaran yang lebih fleksibel atau khusus, misalnya program akselerasi atau dukungan pendidikan tambahan. Sementara itu, variasi dalam GPA pada kedua cluster menunjukkan bahwa ada peluang untuk memberikan

intervensi akademik yang berbeda-beda sesuai dengan kebutuhan individu dalam setiap cluster.

Dengan memahami distribusi fitur melalui boxplot, kita dapat membuat keputusan yang lebih tepat terkait strategi pembelajaran yang personal dan penyediaan dukungan akademik untuk mahasiswa dalam berbagai cluster.

4. Analisis Hubungan Antar Fitur dengan Pair Plot

Pair Plot digunakan untuk memvisualisasikan hubungan antara beberapa fitur, yaitu usia dan GPA, dalam setiap cluster yang dihasilkan dari algoritma Mean Shift. Dengan pair plot, kita dapat melihat interaksi antar variabel dalam setiap cluster, serta distribusi masing-masing fitur dalam bentuk diagonal.




Pair plot menampilkan hubungan antar fitur dengan menggunakan titik-titik yang dibedakan berdasarkan cluster. Beberapa hal yang bisa kita amati dari pair plot ini:

1. Distribusi Usia di Tiap Cluster:
   - Cluster 0 (ditandai dengan warna merah) sebagian besar terdiri dari mahasiswa dengan usia muda, terutama pada rentang 20 hingga 22 tahun. Rentang usia ini sangat terkonsentrasi pada usia mahasiswa sarjana yang relatif muda.
   - Cluster 1 (ditandai dengan warna biru) juga mencakup mahasiswa pada rentang usia yang sama, tetapi distribusinya lebih tersebar. Meskipun begitu, masih ada beberapa outlier yang lebih tua, yaitu di atas 25 tahun.
2. Distribusi GPA di Tiap Cluster:

- Sebaran GPA untuk Cluster 0 menunjukkan bahwa sebagian besar mahasiswa memiliki GPA antara 3.5 hingga 4.0, dengan konsentrasi yang tinggi di sekitar GPA 3.7. Mahasiswa dalam Cluster 0 cenderung memiliki performa akademik yang lebih seragam dan relatif tinggi.
- Cluster 1, di sisi lain, memiliki GPA yang lebih bervariasi. Sebagian besar mahasiswa memiliki GPA antara 3.2 hingga 3.8, dan ada beberapa mahasiswa yang memiliki GPA lebih rendah di bawah 3.5. Ini menunjukkan bahwa Cluster 1 memiliki tingkat performa akademik yang lebih beragam dibandingkan Cluster 0.

3. Hubungan Antar Fitur Usia dan GPA:

- Dalam Cluster 0, hubungan antara usia dan GPA tidak terlihat begitu jelas, karena mahasiswa dalam rentang usia yang sama (20-22 tahun) memiliki performa akademik yang cukup tinggi dan seragam. Artinya, usia tidak mempengaruhi performa akademik secara signifikan dalam cluster ini.
- Dalam Cluster 1, hubungan antara usia dan GPA juga tidak terlihat secara langsung. Namun, mahasiswa dengan usia lebih tua (di atas 25 tahun) cenderung memiliki GPA yang lebih rendah, meskipun pola ini tidak signifikan. Mahasiswa yang lebih tua dalam cluster ini mungkin memerlukan dukungan akademik lebih lanjut untuk mencapai GPA yang lebih tinggi.

**Interpretasi Pair Plot**

Dari analisis pair plot, kita dapat melihat bahwa:

1. Cluster 0: Sebagian besar mahasiswa dalam cluster ini memiliki usia yang muda dan GPA yang tinggi, menunjukkan bahwa kelompok ini mungkin terdiri dari mahasiswa yang berprestasi secara akademik. Mereka mungkin tidak memerlukan dukungan tambahan, tetapi bisa diuntungkan dari program akselerasi atau tantangan akademik tambahan.
2. Cluster 1: Cluster ini menunjukkan variasi yang lebih besar dalam GPA dan usia, dengan beberapa mahasiswa memiliki GPA yang lebih rendah dan berada di rentang usia yang lebih tua. Kelompok ini mungkin memerlukan pendekatan yang lebih personal dan fleksibel, seperti bimbingan tambahan atau program remedial untuk mahasiswa yang berprestasi rendah.

Secara keseluruhan, pair plot membantu dalam memahami kompleksitas hubungan antar variabel, memberikan wawasan yang lebih mendalam tentang pola-pola yang mungkin tidak langsung terlihat pada visualisasi tunggal seperti histogram atau boxplot. Wawasan ini penting untuk membantu institusi pendidikan dalam merancang program yang sesuai untuk mendukung setiap kelompok mahasiswa secara optimal.

## 4. KESIMPULAN

Penelitian ini berhasil melakukan pengelompokan mahasiswa ilmu komputer berdasarkan profil akademik dan keterampilan teknis menggunakan algoritma Mean Shift Clustering. Algoritma ini terbukti efektif dalam mengidentifikasi pola alami dalam data tanpa memerlukan penentuan jumlah cluster di awal, menghasilkan dua cluster utama dengan karakteristik yang berbeda. Cluster 0 didominasi oleh mahasiswa muda dengan usia berkisar antara 20 hingga 22 tahun, serta memiliki GPA yang lebih seragam dan tinggi, dengan sebagian besar berada di rentang 3.5 hingga 4.0, menunjukkan potensi akademik yang baik. Kelompok ini mungkin memerlukan program akselerasi atau tantangan akademik tambahan untuk menjaga keterlibatan mereka. Sebaliknya, Cluster 1 memiliki distribusi usia yang lebih beragam, dengan beberapa mahasiswa berusia lebih dari 25 tahun, serta GPA yang lebih bervariasi dengan sebagian mahasiswa menunjukkan prestasi akademik yang lebih rendah. Mahasiswa dalam cluster ini mungkin membutuhkan pendekatan pendidikan yang lebih fleksibel, seperti bimbingan akademik tambahan atau program remedial. Visualisasi menggunakan teknik PCA dan t-SNE menunjukkan pemisahan yang jelas antara cluster, dengan Cluster 0 memiliki distribusi yang lebih tersebar, sementara Cluster 1 lebih terfokus. Analisis lebih lanjut dengan histogram dan boxplot mengungkapkan bahwa sebagian besar mahasiswa dalam kedua cluster berada di usia muda, tetapi variasi dalam GPA di Cluster 1 mengindikasikan adanya heterogenitas akademik. Berdasarkan hasil ini, institusi pendidikan dapat merancang program yang lebih adaptif, menyediakan akselerasi untuk mahasiswa berprestasi di Cluster 0 dan memberikan

dukungan akademik bagi mereka yang memerlukan bimbingan di Cluster 1, sehingga strategi pembelajaran yang lebih efektif dan tepat sasaran dapat dikembangkan.

## DAFTAR PUSTAKA


J. Han, J. Pei, and M. Kamber, Data Mining: Concepts and Techniques. Elsevier, 2011.

R. Reliovani, N.Nadia, S. Husein, K. Z. Abdurrafi, C. R. Alhusni, and M. A. Khowarizmi. "Algoritma Mean Shift untuk Menentukan Segmentasi Pelanggan pada," Gunung Djati Conference Series, vol. 3, pp. 92-98, 2021.

R. Setyawan, "COMPARATIVE STUDY OF k-MEANS AND MEAN SHIFT CLUSTERING ALGORITHMS FOR WASTE DATA IN WEST JAVA PROVINCE," Journal of Engineering Science and Technology, vol. 19(3), pp. 869-879, 2024.

K. Fukunaga and L. Hostetler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," IEEE Transactions on Information Theory, vol. 21, pp. 32-40, 1975.

Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, pp. 790-799, Aug. 1995.

D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, May 2002.

M. Ankerst, M. M. Breunig, H. P. Kriegel, and J. Sander, "OPTICS: Ordering Points to Identify the Clustering Structure," ACM SIGMOD Record, vol. 28, pp. 49-60, 1999.

L. Abdallah and I, Shimshoni. "Mean Shift Clustering Algorithm for Data with Missing Values," Springer International Publishing, pp. 426-427, 2014.

G. J. Oyewole and G. A. Thopil "Data clustering: application and trends," Artifcial Intelligence Review, pp. 6339-6475, 2023.

D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach Toward Feature Space Analysis," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, pp. 603-619, May 2002.

A. Chhabra, "An Overview of Fairness in Clustering," IEEE Access, vol. 9, pp. 130698-130720, 2021.

M. A. Carreira-Perpin´an, "A review of mean-shift algorithms for clustering," pp. 1-28, 2015.

M. Schier, C. Reinders, and B. Rosenhalm."Constrained Mean Shift Clustering," pp. 235-243, 2022.

A. Al-Rahayfeh, S. Atiewi, A. Abuhussein, and M. Almiani. "Novel Approach to Task Scheduling and Load Balancing Using the Dominant Sequence Clustering and Mean Shift Clustering Algorithms," Future Internet, pp. 1-15, 2019.

Z. Liu, J.Lium, X. Xiao, H. Yuan, X. Li, J. Chang, and C. Zheng. "Segmentation of White Blood Cells through Nucleus Mark Watershed Operations and Mean Shift Clustering," Sensors, pp. 22561-22586, 2015.

G. Wu, "Histological image segmentation using fast mean shift clustering method," BioMedical Engineering OnLine, pp. 1-12, 2015.

L. Ai, "Application of mean-shift clustering to Blood oxygen level dependent functional MRI activation detection," BMC Medical Imaging, pp. 1-10, 2014.